Linköping Studies in Science and Technology

Dissertation No. 1187

Analytical tools and information-sharing methods supporting road safety organizations

Imad-Eldin Ali Abugessaisa

Linköping University
INSTITUTE OF TECHNOLOGY

Department of Computer and Information Science
Linköping University
SE-581 83 Linköping, Sweden

Linköping 2008

ISBN 978-91-7393-887-7
ISSN 0345-7524
Printed by LiU-Tryck Linköping 2008

ABSTRACT

A prerequisite for improving road safety are reliable and consistent sources of information about traffic and accidents, which will help assess the prevailing situation and give a good indication of their severity. In many countries there is under-reporting of road accidents, deaths and injuries, no collection of data at all, or low quality of information. Potential knowledge is hidden, due to the large accumulation of traffic and accident data. This limits the investigative tasks of road safety experts and thus decreases the utilization of databases. All these factors can have serious effects on the analysis of the road safety situation, as well as on the results of the analyses.

This dissertation presents a three-tiered conceptual model to support the sharing of road safety–related information and a set of applications and analysis tools. The overall aim of the research is to build and maintain an information-sharing platform, and to construct mechanisms that can support road safety professionals and researchers in their efforts to prevent road accidents. GLOBESAFE is a platform for information sharing among road safety organizations in different countries developed during this research.

Several approaches were used, First, requirement elicitation methods were used to identify the exact requirements of the platform. This helped in developing a conceptual model, a common vocabulary, a set of applications, and various access modes to the system. The implementation of the requirements was based on iterative prototyping. Usability methods were introduced to evaluate the users' interaction satisfaction with the system and the various tools. Second, a system-thinking approach and a technology acceptance model were used in the study of the Swedish traffic data acquisition system. Finally, visual data mining methods were introduced as a novel approach to discovering hidden knowledge and relationships in road traffic and accident databases. The results from these studies have been reported in several scientific articles.

Department of Computer and Information Science
Linköping University, SE-581 83 Linköping, Sweden

Linköping 2008


إلى روح والدي
1918 - 1990

Acknowledgments

This is the time to write the last pages of this dissertation. Despite the fact that my name stands alone on the cover, these efforts would never have ended without the support and encouragement of others.

I am deeply indebted to my supervisor, Erland Jungert. Erland has been an outstanding supervisor during the last years of my research; his support and encouragement have made this dissertation possible. During the course of my research I learned from Erland how to operate independently.

Besides Erland I have been supervised by Kenneth Asp and Åke Sivertun, both of whom shaped my ideas in different ways. Kenneth introduced me to the field of traffic safety. I am grateful for his support and valuable discussions at an early stage of the research and during implementation of GLOBESAFE.

Starting from the first day in Linköping that I met Åke Sivertun, he deserves a lot of credit. Åke devoted so much time and effort to guiding me through the research landscape and venture. Åke is not only a scientist and advisor but most importantly a kind person. I was more than grateful for his constant and unlimited support, his availability and co-authoring. I enjoy traveling with Åke and our pleasant discussions about everyday life.

My gratitude to Sture Hägglund for his advice on methodological matters and for reading and comments on Paper VI. I thank him for his kind support when things were down and for the midnight conversations.

I also want to thank Michaël Le Duc, for co-authoring some of the published articles in this dissertation and also for reading and valuable comments.

I want to thank the Department of Computer and Information Science at Linköping University for giving me the chance to do my research and use the departmental facilities and for financial support. Special thanks to Arne Jönsson for unlimited assistance and support.

I would like to thank Josef Muklik for agreeing to serve as opponent, and Bo Sundgren, Håkan Alm and Henrik Eriksson for their valuable time in reading my thesis and for accepting Erland’s request to serve on the examination committee.

I also want to acknowledge the great help from Ingela Dellby and Paul Norlén for effective language reviews on both the thesis and a number of the published articles. Your efforts improved my writing a lot.

Furthermore, I want to thank Lillemor Wallgren, Britt-Inger Karlson, Inger Emanuelsson, and Helene Eriksson for being supportive in practical issues concerning my study.

Special thanks to Carita Lilja for her help and assistance at the office and when I'm traveling, and to Stefan Karlsson, Rolf Nilsson and Göran Sedvall for technical assistance and solving my computer problems.

I'm grateful to the former and current members at Human-Centered Systems, where I met many colleagues and I want to take the opportunity to thank them: Henrik Eriksson for his valuable comments and discussion on the final version of the dissertation; Bahlol Rahimi, co-author and travel-mate in Sweden and France; my lunch mate Mustapha Sakhiri, Mutasfa has helped me in many ways during my stay in Linköping; Nury Simfors, Magnus Bång, Per Ola Kristensson, Erik Berglund, Magnus Ingmarsson, Anders Larsson, Ola Leifler, Lise-Lott Andersson, Nils Dahlbäck, Rego Granlund, Pär Svensson, Susanna Nilsson, Mattias Arvola, Åsa Hedenskog, Yu-Hsing Huang, Jonas Lundberg, Björn Johansson, Lars Ahrenberg, Maria Holmqvist, Sonia Sangari, Lars Degerstedt, Stefan Holmlid, Per Sökjer, Mikael Kindborg, and Rita Kovordányi.

I want to thank Abubaker Mustafa, for his encouragement and support during my stay in Linköping. Thanks to Per Lindskog and Ghazwan al-Haji for their comments during GLOBESAFE implementation.

In Linköping I met many friends that gave me a social life; special thanks go to Ibrahim-Bedri Abdelkareim, Rihab Awad Elkarim, Adnane Bouchaib and his family, Sadig Elamin (who introduced me to LiU and IDA), Isam Salih, Amr Kambal, Muhand Gaffer, Muez Dawi, Elfatih Mustafa, Elamin Dabo, Elbashir Elhassan, and Fayez Abuzaid.
My colleagues from Sudan who created a nice social environment are warmly acknowledged: Wagea Allam for his company and his generous advice (Wagea, you are the best!); Nour-Eldin Elshaiekh and Sarra for traveling all the way from Malaysia to Sweden to offer very kind support during the last two months of my Ph.D. study; Abdelhaleem Ahmed, Ibrahim A.Gader, Amin Gasem, Kalid M. Khier, Musab Osman, Mohammed Tahier. Mohammed Elmahgoub, Mutaz Hamed, Mohammd Kheier A., Naser Adil, Mohammed Elmontaser H., Nadir A. Farah, Mohhamed A. Elamin, Ali M. Eltom, and Waleed Sulatn.
The cover of this dissertation has been designed by Mohamed-Elbashier M. Belo. You have been so patient to reflect my ideas and thoughts on the cover, many thanks for everything you did so far.

Last but not least, I'm deeply grateful to my family. I want to thank my parents, brothers and sisters, without whose patience and support I would never have been able to finish this. I'm sure the 12th of June, 2008 will be a special day in Portsudan, my lovely home city!

I especially wish to express my love to Rania for being around and understanding that I sometimes had to leave her either to travel or to work on the weekend.

Imad-Eldin Ali Abugessaisa
Linköping, June 2008

List of papers included in this thesis

This thesis contains the following papers, which will be referred to in the text by their numerals. (e.g. Paper I, Paper II, etc.)

I. **Abugessaisa**, **I.** & Sivertun, Å. (2004). Ontological Approach to Modeling Information Systems. In *Proceedings of the Fourth International Conference on Computer and information Technology (Cit'04),* 14 – 16 September 2004 (pp. 1122-1127). Wuhan, China: IEEE Computer Society, Washington, DC.

II. **Abugessaisa, I.** & Sivertun, Å. (2005). Benchmarking Road Safety Situations Using OGC Model of Portrayal Workflow. In *Proceedings of the 13th International Conference on Geoinformatics (GeoInformatics'5)*, 17-19 August 2005 (CD-Rom). Toronto, Canada: Ryerson University.

III. **Abugessaisa, I.**, Sivertun, Å., & Le Duc, M. (2006). Map as Interface for Shared Information: A Study of Design Principles and User Interaction Satisfaction. In *Proceedings of the International Association for Development of the Information Society 'IAD'IS' International Conference WWW/Internet*, 5-8 October 2006 (pp. 377-384). Murcia, Spain: University of Murcia.

IV. **Abugessaisa, I.**, Sivertun, Å., & Le Duc, M. (2007). GLOBESAFE: A Platform for Information-Sharing Among Road Safety Organizations. In *Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries*, 28-30 May 2007 (CD-Rom). São Paulo, Brazil: IFIP-W.G.

V. **Abugessaisa, I.,** Sivertun, Å., & Le Duc, M. (2007). A Systemic View on Swedish Traffic Accident Data Acquisition System. In *Proceedings of the 14th International Conference on Road Safety on Four Continents (RS4C)*, 14-16 November 2007 (CD-Rom). Bangkok, Thailand: Vti.

VI. **Abugessaisa, I.** (2008). Knowledge Discovery in Road Accidents Database Integration of Visual and Automatic Data Mining Methods. *The International Journal of Public Information Systems*, Vol. (2008: 1), pp 59- 85.

List of abbreviations

ADB	Asian Development Bank
AIS	Androgen Insensitivity Syndrome
APRAD	Asia Pacific Road Accident Database
SRA	Swedish Road Authority
ASEAN	Association of Southeast Asian Nations
ASNet	The ASEAN Region Traffic Safety Network
ASP	Active Server Page
BASt	German Federal Highway Research Institute
CARE	Community Road Accident Database
CGI	Common Gateway Interface
CM	Conceptual Model
CM	Concept Map
DM	Data Mining
DT	Decision Tree
ECMT	European Conference of Ministers of Transport
EDA	Exploratory Data Analysis
EU	Emergency Unit
ETSC	European Transport Safety Council
FARS	National Highway Traffic Safety Administration
FGD	Focused Group Discussion
GIS	Geographical Information System
GML	Geography Markup Language
GPS	Global Positioning System
GRSP	Global Road Safety Partnership
HAC	Hierarchical Agglomerative Clustering
HM	High Motorized
IA	Information Architecture
ICD	International Classification of Diseases
ICT	Information and Communication Technology
I *InfoVis*	Information Visualization
IRTAD	International Road Traffic and Accident Database
KDD	Knowledge Discovery in Database
LM	Low Motorized
MAAP	Microcomputer Accident Analysis Package
MM	Middle Motorized
NTF	The National Society for Road Safety
OECD	Organization for Economic Co-operation and Development
OGC	Open Geospatial Consortium
PDA	Personal Digital Assistance
PU	Percived Usefulness
QUIS	Questionnaire for User Interaction Satisfaction
Radvis	Radial Visualization
RBAC	Role Controlled Access Method
RSMS	Road Safety Management System
RSP	Road Safety Profile

SOM	Self Organized Map
STRADA	Swedish Traffic Accident Data Acquisition
SVG	*Scalable Vector Graphics*
TAM	Technology Acceptance Model
TRL	Transport Research Laboratory
UN-ECE	United Nations Economic Commission for Europe
UoD	Universe of Discourse
VDM	Visual Data Mining
VH	Virtual Hierarchy
Vti	Swedish Road and Transport Institute
VV	Swedish Road Authority
W3C	The World Wide Web Consortium

Contents

1. Introduction
1.1. Research motivation... ………………………………………….. 9
1.2. Research issues ……………………………….……………. 10
1.3. Research design and process ………………………………. 12
1.3. Related work …………………………………….……………... 14
1.4. Contributions …………………….…………………………... 16
1.5. Thesis organization ……………………………………………. 17
2. Background
2.1. Road safety concepts and information system……………….. 18
2.2. The Five E's of road safety improvement……………………. 18
2.3. Framework for road safety performance indicators………….. 20
2.4. Information sharing……………………………………………. 22
2.5. Road accidents data acquisition system………………………. 25
2.6. AsNet system…………………………………………………... 27
2.7. STRADA: Swedish Traffic Accident Data Acquisition………... 30
3. Data mining and system thinking
3.1. Data mining: knowledge discovery in databases……………... 34
3.2. Visual data mining……………………………………………... 36
3.3. Visualization techniques………………………………………. 37
3.4. Knowledge flow in data mining……………………………….. 41
3.5. Tools for data mining: Tangara and Orange…………………... 41
3.6. Thinking and acting in terms of system………………………. 41
4. Internet GIS
4.1. Web mapping and Internet protocols…………………………... 45
4.2. Basic components of Internet GIS……………………………. 46
4.3. Map publishing methods over the web……………………….. 47
4.4. Portrayal model for interactive maps……………….…………. 48
5. Research methods and approach
5.1. Selection of research methods…………………………………. 50
5.2. Requirements identification and prototyping……….…………. 50
5.3. Development of common vocabulary……………………. 52
5.4. Study of STRADA at Swedish Road Administration (SRA)….. 54
5.5. Visual data mining (VDM) and clustering methods……….…. 55
5.6. Information Architecture (IA) for the map design……………... 59
5.7. Evaluation of GLOBESAFE…………………………………... 62
6. Results
6.1. The conceptual model………………………………….………. 66
6.2. GLOBESAFE…………………………………………………. 71
6.3. Road Safety Profile: a diagnostic tool for road safety analysis 79
6.4. Web mapping application: Map as an interface for shared information……………………………………………………… 82

6.5. System evaluation: User interaction study…………………….. 83
6.6. Results of the STRADA study……………………………………. 86
6.7. Application of visual data mining……………………………….. 90
7. Discussion
7.1. Experience from research methods…………………………… 96
7.2. Requirements elicitation, the conceptual model, and prototyping………………………………………………………... 97
7.3. Implementation aspects………………………………………. 98
7.4. Quality of the shared information……………………………….. 100
7.5. Increase trust among organizations……………………………….. 101
7.6. System thinking approach…………………………………… 102
7.7. Implications of Visual data mining……………………………… 102
8. General conclusions and future work
8.1. Conclusions…………………………………………………… 104
8.2. Evaluation of GLOBESAFE………………………………….... 105
8.3. Information quality in the scene………………………………… 105
8.4. Geospatial Data Mining of road accidents database………… 105
List of references

List of papers

Paper I
Ontological Approach to Modeling Information Systems
Paper II
Benchmarking Road Safety Situations Using OGC Model of Portrayal Workflow
Paper III
Map as Interface for Shared Information: A Study of Design Principles and User Interaction Satisfaction
Paper IV
GLOBESAFE: A Platform for Information-Sharing Among Road Safety Organizations
Paper V
A Systemic View on Swedish Traffic Road Accident Data Acquisition System
Paper VI
Knowledge Discovery in Road Accidents Database Integration of Visual And Automatic Data Mining Methods

1. Introduction

This chapter introduces the research and places it in context. This research focuses on the field of traffic safety. The main objective is to develop models and methods that make traffic and accident data accessible and easily shared by a wide range of experts and researchers concerned with road safety management and safety promotion. The research was part of the ASEAN Region Traffic Safety Network (ASNet). ASNet was initiated in 2003 to strengthen regional cooperation in traffic safety between the ASEAN countries, through the distribution sharing of information, experience and so-called best practice solutions. In this chapter, the basic motivation behind this research is introduced and related work is discussed. Thereafter, the contributions of this research to the field of traffic safty management are commented. Lastly, the organization of the thesis is given.

1.1. Research motivation

Road accidents are responsible for a considerable waste of scarce financial and human resources that are needed for the development of countries. In the case of developing countries, motorization and urbanization are growing faster than traffic legislation, institutions and infrastructure, which are needed to solve road safety problems (Trawen et al., 2002). While fatalities are declining in the developed world, they are still increasing in many developing countries (Murray and Lopez, 1996). A prerequisite for improving road safety is information about accidents, fatalities, injures, and roads, to help assess the current situation and also give a good indication of its severity. Many countries have definition problems, no data collection process or simply low-quality data availability, all of which are important for auditing road safety and supporting international comparisons. In many countries there is under-reporting of road accidents, deaths and injuries. On the other hand, under-reporting of injuries is known to be even worse than the under-reporting of fatalities (Jacobs et al., 2000 and Aptel et al., 1999).

Based on a report of the International Road Traffic and Accident Databases (IRTAD)[1], earlier studies have estimated that approximately 50% of road injuries are reported worldwide. To improve reporting, there is thus a need to standardize accident registrations and definitions of accident data, as well as a need for methods to obtain fully reliable accident data. In the following, the motive for this research is summarized:

- Death and injuries due to road accidents are a growing public health issue, disproportionately affecting vulnerable groups of road users. This has economic impact and represents a threat against society. Road traffic accidents have been ranked as the third leading cause of death in the world (Peden et al., 2004).

[1] http://cemt.org/IRTAD/IRTADPublic/index.htm

- Estimates of the annual number of road deaths vary, as a result of the limitations of injury data collections and analysis (Zheng, 2007).
- In order to gain a better understanding of safety problems and challenges, relevant information of high quality is needed. This is not possible without supporting methods and technologies. This also increases the need for research in the area, which is limited by current practices and methods, as described in section 1.4.
- Road safety is of prime concern to many individuals, groups and organizations, all of whom may require data and evidence about accidents (Moon, 2003).

1.2. Research issues

Many groups are interested in road safety data, and they tend to have different needs and reasons for requiring such data. There are also practical limitations on the amount of data that can be collected (Maher, 1991), and information sharing methods are needed to overcome these problems (Chung et al., 2004). Two important characteristics related to road safety data is that their sources vary and they all suffer from under-reporting problems (Aeron-Thomas, 2000).

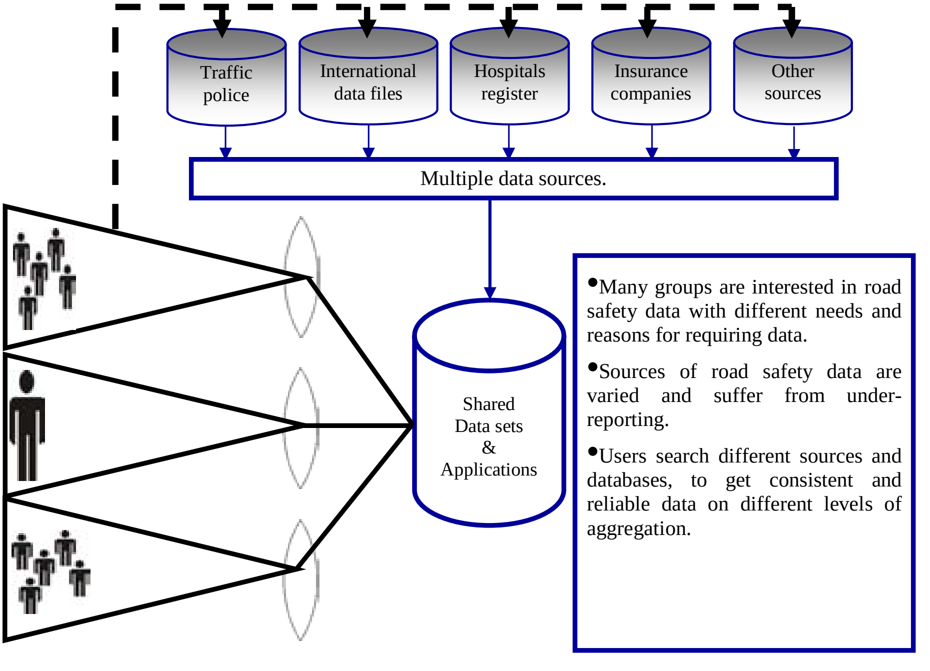


Figure 1. Multiple sources and multiple views of road safety data.

The users of these data should search for different sources and databases to get consistent and reliable information with the required level of aggregation. This reflects the problem of lack of consistency and integrity in the available road safety databases. Figure 1 summarizes the main issues of these aspects.

The above-mentioned problems imply that there is an underlying methodological problem with the collection, analysis and dissemination of road safety information. This could be made more efficient, and consequently lead to better safety practices for specific regions or countries.

Searching through different databases and information sources can be time consuming. In addition, verification of different data stored in different formats and media also complicates the task. All of these factors can have serious effects on the analysis of road safety situations, as well as on the result of the analysis. Additionally, well-defined methods in the field of information sharing, information architecture and web technologies are needed for road safety experts and practitioners. To address these issues, high quality data sets should be defined and agreed upon. These data sets can be regarded as a common vocabulary for road safety information sharing and as tools that can be used by researchers and professionals dealing with road safety to perform international comparisons. The comparison of road safety situations helps to assess the execution of national road safety plans, and it is necessary to compare these situations within an international context (Persaud and Lyon, 2007). This work aims to provide data sets (common vocabulary) and tools that will offer a framework for international evaluations of road safety situations. The requirements of road safety data sets were specified by IRTAD (Reichwein, 2006) as:

- Up-to-date information accessible worldwide
- Detailed and comprehensive data for international comparability
- Consistent time series and computer-assisted updating and processing of data

General aims

As mentioned above, road accidents are a threat to society. The main objective of research in road safety is to improve road safety in a specific region or country. Improving road safety and reducing fatalities and injuries due to accidents requires appropriate measures for particular problems that exist in particular countries or regions. By using reliable data, the magnitude and nature of the different problems related to road safety can be identified.

The overall aim of the research is to build and maintain an information-sharing platform and to construct mechanisms and analytical tools that can support road safety professionals and researchers in their efforts to eventually prevent road accidents.

Specific objectives

In particular, the research aims to explore methods in requirements engineering to determine requirements that will be implemented on the information-sharing platform. A further objective is to investigate information-sharing modes in road safety organizations and to determine how each of them can respond to the specified needs of the different users.

1.3. Research design and process

The ideal way to start building an information system is to determine the requirements of the system to be developed (Sommerville and Sawyer, 1997). For this purpose, the research started with requirements elicitation from road safety experts (Wieringa, 1996) working as researchers at Linköping University in Sweden, Global Road Safety Partnership in Geneva (GRSP), and the Federal Highway Research Institute in Germany (BASt).

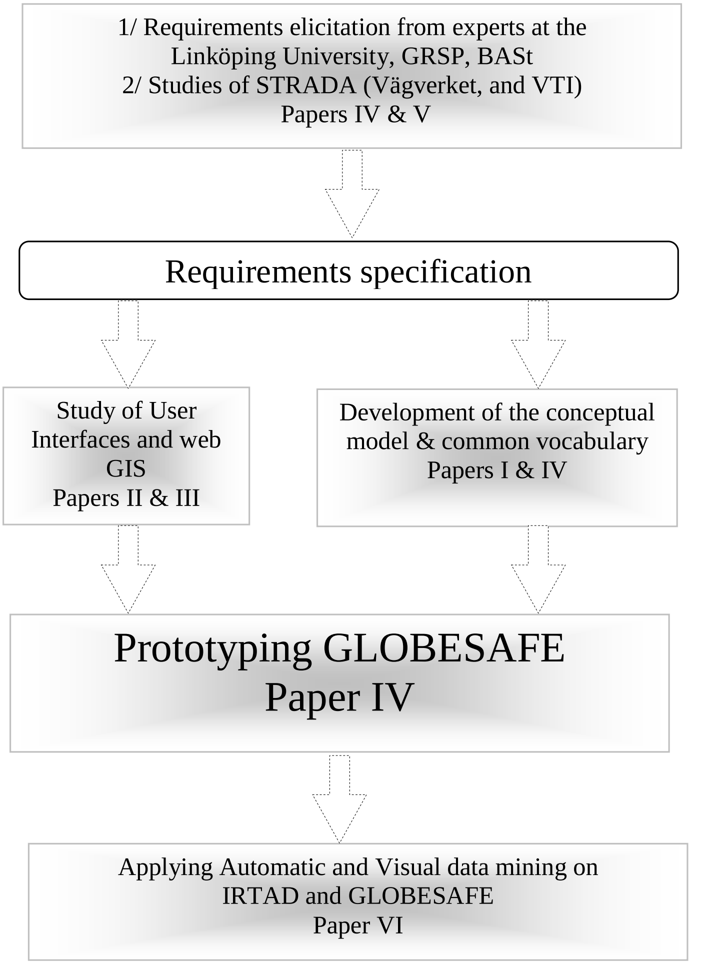


Figure 2. Overall research design.

This stage of research and the rest of the process are shown in Figure (2). Also included in the process flow are the different papers, which are concerned with the various parts of work in focus for the research carried out.

The main purpose of the study was to gain knowledge about how road safety information is collected, stored, manipulated and disseminated among these organizations. This was accomplished by face-to-face meetings, interviews and email correspondence. The results were a set of requirement specifications that helped in developing a conceptual model, built on a required ontology for the domain and explored the appropriate user interface for sharing the information (Papers I, II, III, and IV). All of these aspects are implemented in GLOBESAFE, a platform for information sharing among road safety organizations (www.globesafe.org). The development of GLOBESAFE was accomplished by iterative prototyping of the requirements specifications, then trialing the system with users and thus getting feedback for the next cycle of building a new version of the system. The prototyping process helped to develop a more mature system that could satisfy all users' needs (Livari and Karjalainen, 1989).

During the meetings and interviews with the road safety experts, it was discovered that road safety organizations collect and analyze country traffic safety and socio-economic data from multiple databases. Furthermore, they also look for important indicators when analyzing the traffic safety situation in a country. Such indicators describe traffic risks in terms of fatalities per vehicle; motorization can be measured as vehicles per 1,000 people, and personal risk as fatalities per person. In addition to this, all experts are using methods that can be used to calculate and analyze, so-called performance indicators, that can be seen as diagnostic tools when comparing the traffic safety situations in different countries. These procedures guided the research and the development of a three-tiered conceptual model and an ontology that supported the development of the required database schema. The research also explored different user interfaces for the presentation of the information.

Accident recording systems are used to collect accident information; such systems are operated by traffic police in most countries. To study such systems, a system thinking approach was applied (Lawson, 2006) to the Swedish Traffic Accidents Data Acquisition (STRADA). The study was conducted from two perspectives, the Swedish Road Authority 'SRA' (Vägverket, VV), which is responsible for operating and maintaining STRADA, and the Swedish Road and Transport Research Institute (VTI), one of the main users of STRADA. To apply the system approach, interviews were conducted with focus groups and different stakeholders in the system. To investigate the issues related to the acceptance of the system a Technology Acceptance Model (TAM), a model used to understand user's behavior towards a new innovation, was used (Davis, 1986). TAM is based on two factors: perceived utilities and perceived ease of use (Paper V).

In Paper VI, a novel approach was introduced to discover potential knowledge hidden in road safety databases due to the accumulation of the data. This accumulation limits the exploration tasks and decreases the utilization of the stored databases. In order to help solve these problems Automatic and Visual Data Mining (VDM) methods were explored (Keim, 2002). The main purpose was to study VDM methods and their application to knowledge discovery in road accident databases.

1.4. Related work

Much research has been done and continues to be done to enhance and maintain the sharing of road safety information among countries and organizations. This section discusses the contribution of these research activities and how they support decision-makers and planners with the updated information and knowledge required for evaluation, benchmarking, and development of road safety plans.

The discussion focuses on both accident recording systems that play basic roles at the lowest level of accident statistics, and international and regional initiatives for road accident information sharing.

- **Road accident data systems**

The most recognized international initiative within road accident data systems is the Transport Research Laboratory's (TRL) accident analysis package. Microcomputer Accidents Analysis Package (MAAP) (Transport Research Laboratory, 1994) is a system that basically helps accident investigators to store and analyze accident data. MAAP software consists of two basic parts. The first allows accident investigators to record accident data. One of the main advantages of MAAP is that it facilitates data collection and recording by means of a user-friendly interface and data validation procedures. The second part of the package includes the accident analysis data application, with which different types of analyses can be made. The analysis engine of MAAP provides extensive reporting facilities in different formats that lead to better understanding of the situation and of the causes of accident. The analysis engine uses both graphical and tabular data presentation methods as well as presentation of location information on maps.

- **International road accident databases**

In the development of accident databases on national levels, coordinated efforts were dedicated to making use of the available technologies that could help to develop and implement road safety programs and plans on both local and national levels. The main objectives of these databases are to represent road accident data in compatible and homogenous formats and to reduce the efforts spent by end users in searching for relevant information in the different databases. International and regional road safety organizations annually publish reports and statistics, according to predefined user requirements and agreed variables and indicators (Heinrich and Mikulik, 2005).

  - **Community Road Accident Database 'CARE'**

An example of a still-existing national database is "CARE"[2], developed by the European Commission on Transport. CARE is a centralized database hosted by the EU data centre in Brussels. The purpose of CARE is to provide EU member states with access to the

[2] http://europa.eu.int/comm/transport/care/

central accident databases across Europe. Thus, CARE can be regarded as a tool that can identify and quantify road safety problems in the different member countries. The objectives of CARE are to support the evaluation of road safety measures and to facilitate knowledge exchange between researchers and decision-makers throughout the EU.
The uniqueness of CARE, in relation to other international databases, is the high level of detailed information, collected by the member states, from individual accidents.
After the individual accident data are collected by each member country, an annual report is added to the central database by each member state. There are, however, no agreed standards by which the reports should be compiled, since each state has its own standards and definitions.
Within CARE, a framework of transformation rules, obtained from the original structure of the individual state, has been developed. This framework enables data harmonization that will increase the exchange of data and experiences between the different states. With this framework and standardization, CARE helps researchers and decision-makers to produce road safety programs that can help to reduce the number of road accidents.

- **The International Road Traffic and Accident Database 'IRTAD'**

IRTAD is developed and operated by the members of the Organization for Economic Co-operation and Development (OECD). It provides its members, and other countries, with a database of road traffic and accident information. The database is aggregated at the country level. IRTAD includes tools for decision-makers and researchers to assess the traffic situation in their country within an international context. As with CARE, it is accessible online to the members of the OECD. In addition to traffic and accident data, IRTAD provides demographic data, structured according to age groups.

Many different international and national road databases are available such as the database established by the Economic Commission for Europe (UN-ECE), an organization within the United Nations. UN-ECE annually publishes a report of important road accident statistics.

Research efforts in the area are devoted to the achievement of a schema that well represents the road accident data and other relevant accident details that respond to the users' and decision-makers' needs.

Technically, all available road databases have different ways in which the data are collected and processed (Luoma and Sivak, 2007). Generally, two methods are used to collect the data. The first method uses file transfer procedures to report annual statistics. In the second method, questionnaires are used by the organizations to collect annual accident statistics from the member states of the organizations.

Both data collection methods have limitations as they do not make use of current and different online data entry methods available via Internet protocols.

The operational specifications involve how the organizations that host and operate the databases deal with the operation issues. These specifications include definitions, determined by the users who allow use of the databases and production of publications. Important operation specifications concern the bproduced services and interaction methods.

The output services of the databases provide the users with results obtained after the databases and the statistics had been manipulated. These services can either be paper-based publications that are sent to the member states, or they can be requested via the web.

1.5. Contributions

This thesis contributes to the design and implementation of a platform for information sharing among road safety organizations to support decision-making and the development of road safety plans. Moreover, visual data mining methods for the discovery of hidden knowledge accumulated in road safety databases are implemented. An additional contribution is made by the application of the system thinking approach for the study of the road accident data acquisition systems. The main contributions of the research are:

- A set of requirement specifications, a conceptual model and a frame-ontology were implemented in GLOBESAFE. As mentioned, GLOBESAFE is a web-based platform, which can support road safety communities with required information and knowledge required to help in the work of preventing road accidents and improving road safety. The development of the conceptual model and the implementation of GLOBESAFE are described in Papers I and IV.

- A user-interface have been implemented using geographical maps for benchmarking road safety situations in specific regions or countries together with the application of the Open Geospatial consortium standards to maintain the interface. A usability study was conducted, which helps to compare the map as a user interface for shared information and its suitability to increase the awareness of the road safety situations (Papers II and III).

- Application of a system thinking approach and technology acceptance model to STRADA to help identify the real acceptance and use of STRADA in Sweden, and also to provide practical guidelines for the future development of a new version of the system. This will be useful both to the Swedish Road Authority (SRA, Vägverket) and the developers of the system. It is recommends that the users and developers involved during the life cycle of STRADA can use an enabling system to overcome the problems related to system usability and complexity. Also suggested is the use of an iterative development technique to govern the life cycle of the system (Paper V).

- An approach to discover the knowledge hidden in the road accident databases by combining Automatic and Visual Data Mining (VDM) methods has been determined. This helps to involve the users further in the exploration process. This approach has been applied to two different data sets. The first data set comes from GLOBESAFE and concerns the ten ASEAN countries, and the second concerns the twenty-five OECD countries. The latter data set is presented by special agreement from IRTAD (Paper VI).

1.6. Thesis organization

This thesis is divided into two parts. The first part, composed of eight chapters, introduces the reader to the research problems and the main issues that have been identified and discussed through the research period. In Chapter 1 the research is discussed and linked to the publication process and the methods for overcoming the problems this presented. Related work in the field is discussed and compared to the approach presented here.
Chapters 2 to 4 introduce the basic theoretical concepts in the area of the research and present the theoretical framework of this research. Chapter 5 covers the research methods and the approach taken, and justifies the use and validity of these. Chapter 6 presents the main findings and results of the different studies. Chapter 7 discusses the research findings and results, comparing them to other current research in the area. The last chapter of the thesis, Chapter 8, gives general conclusions and some suggestions for future work.

The second part of the thesis is a collection of the six papers published during the research.

2. Background

The present chapter aims to present the basic concepts in traffic safety and the role of road safety systems. Different components of road safety systems and the “Five E's” of safety management are presented. This chapter also describes the framework for performance indicators as countermeasures for safety situations. Information-sharing modes in general are discussed, with a focus on current methods and research into information sharing in road safety organizations. Finally, the framework for ASNet system and STRADA as an example of a road accident recording system are introduced.

2.1. Road safety concepts and information system

Different types of road accidents have different definitions. A commonly accepted definition is given by the Transport Research Laboratory (Baguley, 2001) as “*a rare, random, multi-factor event which is always preceded by a situation in which one or more road users have failed to cope with their environment*”. It is clear from the definition that road accidents are events or a series of events that rarely happen, in terms of the passage of time and the number of traffic movements at a particular location in the road grid (WHO technical report, 1998).

Road accidents are characterized as random events (David and Branche, 2004), which means that they are impossible to predict (i.e. they are unpredictable events).

Many factors contribute to accidents, such as weather (e.g. rain and/or darkness), behavioral factors (drunken drivers), vehicle factors and road conditions. All of these factors may lead the drivers to fail to cope with the situation and result in an accident (Peden et al., 2002).

2.2. The Five E's of road safety improvement

Improving road safety requires appropriate measures for particular problems that exist in particular countries or regions (Elvick and Vaa, 2004).

The three E's (South Walk Road Safety Plan, 2004) of road safety improvement are known as: *Education*, *Enforcement* and *Engineering*. Education can target different groups of road users and traffic policy by use of different group campaigns. Enforcement concerns the execution and enforcement of the road legislation. Through engineering work, road safety will be improved by means of traffic engineers, who play a fundamental role in the performance of these activities.

Recently *Encouragement* has been added as a fourth E (WHO Technical report, 1998), which refers to the government role in road safety improvement.

TRL added a fifth E for *Evaluation*. That is, evaluation should be seen as a pilot scheme for improvement, as this will identify the resources available for road safety. The evaluation results can be used to take proper decisions, primarily by the policy makers and the planners to:

- Prevent accidents
- Reduce the causes of accidents

A general framework for the two processes implies that accident data systems should be at the heart of the improvement (IRTAD special report, 1998). Figure 3 shows the framework for the road safety improvement by using the five E's.

<u>Road Safety strategy</u>

A. Accident prevention.
B. Accidents reduction.

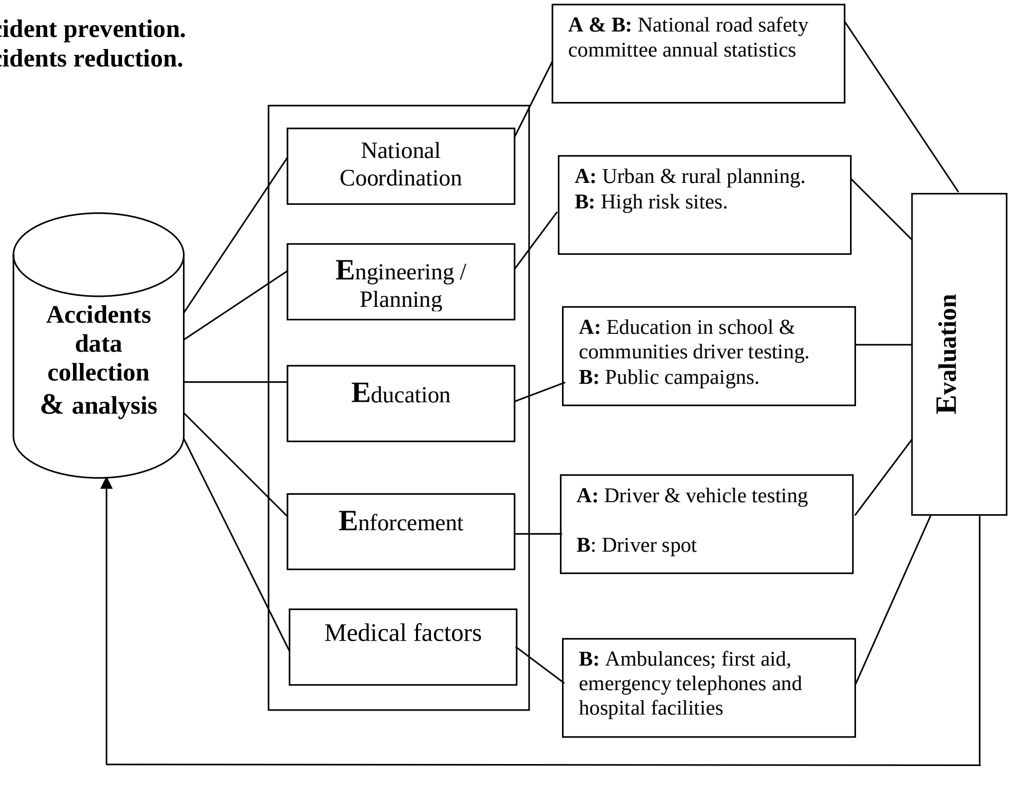


Figure 3. Framework for the road safety improvement using the five E's (Baguley, 2001).

At the heart of the framework for road safety improvement, information flow should support the implementation of the plans and work flow processes. The data to support this should be collected from different sources and should be available to, and shared by, all concerned users.

- **Requirements for road safety improvement**

When policy makers and planners develop road safety strategies and programs, they should acquire an understanding of the problems while considering all dimensions and causes of accidents (Ghee et al., 1997).
By using reliable data, the magnitude and nature of the different problems related to road safety can be identified. The data can, in this way, help to identify different road user groups. Moreover, the risk factors should also be identified, for example, road design, driver education and training, and vehicle conditions (SWOV research activities report, 2001).

- **Information on accidents**

The shortcomings of the available accident data are a major constraint on actions to ensure safety (Koster and Langen, 2000), because the decision-makers can be misled in underestimating the problems, through the analysis processes in which accident causes are identified. This should be taken into account in the design of countermeasures and the initiation of new approaches to accident prevention.
Data on road accidents are generally collected at a local level. In most countries, detailed investigations are carried out, often by the police, at the scene of the accidents (Wegman, 2001). The collected data will serve statistical, legal or research purposes. So, even when statistics on accidents are not collated at the central level, it is often possible to gain access to detailed data for in-depth analyses. Whether all or only a small sample of accidents are described in this way, invaluable information on the factors that generate them or contribute to their severity can be obtained and, to some extent, general conclusions can be drawn from them.

2.3. Framework for road safety performance indicators

Road safety indicators are used to provide the policy makers with a means of measuring the effectiveness of the safety programs and the utilization of public resources. Generally, road safety performance indicators are required (European Transport Safety Council, 2003) to monitor the progress and results of the road safety programs in each country.

- **Performance indicators**

The European Transport Safety Council (ETSC) (ETSC, 2001) defines road safety performance indicators as “*any measurement that is causally related to crashes or injuries, used in addition to a count of crashes or injuries, in order to indicate safety performance or understand the process that leads to accidents*” (Figure 4). Safety performance is defined by ETSC as “*the changes over time in the level of transport* *safety, with a reduction in the number of accidents or the number of killed or injured people defined as an improvement in safety performance*".

Road safety indicators are needed by the people involved in road safety planning, because accident and injury counting cannot be a perfect measure to determine the level of safety. For this reason, accident and injury counting will not be useful unless it is assessed in terms of their social cost. Different road safety performance indicators were proposed during the requirements elicitation process, which were later implemented in GLOBESAFE. The reasons behind the use of these indicators are summarized as follows (Brouwer, 1997):

- Both the number of accidents and injuries on roads are subject to random fluctuations, i.e. short-term changes in the number of accidents and injuries will not change the underlying, long-term expected number
- Under-reporting of accident and injury statistics in international databases

Counting the crashes will not help to understand the causes of the crashes. The road safety community needs effective measures to understand causes of accidents.

- **The wider context of road safety indicators**

The theoretical frameworks in which road safety indicators are used are the safety programs within certain regions or at a national level. In this section, performance indicators are discussed, as they can be used to serve and advise policy makers and road safety program developers in identifying weak and strong sectors in safety (see Figure 4).

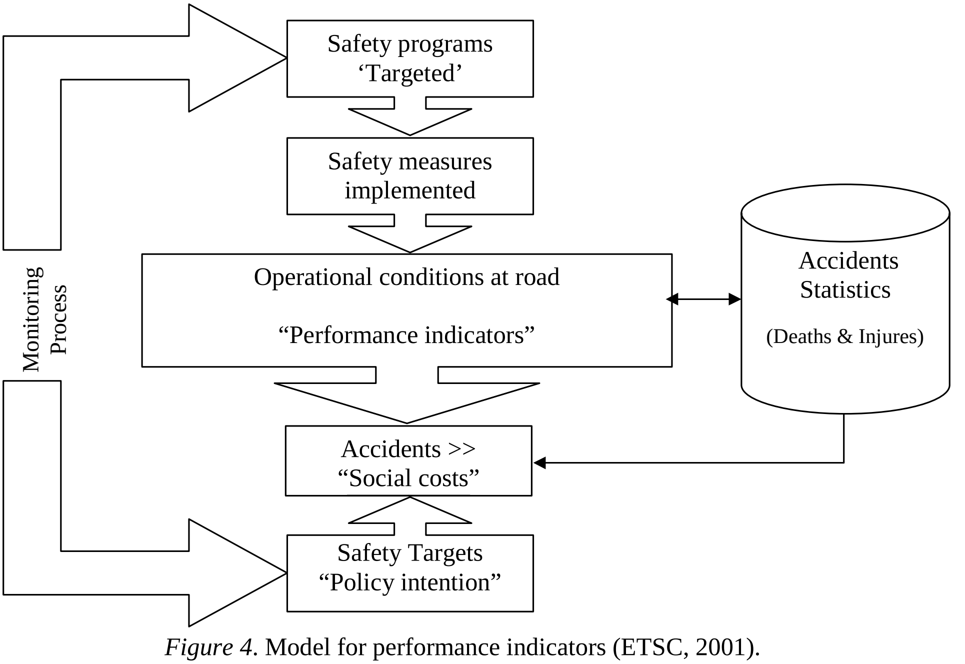


Figure 4. Model for performance indicators (ETSC, 2001).

The performance indicators model (ETSC, 2001) is composed of different processes and information flows; the basic processes take place at the top of the model, where the safety programs are identified and set by safety policy and decision-makers. Identification of safety indicators is the major process, which results in certain measurements, rules and regulations. The operation part of the model is linked to these measurements in the safety program, which also can be influenced by various factors, including:

- Environmental and social factors
- Technical conditions of the roads
- Vehicles and other means of road transportation

Operational conditions result in statistical data relating to accident rates and casualties. The consequences of the operational conditions may lead to accidents with high costs to society.

The backward cycle of this model takes place between the bottom and the top of the model, in which the safety targets are compared against the rate of accidents and the social costs.

- **Road safety indicators as elements of safety management systems**

The performance indicators are required during the safety auditing process and can be used as means to characterize the safety quality of the road components. To establish road safety performance indicator systems, causal relationships should be identified and the relationships should be converted to quantitative indicators as prerequisite requirements for these conditions. Thus all road safety problems should be transformed into specific indicators.

Road safety performance indicators should be compiled using the following conditions given by ETSC:

- Each indicator should be significant, available, reliable and easy to collect
- Any quantity or number (absolute figure) must refer to a full calendar year
- Within a given country, the indicators must be comparable from year to year, so that progress can be monitored separately from other countries.

2.4. Information sharing

In this section, information-sharing concepts and the need for sharing road safety information at both horizontal and vertical levels is discussed. Road safety information has special features that make it conducive to sharing (Mitchell, 2002). This is, of course, due to the fact that different agencies and organizations (traffic police, hospitals, etc.) are collecting road accident and injury statistics and many other groups are requesting this information to carry out their tasks (decision-makers, insurance companies, road engineers, etc.). Also, road safety information can be reused in different ways by these groups and others, such as researchers and experts.

Sharing road safety information across different organizations or between the above-mentioned groups is constrained by two factors. The first is organizational behavior and the second comprises technical constraints and barriers (Yannis et al., 1998).

- **Concept of sharing**

The definition of the concept of sharing information is based on the associations made by Roget's Thesaurus (Roget and P. Mar, 1987) with the words and certain types of human activities and with entities such as information and physical resources.
The types of activities in these associations are common, cooperative, and participatory activities. The entities that could be merged, composite, and coincident, in partnership are possible to be shared.
The above association could be extended to *information sharing*. Figure 5 shows the processes and activities associated with information sharing.

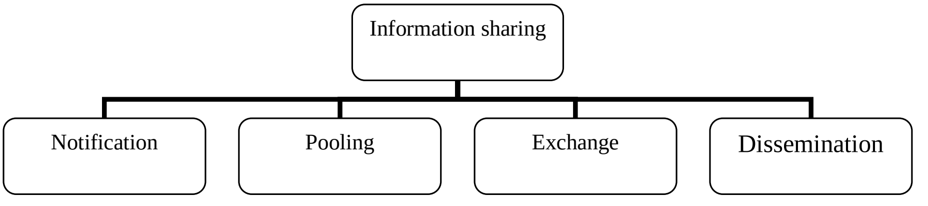


Figure 5. Information sharing and associated activities.

- **Modes of information sharing**

The modes of information sharing that could be performed in different situations can be found in one or more of the activities shown in Figure (5).
The first two modes, proposed by Carter (1992), are where one organization can sell its information to another organization. Here, information is regarded as a product. Another mode of information sharing is the partnership mode, in which multipurpose information is shared with other partnership members.
Two non-commercial modes of information sharing are proposed by Tosta (1992), who emphasized organizational relationships and information flow. The first mode is the vertical relationship that exists mostly within a single country, in which lower administrative levels are reporting their information to higher levels. The second mode is the horizontal relationship in which different organizations are cooperating together to achieve their tasks (an example of this mode is the Traffic Incident Management system of the USA, which involves different public and private parties sharing information).

- **Methods of information sharing in road safety**

Road safety information is often shared on different aggregation levels. Disaggregated information contains detailed information for every accident (e.g. CARE and WHO), while aggregated information contains statistical summaries of accidents without providing details (e.g. IRTAD, ECMT). In both cases, the methods of information sharing will affect the data quality, which also has an influence on the utilization of the information and also on the users. Available methods can be categorized according to the way in which the users are dealing with the sources of information (the information providers).

Current methods use the World Wide Web for online data sharing while protocols such as FTP are sometimes used to transfer both aggregated and disaggregated information. The users have, in this case, the ability to perform online queries and this method enables them to select information according to their requests. Examples of online queries are retrieval of information from countries with equivalent motorization level or countries from the same region.

Different countries use storage media such as CD-ROMs to report road accidents, both in horizontal and vertical levels. With these methods, data compatibility and heterogeneity are required.

Annual publications are commonly accepted methods for information sharing, in which the responsible body (e.g. IRTAD or WHO) will produce annual reports that contain analyses as well as raw data for each member; in this case, data confidentiality issues are considered and some sensitive information is removed.
Personal contacts via email or face-to-face communications could be used to share non-statistical information such as experiences in different countries with similar situations and road safety problems.

- **Facilities for information sharing**

The methods discussed in the previous sections need certain facilities and tools to make the information interchange between the different parties easier and more likely. These facilities can be permanent, if the information is shared in large volumes or in higher frequencies than *ad hoc* interchanges. In this thesis, a platform was built that supported sharing of information between organizations and individuals (Papers III & IV).

2.5. Road accidents data acquisition system

The basic element needed for coordination, maintenance and auditing of road safety information is the accident data system (ADB, 2000); see Figure 6.

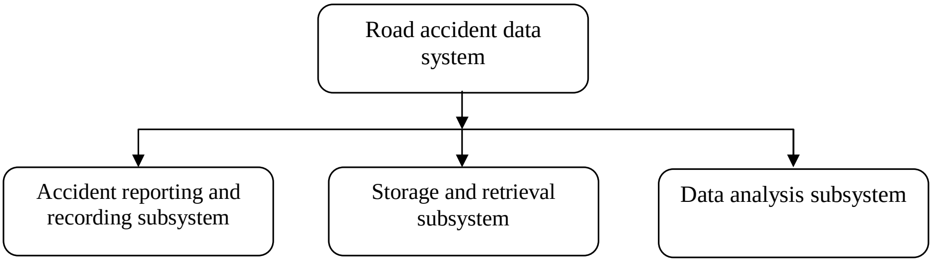


Figure 6. Components of the road accidents data system.

The road safety community makes use of the system output and requires input to perform their tasks. For that purpose, different requirements should be supported by the road accident data system.

Subsystems in Figure 6 can be integrated or eliminated within the broader context of the road safety information system.

A general model of the road safety information system is required to give a framework for accident causes and trends (GRSP, 2003). Road accident reporting and recording subsystems should be able to answer question such as:

I. Where did the accident occur? (Location data)
II. When did the accident occur? (Attribute data)
III. Who was involved? (Deaths, injuries, etc.)
IV. What were the consequences of the collision? (Costs)
V. What were the environmental conditions? (Descriptive)
VI. How did the collision happen? (Descriptive)

If these questions can be answered by the accident reporting and recording subsystem, then the different actors and governmental departments can indicate a level of safety in their country.

- **Level of use of accident data system**

Road safety data is used by different actors in road safety, who have different needs and requirements. The data collected should give the user a general understanding of the problem in a way that can help to develop plans to solve or reduce the road safety problems. For user sub-groups, changes in the trends and in the progress of road safety situations are needed. Hence, the information should help capture these trends. In addition, the data system should help to (Proctor et al., 2001):

- Identify high-risk groups and the problems facing them
- Identify hazardous locations

- Enable objective planning and resource management
- Evaluate effectiveness and monitor achievements of targets
- Make international comparisons

Road data systems can be used in each country on national and on local levels. On national levels, users will try to get an understanding of the nature, characteristics and scale of the existing problems. On the other hand, at local levels, the users will tend to investigate road user groups, use data for designing safety schemes, justifying highway planning and increasing public awareness. Furthermore, the information will be used in education and training programs, while insurance companies will make use of the accident information for legal purposes and to assess the amount of lost and damaged equipment.

- **Accidents reporting and recording system**

Normally the traffic police take responsibility for registering accidents, as they will be among the first to reach the scene of an accident and to get the report from those involved. However, substantial numbers of accidents are not reported to the police. This situation is referred to **under-reporting** (Sluis, 2001).

- **Accident registration form**

Accident reporters (the police) usually use paper forms that are very condensed and contain all the necessary details. Also, the reports look very different from country to country according to the legislative procedures followed by the courts in the respective countries. Reporting forms are usually supplemented with other information, in the form of attachments such as statements from drivers, pedestrians and witnesses (ADB, 2000).
The forms normally contain coded information that makes it easy for the police to write down all the information quickly; the rest of the form will be left empty. Then at the police station another specialist will continue the work and perform two further tasks. The first is to complete the form and the second is to enter the data into the computer system. In all cases, the information should be sent to the national accident database, either as paper or via the network, where the accident records will be collated (Sluis, 2001).

The data sets collected in the accident reports should answer the following queries:

I. Where (this section should give the details of the location of the accident)? According to the supportive technology, the location can be map coordinates, road names or road segments
II. When (the time of the accidents)?
III. Who was involved in the accident (people, vehicles, or animals)?
IV. What is the result of the collision?
V. What are the environmentally related conditions (e.g. poor light, weather and/or road surface conditions)?
VI. Why or how did the collision occur (including collision type, drivers, or fault types)?

- **Accident location and location coding system**

To identify black spots on the national grid, accident locations are required. Accident location information can also be used for evaluation of road safety strategy and policy; this implies that the accident location should be coded accurately as a part of the accident report (Wegman, 2001). Location information is used by the local authority (low level) users to pinpoint the locations via the network. Different methods are used to code accident locations. The first method is a general map sketched by the traffic police officer or the investigator. The second method can be based on the use of a GIS reference system to identify the location. In the first case, the sketch is attached to the accident report. If the second method is used, it is easier to computerize the accident location and to geo-reference the information with help of geo-information capturing methods (Hanakawa, 2004).

- **Accident report's storage and retrieval system**

The second level of the accident data system concerns the storage and retrieval process (ADB, 2000). As the accident reports are filled at the scene of the accident, the contents should be stored in a system that will keep the reports as they were recorded and so it will be possible to refine them in the future for analyses and other purposes. User requirements will put certain specifications on the data structures used, the types of analyses performed and the reports generated. Different software packages are used in different countries (Hills and Elliott, 1998) to track the accident records and to apply different standard analysis techniques.

- **Accident information dissemination**

Accident information should be disseminated widely within the road safety communities. This will help to increase the awareness of road safety activities, and show the magnitude and nature of safety problems (Hills and Baguley, 1994). All actors and road safety agencies should receive annual and *ad hoc* published reports from the traffic police showing the actual situation during certain periods and at different sites.

The published reports will also help to evaluate the accident prevention plans to help local authorities to improve their plans.

The requirements here are that the accident databases should be accessible by all organizations participating to improve overall road safety.

2.6. AsNet System

The information-sharing platform will be integrated as a part of an existing Internet-based education and training system at Linköping University ASEAN Safety Network. Current Internet technologies support sharing of knowledge and experiences that can only be captured within the traditional classroom (Makedon et al., 2003). As a response to the health and economic impact of road crashes, ASEAN Development Bank (ADB) and Transport Ministers (ATM) launched a project intended to provide technical assistance to improve the traffic safety situation by means of:

- Strengthening the different institutional work with traffic safety
- Building the capacity to provide courses and training programs for human resource development

In addition, the aim of the project is to encourage regional cooperation between the member states of the ASEAN organization.
An approach to achieving these goals is through a project composed of:

- A network system supported by modern information technology to exchange knowledge and best practice between the members
- An education and training system to increase capacity

The objective of ASNet is to create a sustainable support system to increase capacity building by providing modern tools that enable communication between practitioners, knowledge sharing and best practice solutions. Furthermore, ASNet provides sustainable system and training courses for trainees, and analysis systems that can help to measure the road safety situation and carry out the safety programs. The main objective of the ASNet system is to increase road safety education for postgraduate students as well as for professionals working in the field of road safety. The general structure of the framework is presented in Figure 7.

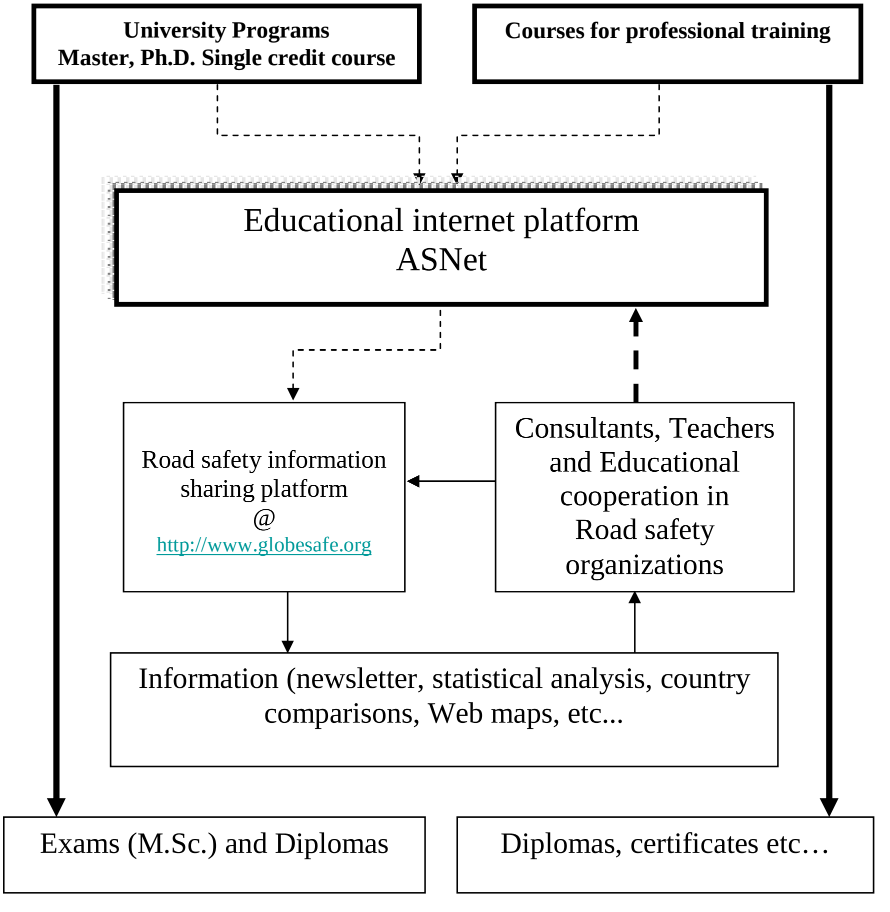


Figure 7. Structure framework of ASNet system.

2.7. STRADA: Swedish Traffic Accident Data Acquisition

STRADA was introduced ten years ago, when the counties of Stockholm and Skåne started to test the system. In December 2002, the Swedish Road Administration (SRA, 'Vägverket') accepted STRADA as an accident reporting system, to be used by the police to replace the old system. (Until the end of 2002, RSA had a system called OLY/VITS. Accident data are linked to traffic data in VDB.)
According to AerotechTelub STRADA incorporated the following features:

- Helped the coordination of accident reports between the police and emergency units (*akutmottagningen*) at the hospitals
- Maintained the exact location of the accidents on the map
- Increased/improved the statistics of road accidents with respect to reliability and consistency (by matching similar accidents together)
- Allowed data retrieval from a common database

- **Overall system structure**

STRADA mainly consists of three subsystems shown in Figure 8:

I. Hospital client
II. Police client
III. Output client

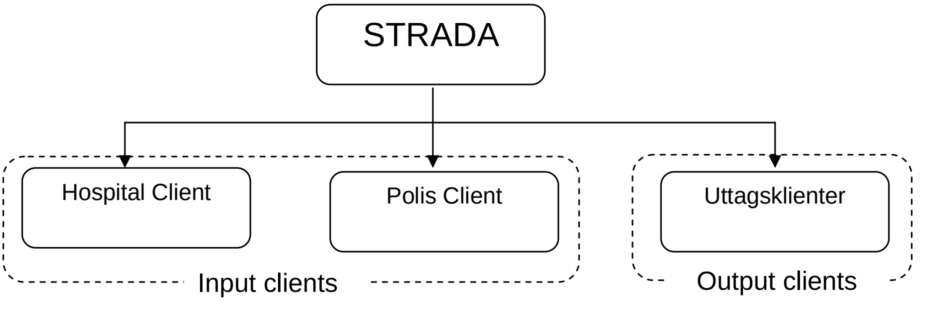


Figure 8. STRADA subsystems.

Since 2003, STRADA has operated in the seven administrative regions of SRA (Björketun, 2005). Today, not all hospitals use STRADA;some of them still refuse to use the system. SRA is encouraging hospitals to join STRADA by contributing to the operation costs of the system and training of the staff.

- **Information flow in the system**

Figure 9 shows that there are three types of organizational entities interacting with STRADA: the data collection entities (police and hospitals) and the STRADA coordinators at SRA that use STRADA clients to input reported accidents. Each of them has reporting procedures that are different from the others. The rest of the entities use STRADA to carry out their tasks. They have clients that only allow them to query the database and retrieve information. Here information such as vehicle numbers and personal information is removed.

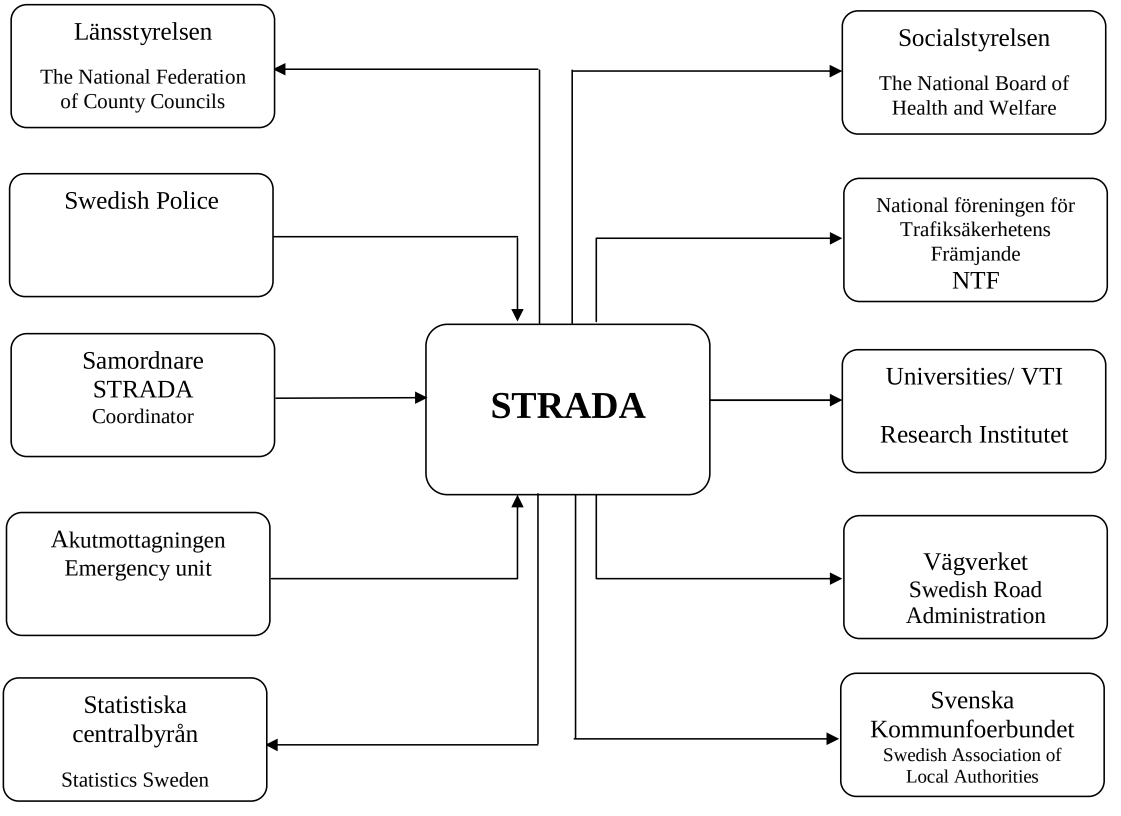


Figure 9. Information flow and user entities in STRADA.

- **Reporting procedures at the hospital clients**

When a hospital accepts STRADA, the client software is installed by the person responsible at SRA in that region and the nurses in the emergency unit are responsible and trained for data input to the client. At the hospital, two forms are used as input to STRADA. Those are:

o Traffic reports (*Trafikskadejournal*)

This is a paper form to be filed by the persons who are involved in an accident and arrive at the hospital either by ambulance or by any other transportation (see Figure 10).

- Doctor journal

This form is filled by the doctor who meets the patient and reports their health status.

These two forms are used only if the accident involves injuries. In the case that the accident has led to the death of a person, then this type of accident will not be reported at the hospital.
The nurses use both forms to complete the information required by the hospital client. Three groups of data are entered to the system:

- Basic information about the patients
- Injury descriptions
- Location information

Injures are classified according to International Classification of Diseases (ICD) and Androgen Insensitivity Syndrome (AIS) codes. By entering this information, the nurse will be able to send the report to the main STRADA server.

- **Reporting procedures at the police client**

In the case of an accident, the police will be informed and the investigators take the paper form and go to the accident site. The form that is filled in by the police is given to a STRADA-trained police officer, at the police station, who is responsible for the entering of the data to the police client.
If there is any missing information the data input officer will contact the investigator to clarify the errors. Also, the information acquired represents the basic information about the accident and the location descriptions.
Police reports from the accident sites include information about when, how, and where the accident took place, the traffic environment, the speed limit, circumstances of the accident, light and road surface conditions, passive safety systems used, and some facts about the injured persons.

- **Matching process**

Accidents that are reported in the same region by the hospital and the police are matched by the system. When similar accidents have a high matching ratio and appear at the same date, time and location, this implies that the accident was reported by the police and the hospital, and involves injuries. This, however, relies on both the hospital and the police using the system properly at the regional level. The matching procedures used by the system include the following information:

  - Circumstance (circumstance code)
  - Person information (person, personal ID number, driver, passenger)
  - Vehicle information (vehicle registration number)
  - Accident (accident number)

Unmatched accidents are treated manually during the quality control procedures by SRA.

- **Quality control procedures**

Quality control takes place twice a month. The purpose of the work is to assure the quality of the database. Routine controls are performed by the regional coordinator and the results are sent to the STRADA main server.

Trafikskadejournal – person.

Ifylls av samtliga patienter som skadats i trafiken.
OBS! Gäller även fotgängare som har ramlat och cyklister som har kört omkull.

Jag kom till akutmottagningen
Datum (år/mån/dag) ___/___/___ Klockslag ___:___

Olyckan inträffade
Datum (år/mån/dag) ___/___/___ Klockslag ___:___

ID-uppgifter

Beskrivning av olycksplats
Ort: ____________
Väg/gata: ____________
Olyckan inträffade i tättbebyggt område ☐ Ja ☐ Nej

Olyckan inträffade på
☐ Cykelöverfart ☐ Övergångsställe

Platstyp
☐ Gatusträcka ☐ Korsning ☐ Rondell
☐ Cykelbana ☐ Trottoar ☐ Hållplats
☐ Annat, t.ex.park,torg,p-plats: ____________

Beskriv utförligt hur olyckan gick till
T.ex. orsaker och händelseförlopp, i eller på väg till/från skola eller arbete, på fritiden etc.
Rita gärna en skiss över olycksplatsen.

Vid olyckan var jag
☐ Fotgängare
☐ På cykel
☐ På moped
☐ På mc
☐ I personbil
☐ I lastbil
☐ I buss
☐ Annat: ____________

Och jag var
☐ Förare
☐ Passagerare med placering (fram, bak, höger, vänster, stående) ____________

Jag kolliderade med
☐ Fotgängare
☐ Cykel
☐ Moped
☐ Mc
☐ Personbil
☐ Lastbil
☐ Buss
☐ Annat, t.ex träd, stolpe, älg: ____________

Skyddsutrustning
☐ Ingen
☐ Hjälm
☐ Bälte
☐ Barnstol/kudde
☐ Airbag utlöst
☐ MC-ställ
☐ Belysning/reflex
☐ Annat: ____________

Vägomständigheter
☐ Vägen/gång-/cykelbanan var hal pga snö/is
☐ Vägen/gång-/cykelbanan var hal pga löv
☐ Vägen/gång-/cykelbanan var hal pga vatten
☐ Vägen/gång-/cykelbanan var hal pga annat
☐ Vägen/gång-/cykelbanan var ojämn
☐ Vägen/gång-/cykelbanan hade hål och gropar
☐ Vägen/gång-/cykelbanan hade löst grus
☐ Vägen/gång-/cykelbanan har spårvägs-/järnvägsspår
☐ Annat: ____________

Jag kom in med ambulans ☐ Ja ☐ Nej

Polis har varit på olycksplatsen ☐ Ja ☐ Nej

Får kontaktas per telefon för komplettering
☐ Ja Tel nr ____/________
☐ Nej

Uppgifterna kommer att databearbetas och hanteras enligt Sekretesslagen och Personuppgiftslagen.

Figure 10. Traffic report (*Trafikskadejournal*).

3. Data mining and System thinking

3.1. Data mining: knowledge discovery in databases

Data acquisition methods and storage technology have resulted in the growth of a huge amount of data stored in different types of databases. With this advancement in database technologies, the need to extract useful information from the databases has increased (Larose, 2006). The field concerned with these tasks has become known as “data mining” (DM; Pang-Ning et al., 2006).
Data mining is the analysis of large observational data sets to find unexpected relationships and to summarize the data in novel ways that are both understandable and useful to the data owners (Hand et al., 2001). A more relevant definition from Demšar (2006) is that *data mining is the process of identifying useful and as yet undiscovered structures in a database*. The relationships and summaries derived through a DM process are models, patterns or relationships (Figure 11).

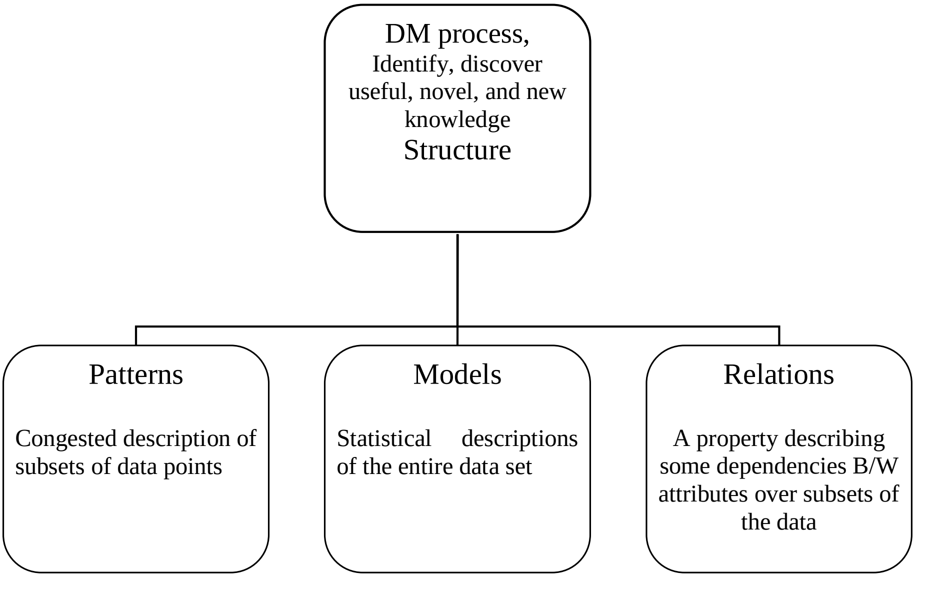


Figure 11. Mining process results in pattern, models, and relations.

The structures found within data sets must be novel, novelty must be measured relative to the user's prior knowledge, and the structure should be understandable.

- **Knowledge extraction through data mining**

Knowledge Discovery in Databases (KDD) originated in the field of Artificial Intelligence (AI), and data mining has been placed within the context of KDD. As a process, KDD involves several stages (see Figures 12 and 13) (Hirji, 2001). The process starts by selecting the target data set and transforming it into the required format by the data mining software (optional stage). The impoertnat stage when performing DM is to extract new structure in the data set, and to interpret and assess the discovered structure (Hoffman et al., 1997).

As previously defined, DM is the process of sifting through and analyzing rich sets of domain-specific data and then extracting information and knowledge in the form of new relationships, patterns, or clusters for different purposes.

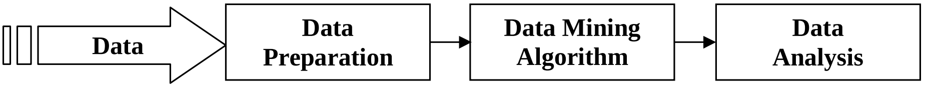


Figure 12. Overall data mining process.

- **The KDD process iterates five steps**

  - Data selection

  This step consists of choosing the goal and the tools of the data mining process, by first identifying the data to be mined and then choosing the appropriate input attributes to output the information representing the task.

  - Data transformation

  Transformation operations include the following:
  Organizing data files into desired format
  Defining new attributes
  Reducing the dimensionality of the data file
  Handling missing data

  - Data mining step *per se*

  The transformed data are subsequently mined, using one or more techniques to extract patterns of interest. The users can significantly aid the data mining process by correctly performing the subsequent steps.

  - Result interpretation and validation

  For the understanding of the meaning of synthesized knowledge and its range of validity, the data mining application tests its robustness, using established estimation methods and unseen data from the database. The extracted information is also assessed by comparing it with expertise in the application domain.

  - Incorporation of the discovered knowledge

This step consists of presenting the results to the decision-makers who may check or resolve potential conflicts within previously extracted knowledge and apply the new discovered patterns. The above-mentioned steps are illustrated in Figure 13.

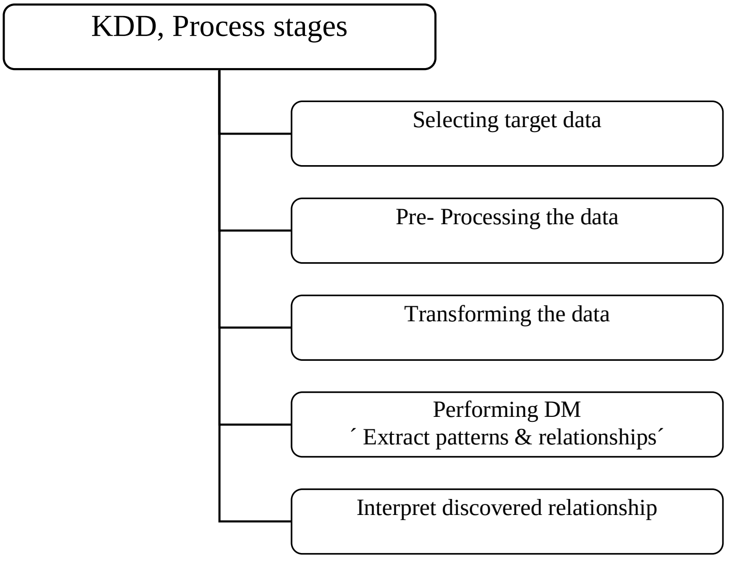


Figure 13. Steps in knowledge discovery in database.

3.2. Visual Data Mining

Conventionally, DM refers to the act of extracting patterns or models from the data. Many steps precede the data mining step: retrieving the data from large databases, selecting the appropriate subset to work with, deciding on the appropriate sampling strategy, cleaning the data and dealing with missing fields, and applying the appropriate transformations, reducing the dimensionality, and making projections. To decide whether the extracted information does represent knowledge, the information must be evaluated, which can be done by visualization of the result. *The approach in DM where visualization and visual interaction between the computer and its users is used is known as Visual Data Mining (VDM)* (Keim, 2001).

- **Classification of VDM techniques**

Information Visualization (InfoViz) focuses on the mapping of the abstract data onto the screen space (Shneiderman, 1996). Many techniques have been developed for this purpose. Some of these techniques can be used only to visualize one or two-dimensional data and give good exploration abilities, but can only be used for low dimensional data sets. The need for visualization techniques that help in visualizing multi-dimensional data sets continuously has increased lately and many tools have been developed for this purpose. These techniques can be classified according to three aspects (Table 1) (Keim, 2001):

- The data to be visualized
- The visualization technique
- Interaction and distortion technique used

Table 1

Classification of VDM techniques

Data type to be visualized	Visualization technique	Interaction and distortion techniques
One-dimensional data	Standard 2D/3D displays	Interactive Projection
Two-dimensional data	Geometrically transformed displays	Interactive Filtering
Multidimensional data	Icon-based displays	Interactive Zooming
Text and hypertext	Dense pixel displays	Interactive Distortion
Hierarchies and graphs	Stacked displays	Interactive Linking and Brushing

3.3. Visualization techniques

Knowledge discovery and mining tasks can be easily performed with graphical presentations, and this is the role of visualization. The main function of visualization according to Fayyad and Grinstein (2002) is to:

- Provide overviews that simplify complex data sets
- Summarize the data and help to identify the relationships and patterns in the data set

InfoVis principles can be applied to visualize the results obtained from data mining algorithms. Methods such as Scatterplot (Harris, 1999), linear projection, Polyviz, and survey plot provide information on the relationships between two numerical attributes and a discrete one. All are applied as geometric visualization methods where certain value(s) are visualized as points in a two-dimensional space and where the values of the attributes only influence the positions of the points and not their size, shape or color (Grinstein et al., 2001). In this research these methods were implemented. Furthermore, for visualization of massive data sets, we used graphical visualization methods such as:

- Histograms used to illustrate the probability distributions of certain numerical attribute values of a given set of objects (Figure 14).

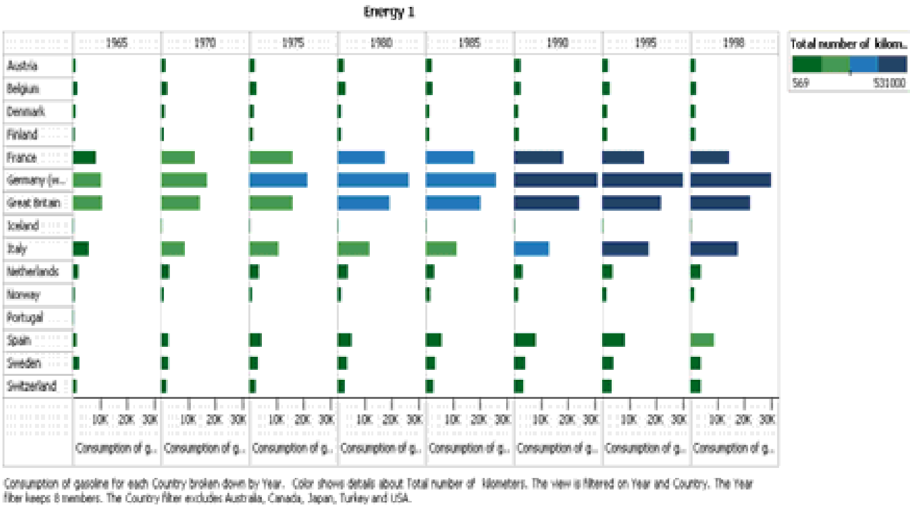


Figure 14. Histograms of a visualization method., showing consumption of gasoline for each country by year; the view is filtered on year (1965-1998) and the countries filtered exclude Australia, Canada, Japan, and USA.

- Dendrograms used to visualize and illustrate different clusters produced by automatic data mining alogorithms. In this research they are used to illustrate the clusters of K-means (Paper VI).

- Parallel coordinates (Inselberg, 1981) used as VDM tool and originally used as a new way to represent multi-dimensional information. In parallel coordinates, all axes are parallel to one another and equally spaced. Figure 15 shows results from the research experiment on the GLOBESAFE data set; see Paper VI.

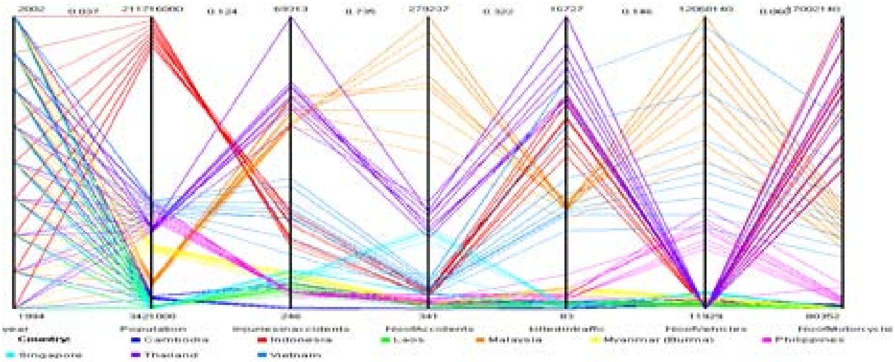


Figure 15. Parallel coordinates, all axes are parallel to one another and equally spaced.

- Radviz (Hoffman et al., 1997), or radial visualization. In radviz the data are represented as points. These points are plotted inside a circle. Figure 16 shows a radviz plot of five attributes from nine Asian countries. The data set is taken from GLOBESAFE (1994-2002).

It can be seen that the attributes (number of motorcycles, number of accidents, number killed in accidents, and country population) separate all nine countries and provide a clear interpretation of each country. As an example Indonesia (red points) has a high value for “population" and low value for the other attributes. As for Singapore, it is distinguished by “population” and “no. of motorcycles" attributes.

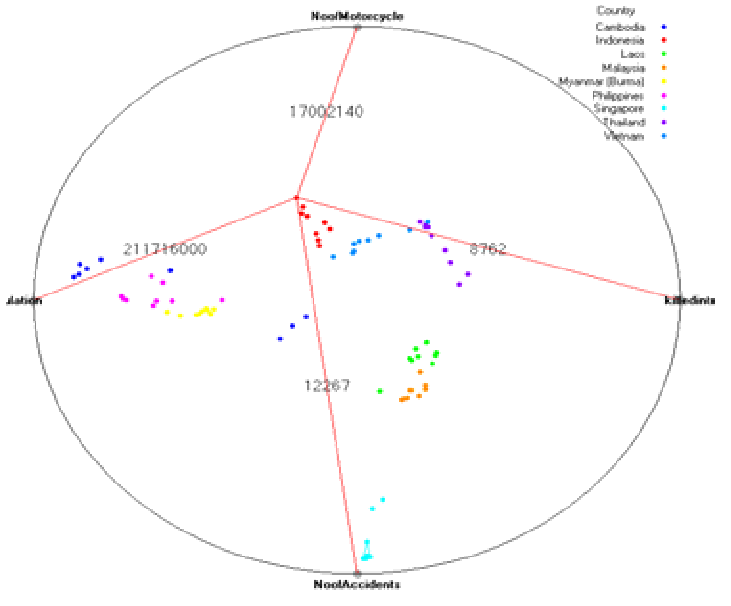


Figure 16. Radviz plot of five attributes from Asian countries, used to explore Indonesia.

- A scatterplot (Easton and McColl Statistics Glossary v1.1) is used to summarize a set of bivariate data. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model (see Figure 17). The scatterplot shows the relationship between the number of vehicles and the number of people killed in traffic in the ASEAN region.

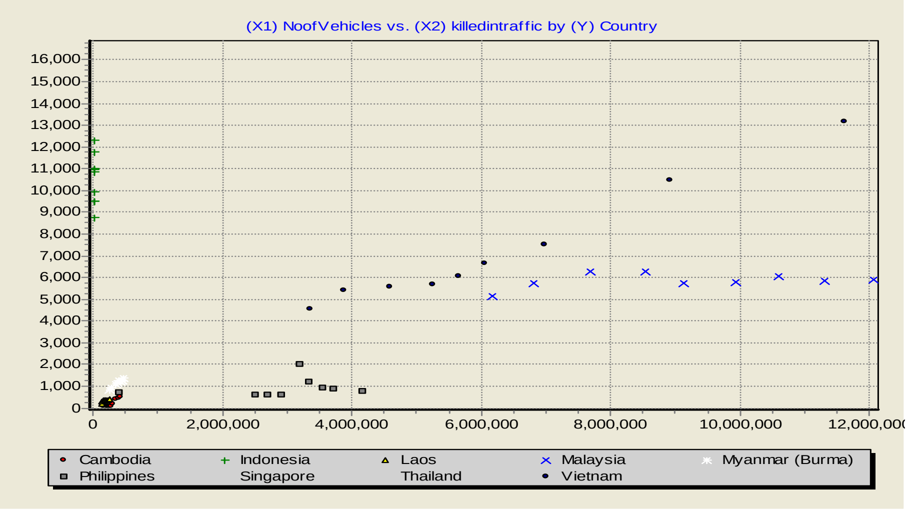


Figure 17. Scatterplot represents the relationship between number of vehicles (NoofVehicle) and Killed in traffic (Killedintraffic) as per country in ASEAN.

- Decision Tree (DT) is a classification tree, which determines an object's class by following the path from the root to a leaf node (Poole, 1998). It chooses the branches according to the attributes value of the objects. Decision trees are induced from training sets. Classification rules can be extracted from the decision trees (Figure 18).

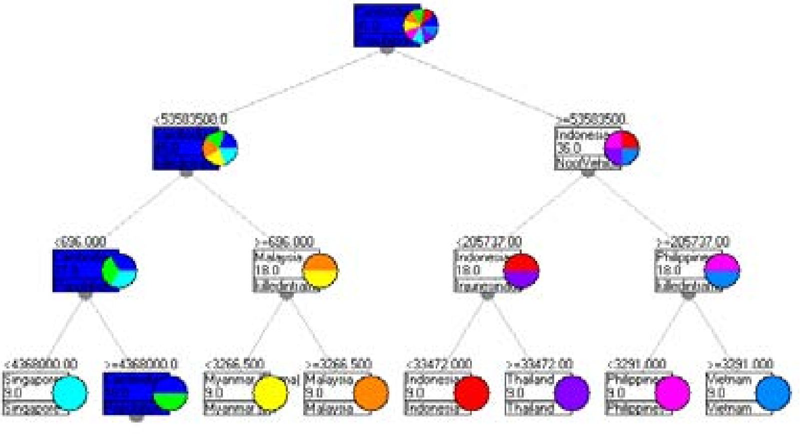


Figure 18. Decision tree with four levels and fifteen nodes.
Eigth of the nodes are leave nodes.

3.4. Knowledge flow in data mining

The first step in data selection in a data mining process is to build up the knowledge flow. A knowledge flow represents the knowledge discovery process as a stream diagram (Demsar et al., 2004). The main step in building the knowledge flow is to:

- Import the data set
- Compute the descriptive statistics
- Select the target and the input attributes (discrete and continuous)
- Choose the learning algorithm
- Build the prediction model and visualize the results

Both Tanagra and Orange (data mining tools used in the research, see section 3.5) support the building of knowledge flow using a drag-and-drop interface.

3.5. Tools for data mining: Tanagra and Orange

To perform machine learning and apply DM algorithms, two tools were used; the first was Orange, which is a machine learning and data mining open source software (Demsar et al., 2004). Orange is a comprehensive, component-based framework for both data mining and machine learning. Among other features, Orange has the ability to accept various popular data formats as input. It supports different visualization techniques that are useful in this research as well.

The second tool used was Tanagra, free DM software for academic and research purposes. It supports several data mining methods that include, but are not limited to, exploratory data analysis (EDA), statistical learning, and machine learning. The disadvantage of Tanagra is that it includes limited visualization techniques (dendrogram and scatterplot) (Rakotomalala, 2005).

3.6. Thinking and acting in terms of system

- **System: Definition, movements, and taxonomy**

The definition of a system is related to the perspective of the terms used to describe something that is essential in our life (IEEE 1220, 1999). Examples of systems are the university system, the transport system, the vehicle system, and so forth. Systems are always present and affect us. Some systems are part of nature and others are man-made.
There is a growing awareness amongst researchers about the importance of studying and understanding complex entities, which are composed of groups of entities, in a holistic manner (Arnold and Lawson, 2004). The search for holistic methods has its roots in

history, going back to the Greek philosophers. Scientific methods have led to the need for isolation of the phenomenon that is to be studied.

Any system can be classified in a certain way. The taxonomy of the system (Checkland, 1993) helps as a useful frame to understand different properties of a system. The following are the main categories of a system:

- Natural systems have their origin in the universe, and are a result of the force, which characterizes the universe. These types of system cannot be different than they are; at least not from a short-term perspective.
- Defined physical systems were created to satisfy and meet specific human needs and purposes. The elements of these systems have strong relationships.
- Defined abstract systems were designed by humans to serve some explanatory purposes. They are different from other systems in that they have no physical artifacts. An example of an abstract system is a mathematical description.
- Human activity systems, such as road transport systems, are observable in the world of innumerable sets of human activities that are ordered in wholes, as a result of some underlying purpose or mission.

- **System focus and achievements**

Systems are essential. They are created for specific purposes and work to achieve these purposes and goals (Lawson, 2006). The organization must focus upon the institutionalization of its portfolio of system assets. It is an essential requirement to have a sustained portfolio and system assets in good condition to manage and organize. Assets are any type of product produced by the organization, or the infrastructure that is essential to produce the services or the product.

- **Road safety management as a system**

Improving road safety in any country is the responsibility of the Road Safety Management System (RSMS). In Sweden, the Swedish Road Administration (SRA- http://www.vv.se/) is the authority assigned with overall responsibility for the entire road transport system. In accordance with the system classification of Checkland (Checkland, 1993), RSMS can be classified as a *human activity system* that has underlying missions, purposes, and goals. RSMS can be defined as a *sustained* system that has been institutionalized for a relatively long time with system assets that help in managing the road transport system (Beer, 1985).

The mission of SRA is to provide a safe, environmentally sound and gender-equal road transport system. The main purpose of the system is to offer individuals and business communities easy accessibility and high transport quality. The main goals are that there should be no fatalities or serious injuries in road traffic, based on the "Vision Zero" program (Safe traffic – Vision Zero on the move, 2000). The basic activities taking place in RSMS represent the system with its elements, adapted from Baguley (2001), shown in Figure 19.

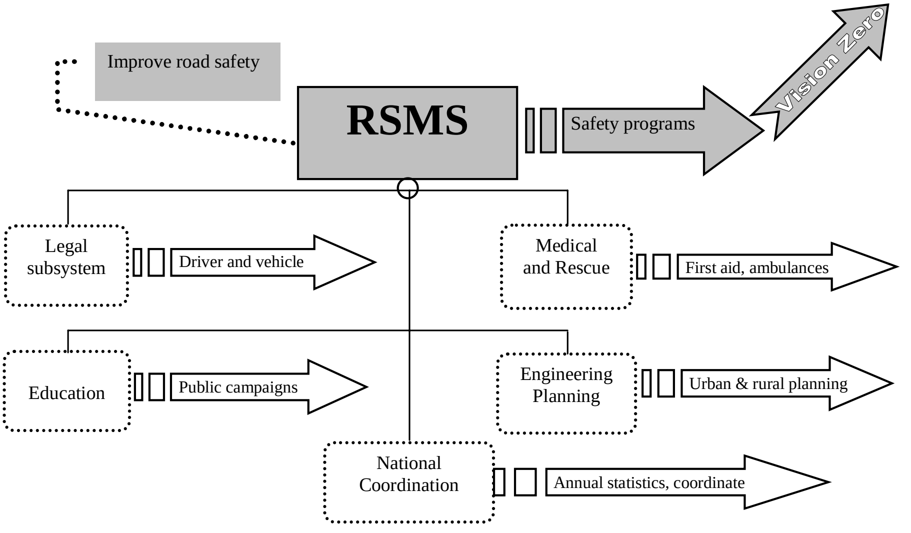


Figure 19. RSMS's Structure, needs, services, and effects on its elements.

Each of the subsystems of RSMS delivers a service and has a specific effect on the overall system. Figure 19 shows a need for RSMS to improve the road safety in Sweden. This need represents the *threshold* that keeps the system running. Each of the system elements in Figure 19 contributes to RSMS by providing services. When these elements work together, they provide the general effect of the system (offering a high quality road transport system that results in a safe society). In Sweden, Vision Zero is the main effect of the system. RSMS behave in a way that helps to prevent accidents and reduce their occurrence.

Policy makers and planners are developing road safety strategies and programs. They need an understanding of the problems, including all dimensions and causes of the accidents (Ghee et al., 1997).

The magnitude and nature of the different problems related to road safety can be identified; the knowledge extracted from the road accident database helps to identify exposure groups and risk groups.

Moreover, knowledge about accidents is required to identify risk factors that lead to the accidents (SWOV, 2001). STRADA was the main source of knowledge for SRA.

- **How we view the system**

Different individuals have different perspectives, roles and responsibilities with respect to specific systems. STRADA stakeholders have different perspectives on the system and they play different roles in RSMS throughout the country. Therefore, STRADA can be viewed in three ways (Paper V):

- An **asset**, used by stakeholders, which helps them to achieve their goals

- A **product**, produced by the development company on behalf of SRA
- An information **service**, in terms of annual statistics and publications

The different viewpoints influence the way the system will be managed during its life cycle.

- **Changes in the systems**

For any man-made system in an organization, there is a time when the organization needs to introduce changes to the system. The change model presented in Lawson (2006) is either a change in the system description (structural change) or changes in the operational parameters of the system. Structural changes can include recreation or retiring of the entire system, adding or removing system elements, adding or removing services of the system elements, and/or defining relationships between system elements.

On the other hand, operational changes are related to alternation of the operational parameters. This type of change can affect the behavior of the system services. Another type of operational change is a change of operation mode when the system in operation provides multiple modes.

Data and information are always required and they are key to making proper decisions. The knowledge is utilized in providing feedback from decision-making and for planning the change itself.

The knowledge is captured and collected using an accident data acquisition system. Such a system should be at the heart of safety improvement (IRTAD, 1998). The focus in this study was to identify the knowledge, supporting decisions concerning the changes, and to measure the effects of the different changes by applying the basic change management model. Figure 20 depicts how such a model can be introduced and the positions of knowledge in the model.

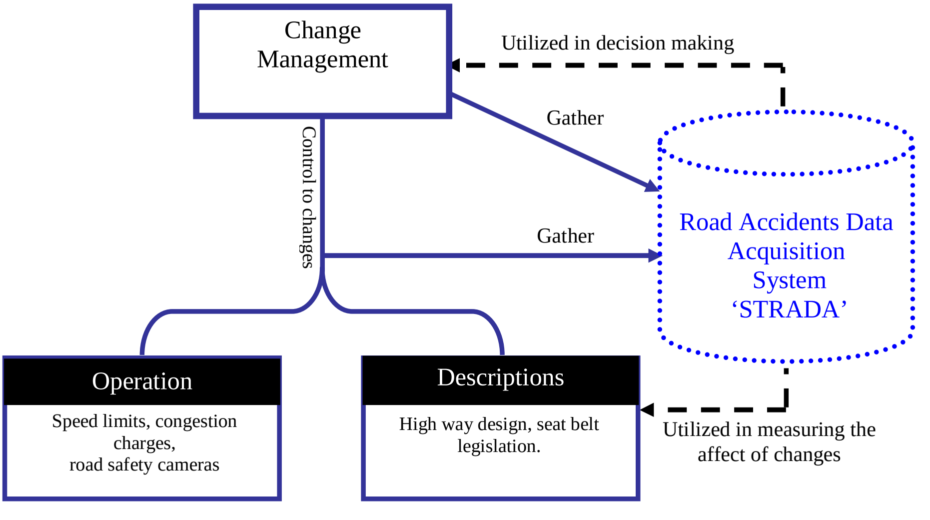


Figure 20. Application of fundamental change management model (Lawson, 2006).

4. Internet GIS

In Papers II and III a new method to benchmark road safety was introduced. This was accomplished by applying techniques from Internet GIS. The main model used was the portrayal model for interactive maps introduced by Open Geospatial Consortium. In this chapter, concepts from Internet GIS are introduced and the model used is discussed.

4.1. Web mapping and Internet protocols

The increasing demand for Internet services has made the Internet an integral part of society. Internet GIS is defined as "*a research and application area that utilizes Internet and other Internetworking systems to facilitate the access, processing, and dissemination of geographic information and spatial analysis knowledge*" (Peng and Tsou, 2003).
There are two reasons that force geographical information providers to explore the Internet to disseminate geospatial data (Gillavry, 2000). First, Internet architecture and protocols allow the users to interact visually with the spatial data. Second, the near ubiquitous nature of the Internet gives it accessibility.
During the development and evolution of geospatial data storage, retrieval and dissemination for various users and organizations has moved from stand-alone platforms to web services, offered by GIS software vendors to the users over the web (Taylor et al., 1997). Hence, web map services need to be maintained through agreed concepts and with tools that can be useful and satisfy the following requirements (Douglas, 2000):

- The organizations or the developers may need to have their maps online; these maps can either be simple ones that can be displayed as a single map at a time or maps produced from different sources by overlaying the different views.
- Users need to view the metadata, to help clients get a picture of the data and the maps offered to them.

Web mapping, as a tool for development in GIS, makes it possible to share geospatial information using available web browsers. Maps can be presented as general-purpose maps that mainly can display locations and other geographical features (cartographic elements). In this case, the maps can be regarded as static maps, and more advanced web mapping should provide interactive mechanisms.

4.2. Basic components of Internet GIS

The uniqueness of Internet GIS lies is its components and architecture, compared with other web applications, as publishing maps over the Internet requires special architecture. The four components of Internet GIS (Peng and Tsou, 2003) can be seen in Figure 21.

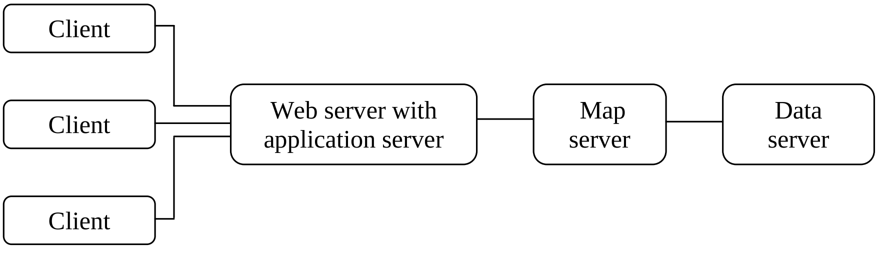


Figure 21. Basic components of Internet GIS (Peng and Tsou, 2003).

- **Clients**

Internet GIS users use clients to interact and perform different operations and tasks. Traditionally users of GIS use Graphical User Interfaces (GUI) to interact with different applications. In Internet GIS the situation differs when the users (clients) rely on HTML and add-once clients. HTML is not sufficient for user interaction. However, this can be overcome with the use of DHTML, which allows more efficient interaction with spatial objects. In addition to DHTML a client side application can be used, such as plug-ins, Java applets, Java beams and ActiveX controls.

- **Web server and application server**

The web server is called the HTTP server; its function is to respond to the requests made by the browser (client). Three methods can be used by the web server to respond to the clients' requests: first, by sending Java applets or ActiveX control to the client; second, by passing requests to other programs and invoking other programs capable of processing the queries; and third, by sending existing HTML documents to the clients. The application server is a glue program that receives client requests; when it is passed to the web server the application server acts as middle ware that connects both web server and any other server-side applications, such as the map server.
Basic functions of the application server are:

I. To establish, maintain, and terminate the connections between the web server and the map server
II. To act as a translator between the web server and the map server
III. To interpret the clients' requests and pass them to the map server
IV. To manage concurrent processing and requests
V. To balance the load among the map servers and the data servers
VI. To act as a managers of the state, the security and the transactions

- **Map servers**

Map server components (also known as a spatial server in some commercial products) perform traditional desktop GIS functions. The function of the map server is to respond to spatial queries by performing spatial analyses and generating maps for the clients.

- **Data servers**

Spatial and non-spatial databases are available at a data server. Other components such as web servers and map servers can access the databases using traditional SQL.
It is clear from the discussion above that Internet GIS relies on extensive use of graphics that can reduce the performance of the users. On the other hand, Internet GIS reduces the requirements needed in desktop GIS such as advanced computers, user training, and site-licensed software. None of these, however, are required to operate a web GIS (Strand, 1998).

4.3. Map publishing methods over the web

- **Static maps**

Static maps are published over Internet web servers: In the case of static maps, the servers will handle the maps without the need of any GIS programs. The server can be any HTTP server (Plewe, 1997). The maps need to be in a format that can be understood by the browser. The server responses to the clients' requests which are sent in HTTP format.

- **Dynamic maps**

Dynamic maps need GIS software connected to the web server. The software can generate maps upon clients' requests. The information sent by the clients to the web server should include information such as the requested layers and the map extension. All of this information will be processed by the server and sent to the browser in JPEG or GIF format. Such architecture is known as an Internet Map server (Toon, 1997).

- **Interaction modes with web maps**

Web browsers can be discussed from the user point of view, as graphical interfaces that give a page layout or could help to render and display the images downloaded from a certain web site. Basically, for Internet users browsers help to view web pages and navigate through the web easily (Richmond**,** 2002).
Different types of browsers have different functionalities. Browsers can be categorized as *graphical browsers* or as *text browsers.* Table 2 shows the differences between graphical and text browsers (Jacobs et al., 2002).

Table 2

Differences between graphical and text browser interfaces

Feature	Graphical	Text
Presentation	Text, images, audio, and video can be retrieved	Only text is available to the users
Navigation mode	Clicking and pointing, by means of the mouse, on highlighted words or images	Highlighting emphasized words with the arrow up and down keys, and then pressing the Enter key to follow the links.
Software program	Internet Explorer, Netscape Navigator, and Mozilla	Lynx
Script support	Both Java applets and Java scripts are supported.	No support for Java and scripting languages
Operating system	Mac OS, Windows, Unix	Unix and VMS

In addition to the graphical and text-based browsers, desktop browsers differ from handheld browsers in that they have limited functionality.

4.4. A portrayal model for interactive maps

Different types of maps can be published over the Internet (Cobb and Olivero, 1997): static web maps (not useful for user interaction) or dynamic maps over sophisticated sites that can help the user to customize the maps on most computer platforms. The Open Geospatial Consortium (OGC), Special Interest Group (SIG) on WWW mapping produces a portrayal model for interactive maps.

- **OGC four tiers**

I. Select, representing the selection process performed by the users to retrieve data from geospatial data sources. This can be accomplished by using map query constraints. Also, this allows the users to perform thematic search and selection, i.e. adding from available layers to the current map view.

II. Display element generator, this tier turns the selected geospatial data into individual display elements. This is performed by displaying a sequence of elements. During this process, styles are attached to the maps.

III. Render, in this tier the displayed element(s) will be rendered.

IV. Display, this process is responsible for making the map rendered in the previous process visible to the users of the web browser.

The data flow between these four tiers shows the conversion process from one data type to another, according to the specific tier requirements. It features data types returned as result of a user query in the selection tier, display elements generated by the second tier, and images produced by the rendering process. The conversion process is shown in Figure 22.

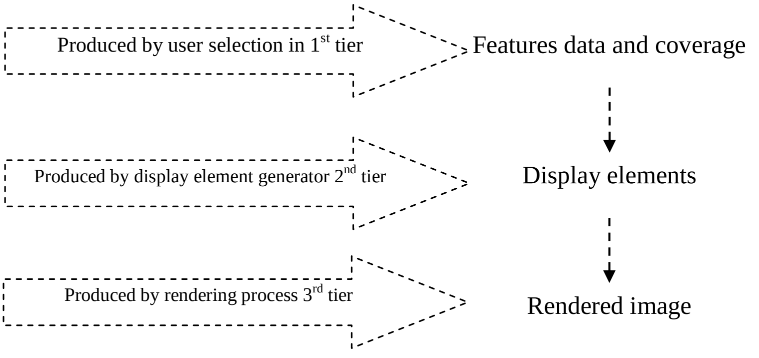


Figure 22. Information flow over OGC portrayal, www.opengeospatial.org.

- **Levels of interaction and GIS functionality**

As previously mentioned, different browsers provide different capabilities to users. The differences in browsers lead to different levels of interaction and GIS functions.
ArcIMS, a server-side Internet GIS (http://www.esri.com/), provides the developers with two options. The first option is to use a Java Viewer and the second is to use an HTML Viewer. Both viewers provide the users with web maps over the Internet, but they differ in functionality and in the operations that can be performed while using Java plug-ins.

Java Viewer allows drawing and editing of different objects. Rectangles, circles and polygons can be drawn but the user cannot draw either lines or points over the map to select specific features. Some spatial analyses cannot be accomplished, which shows the limitations of Internet GIS compared to desktop GIS. HTML viewers are lightweight and do not need any java plug-ins. Additionally, HTML viewers can be accessed with low bandwidth. In contrast, Java applets and ActiveX require broadband to run.

5. Research methods and approach

The aim of this chapter is to introduce and discuss the methods and approach of the research. The chapter also describes the techniques and procedures used to identify the system requirements and the way in which these requirements were converted to the specifications implemented in the platform GLOBESAFE.

5.1. Selection of research methods

Suitable research methods are required to help the researcher improve his/her knowledge of the problem surrounding the research (Easterby-Smith et al., 2002). Early in section 1.3, research design and process were discussed, showing how different methods are combined and the results obtained. Research methods in road safety range from active safety methods to passive safety methods. Some research focuses on engineering methods to improve road equipment and conditions, improve vehicle safety, study different accident models, and also human factors and their contributions to safety.
For this research in particular, the methods were implemented in order to develop analytical tools to support road safety organizations in evaluation and auditing of road safety. Initially data was collected using qualitative methods to identify user's requirements. To implement these requirements methods such as iterative prototyping were used during the implementation. The main advantage of this method is the opportunity to get user feedback to consider for the next cycle of the implementation. Data mining methods were also used to discover knowledge in road safety databases. In the following section all of these methods are introduced.

5.2. Requirement identification and prototyping

In the system development life cycle, the first phase is the process of discovering the requirements of the system by communicating with the stakeholders. The requirements are defined as "*the features that the system must have or a constraint that it must satisfy to be accepted by the client*" (IEEE Std 1233, 1998 and 1233-1996). The main objectives of the requirement identification are to:

- Find out about the application domain (road safety)
- Determine the services to be provided by GLOBESAFE
- Specify and understand the operational constraints

In order to identify the requirements, certain techniques were employed to collect information about the system requirements. The guidelines from Sommerville and Pete (1997) were followed. To start with, the system stakeholders were identified.

The system stakeholders are those who collect, analyze or use road safety information in their professions or for other related research purposes. Initially, a list of optional users was maintained, which included their positions. This helped to avoid missing important requirements. The advantage of this step is that it makes users feel that they are part of the research process (Gottesdiener, 2002). Requirements were determined initially by interviewing the stakeholders and the experts in the field (Myers, 1997). Discussions during the interviews with different stakeholders helped to build up an understanding of the system and its requirements (Murch, 2000).

Requirement prototyping

To enhance the understanding of the vague requirements, a prototype was used to complement the interviews. The prototype is defined as "*a demonstration system which shows end users and system stakeholders what facilities the system requires*" (IEEE Std 1233, 1998 and 1233-1996). The reason prototyping is used is to simulate the behavior of the system, enabling end users to refine their ideas about the requirements of the system (Gomaa and Scott, 1981). From the research point of view, the value of building a prototype and its contribution to the scientific process should be considered. Reference to the literature shows the scientific value of prototyping: "*Systems are good science*" (Hendler, 1995). Other reasons for using prototyping as a method include pragmatic and non-scientific reasons such as technical implementation during software life cycle management. During the research process, prototypes can be used as an object or setting for an experiment that helps to identify user requirements and perceived preferences about the system. Also, prototypes can be used by the researcher to get better understanding of the research questions and the proposed solution. Such use of prototypes does not provide any new knowledge, but helps to get new ideas and insight (Forslund, 1997). As a research method, a prototype can be used either as a proof of something or as an object of study (Forslund, 1997). In this research, the prototype was used as an appropriate method to find out more about the services and functionality the system should provide.

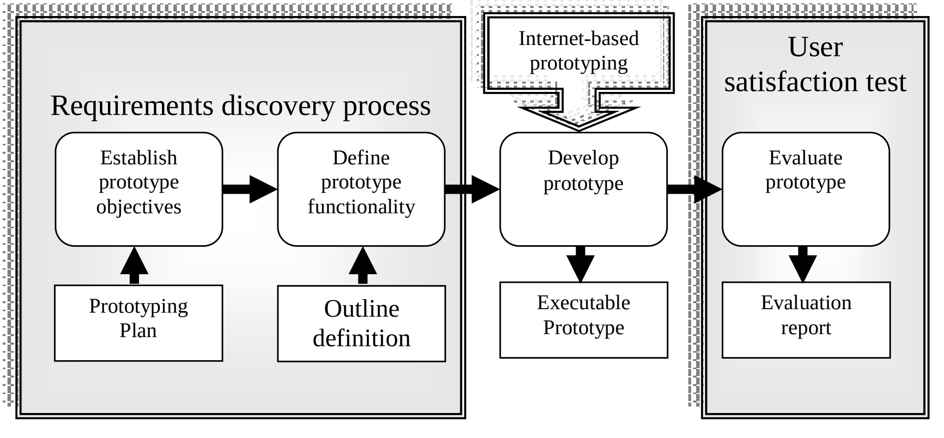


Figure 23. Prototype development during the research process.

During the prototyping process considerations were given to:

- study the properties of the prototype and its operations
- build the prototype from back-end to front-end including user interface development
- evaluate the usability of the prototype with respect to the user's perspective and experience; here user interaction and usability testing techniques were used

The prototyping process used followed guidelines described by Sommerville and Pete (1997). The many steps of the process are shown in Figure 23.
During the requirement identification process, the objectives of the prototype were established, together with the prototype plan. This led to a clear definition of the functionalities of the prototype and its different components and subsystems (Papers II, III, and IV). The second step was to implement the prototype. The operational requirement of the system was to use web technology, making it web-based. The evaluation process followed, to get user feedback to determine the level of satisfaction about the system. This process ended with a report that contained suggestions from the users on the modifications that were required. User satisfaction was measured using a Questionnaire for User Interaction Satisfaction (QUIS; see section 5.7) and usability methods that are discussed later in this chapter.

Advantages of prototyping

The main advantage of the prototyping process is to help in understanding unclear and vague requirements (Andriole, 1994). In addition, when the stakeholders understand the prototype better, they will start to pay more attention to the details than before while working with the prototype. The comments and the input they provide are in much greater depth, more insightful, and far more thorough than previously.

The prototype helps to measure the usability issues, as the users cannot think about usability until they experience the prototype. Another aspect of giving the stakeholder something to experience and interact with (prototype) is that it improves their understanding and helps them to imagine how the system could work. Moreover, the prototype makes it easy to understand the real requirements of the stakeholders and helps to develop effective user interfaces. The requirements of the information sharing platform are listed in the results chapter and Paper IV.

5.3. Development of a common vocabulary

To overcome definitional and semantic problems, a common vocabulary was constructed to help to identify the concepts used in the field and then taken as entities in the database. Ontologies are formal and explicit descriptions of the concepts in a universe of discourse (UoD; Gruber, 1993). Here ontology was used to develop a common vocabulary for road safety (Guarino, 1998). This helped to standardize reporting, classification and analyses of road safety data.

There is no one "correct" way of developing ontologies (Gomez-Perez, 2004). The process is iterative, comprising the following steps:

- Determine the domain, scope and users
- Competency question scenarios
- Domain analysis
- Knowledge acquisition
- Knowledge coding

Each of the above steps was applied to model ontologies, resulting in a generic domain model. The ontologies were used to build database schemas (Svedjemo and Jungert, 2006). The resulting model provided a common vocabulary used to develop the database schema in GLOBESAFE.

Conceptual map for representing and communicating knowledge

Conceptual map (CM) tools were used as a way to convey concepts to the experts in a clear and understandable way (Albert and Steiner, 2005). They have many features, represent the concepts in graphic and visual form, and are practical knowledge representation tools (Novak, 1998).

In the CM, concepts are included within nodes and the relations between the concepts are illustrated by means of lines or arcs connecting them (Castro, 2006). The lines, in turn, have associated words describing the nature of the relationship that links the concepts (see Figure 24) and chapter 6.

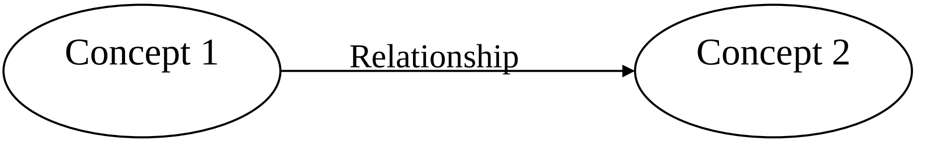


Figure 24. Conceptual maps are graphs that consist of nodes, with connecting arcs that represent relationships between nodes.

Domain information and knowledge level

The domain information, knowledge level, basic entities in the domain, and competency questions were identified before developing the ontology.

The ontology constructed for road safety was modeled in Protégé support frame-ontology paradigm (Gennari et al., 2002). The metadata were generated in formats that can be reused in different applications. After building the CM, the next step was to code the knowledge in Protégé, which included the following:

- Definition of the classes
- Arranging the classes in taxonomic order (Subclass - Super class)
- Defining the slots and allowed values

5.4. Study of STRADA at the Swedish Road Administration (SRA) 'Vägverket'

The focus is to study accident data acquisition systems. A system like STRADA is a prerequisite to the organizations to share their information with others and to collect the necessary statistics about accident and road traffic in general. In Paper V, we used system thinking approach to analysis and study the results from Focused-Group-Discussion (FGD) and interviews. Our intention was to investigate in-depth the low acceptance of STRADA by health authorities in some counties and provinces. The importance of STRADA was discussed within the context of change management, by viewing road safety management from a system perspective.

The most serious concern in road safety research is how to obtain adequate data about accidents in an effective manner (Larsen, 2004). The study of road accident data systems includes reporting and data collection procedures, the analysis and quality control measures employed, and the communications systems used for data transmission (DLTR, 2002; TRA, 1999).

The accident recording system forms the foundation for the conceptual model developed in this work (Paper IV) for information sharing among road safety organizations. STRADA is an example of such a system. For a long time, the Swedish Road Administration (SRA) has developed standards for accident data acquisition (Forward and Samuelsson, 2007). Many systems have been developed and abandoned over time. In December 2002, the SRA accepted STRADA as an accident reporting system and it is used by the Swedish police and at hospitals.

The objective of studying STRADA is to gain knowledge about road accident data collection systems. This includes basic data sources, users and different data acquisition procedures (personal information, vehicle information, injury classifications, and location information).

STRADA was studied at the regional office of SRA Region Southeast in Jonköping, one of the seven main regions in Sweden. The aim was to gain more knowledge about the current system and how to utilize it in the operation phase. Qualitative methods were used and interviews conducted with the STRADA group. The way in which the system operated and the routine work, both at the police offices and at the Emergency Unit were discussed and studied. The types and components of the data entry forms, relations to other systems, user interface, usage of the system in the region, quality assurance routines were all considered during the study.

Analysis

A system thinking approach was used to analyze the findings of the study (Checkland, 1993). This involved thinking in terms of the system and observing the dynamic behaviors of the system in operation (utilization stage). This helped to identify the system structure, infrastructure, and stakeholders. In that way, the structure, needs, services, and effects on road safety management and associated elements were identified (Paper V).

Technology acceptance model 'TAM'

The Technology Acceptance Model (TAM) shown in Figure 25 (Davis, 1989) helps the end users and the customers to study and predict the use of the information system and other technology . In this work, TAM was used to study the acceptance of STRADA. The main concept behind TAM is the concept of *acceptance*. Dillon et al. (1996) defined acceptance as "*the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support*". The main components of TAM is its perceived ease of use, defined as the degree in which the users believes that using the system will be easy and take less effort. This study applied concepts from the software engineering quality matrix system usability. Usability is managed and obtained during system life cycle management (in the specifications stage). Perceived usefulness (PU) also reflects the degree to which the users believe the system will help to improve the performance of the organization. PU relates to the operability of the system with the existing system at the Emergency Unit at hospitals. The emergency unit uses their own system to report injures and accidents.

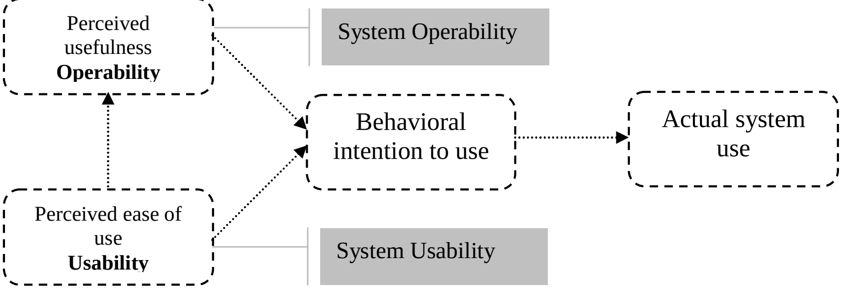


Figure 25. Incorporating system usability and operability with TAM.

The aim of the study of the system users was to examine the reasons for actual system use being less than 50%. The respondents, during the focus group meeting and the interviews, explained that STRADA has a complexity that makes it difficult to use and extra resources and training are needed to make it useful (Paper V).

5.5. Visual Data Mining (VDM) and clustering methods

Road accident statistics have the characteristic of being collected and used by a large group of users. This results in a huge volume of data, which needs to be explored to capture the hidden information. There is a potential problem that information is hidden, which is due to the accumulation of the data. This limits the exploration task to the road safety experts, which decreases the utilization of the data stored in the database. In order to help solve these problems, VDM methods were applied and their application to information discovery in road accident databases was investigated (Paper VI).

To perform machine learning and apply VDM algorithms, two open source tools were used. The first was Orange, machine learning and data mining software (Demsar, 2004).

The second tool was Tanagra, free DM software for academic and research purposes. It supports several DM methods that include, but are not limited to, Exploratory Data Analysis (EDA), statistical learning, machine learning and visualization techniques such as dendrogram and scatterplots (Rakotomalala, 2005).

The data set used in this part of the research was from two international road databases, which have good information quality and have been used for quite a long time, i.e. the International Road Traffic and Accident Database (IRTAD) and GLOBESAFE. IRTAD gathers data on traffic safety. The main part of the database, with around 500 data items, includes aggregated data on injury accidents, road fatalities, vehicle populations, and network length. The data are collected from 28 countries, for 1965 and every year since 1970 (Heinrich and Mukulik, 2005). The second data set used was GLOBESAFE, which was developed during the first two years of this research. GLOBESAFE stores data for the nine ASEAN countries from 1994-2002 (Abugessaisa et al., 2007).

Paradigm for VDM

The paradigm for VDM is described in Stuart (1999). Figure 26 shows the steps in this paradigm.

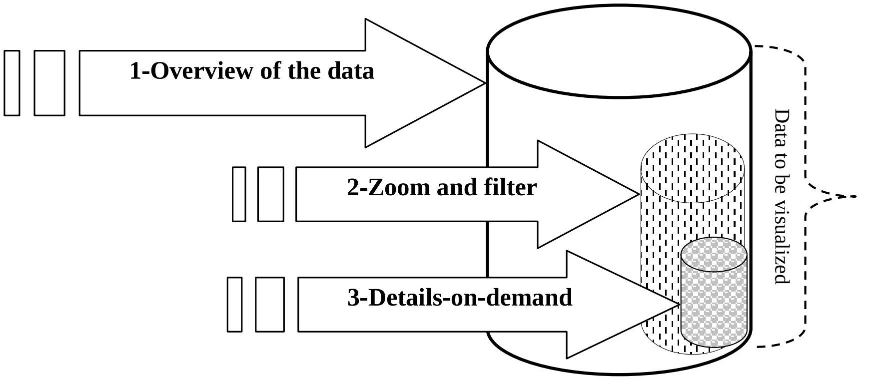


Figure 26. Paradigm for VDM.

The users start by getting overviews of the data context. This step helps to identify areas of interest (patterns) in the data and then it focuses on specific patterns. After this the users can start to analyze the selected patterns. In step 2, this is performed by accessing the details of the data.

Clustering methods

A clustering problem is an unsupervised learning problem (Soukup and Davidson, 2002). The aim is to identify a cluster of similar countries sharing a number of interesting properties (level of motorization, number of fatalities, same network length, etc.). It is used to find the similarities among countries in the same or different regions. Clustering methods help to automatically represent a data set using a small number of regions and preserving the topological properties of the original input space.

- **K-means algorithm**

With its distance-based clustering methods, the K-means technique is considered a classical clustering method (Jain et al., 1999). The main feature that made K-means useful in this research was the ability to specify in advance how many clusters were being sought. This number refers to the parameter K (here it referred to the number of countries). From these, K-points were chosen from random clusters as cluster centers. All instances were assigned to their closest cluster center according to Euclidean distance metric. The next step was to calculate the centroids or the mean of the instances in each cluster. These centroids were regarded as the new center values.

- **Hierarchical Agglomerative Clustering 'HAC'**

HAC is a clustering method that produces "natural" groups of examples characterized by attributes. HAC uses a bottom-up strategy, which starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters until all of the countries are in a single cluster (Rakotomalala, 2005). A tree structure called a "dendrogram" (Figure 27) is commonly used to represent the process of hierarchical clustering. The dendrogram tree represents the successive agglomerations, starting from one country per cluster, until the whole data set belongs to one cluster. The main advantage of HAC is that the users can guess the right partitioning by visualizing the tree. A disadvantage is that it requires computation of distances between each example, which is very time consuming when the data set is large.

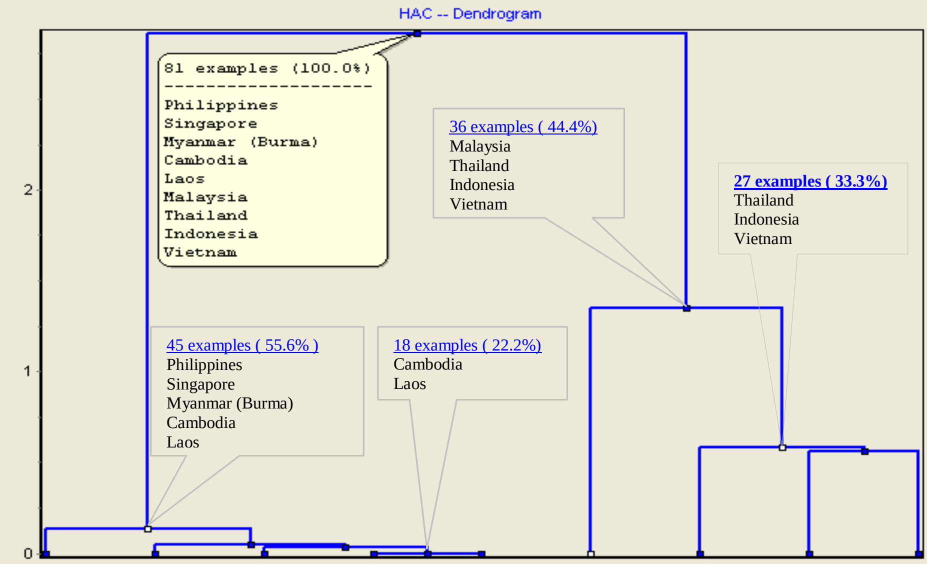


Figure 27. HAC / Dendrogram.

- **Decision Tree 'DT'**

A decision tree (DT) is a classification tree, which determines the class of an object by following the path from the root to a leaf node (Poole, 1998). It chooses the branches according to the attribute value of the objects.
The DT method uses a 'divide-and-conquer' strategy. The nodes in a DT involve testing a particular attribute. The test at a single node compares an attribute with a constant. The leaf nodes give a classification that applies to all instances that reach the leaf, or a set of classifications.

Furthermore, for visualization of large quantities of data (Tufte, 1986), the following graphical visualization methods, introduced in section 3.3, were used:

- Histograms (estimate the probability distribution for a certain numerical attribute of a given set of objects).
- Dendrogram for visualization of K-means clusters (a correlation analysis between attributes or objects) to help in visualization.
- Parallel coordinates (Inselberg, 1981) used as a VDM tool and originally used as a new way to represent multi-dimensional information. In parallel coordinates, all axes are parallel to one another and equally spaced.
- Radviz (radial visualization); in Radviz the occurrences are represented by points plotted inside a circle (Hoffman et al., 1997). See chapter three.

5.6. Information Architecture (IA) for the map design

Information architecture (IA) principles aim to create environments with logical structures that may help the users to find answers and complete their tasks (Rosenfeld and Morville, 2002). IA concerns organization, labeling, navigation in, and search for the information (Dong and Agogino, 2001). Figure 28 depicts the way information on the Internet should be organized and presented in a web page. The principles are displayed in the user interface for the web map that has been reproduced in Figure 29. Road safety statistics tables and sheets are difficult to interpret; experts and professionals prefer to use visualization and mapping tools, rather than interpret different kinds of lists and numbers.
The maps for the ten countries located in Southeast Asia, i.e. the "ASEAN" countries, were selected by using the "select by theme" command available in the ESRI Arc GIS world database. Figure 29 shows one example of this map and how information architecture principles are applied.

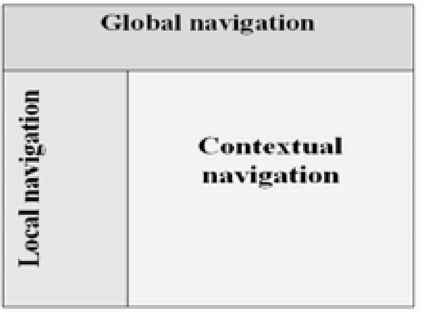


Figure 28. Information architecture principles.

Map as presentation and communication media

Commonly, maps are used to present facts to people in a cartographic form. In order to do that, maps need to be designed and organized in a way that could be easily understood by the users.
The different uses of maps show that there should be some design principles and guidelines for developing web maps (static and dynamic). These guidelines were used to design all of the maps in GLOBESAFE.

Design principles

If the maps are designed following good design principles and good practices, then they will convey relevant information to the users. Also they will inform, convince and persuade decision-makers, giving a full understanding of the problems presented. Moreover, they will help to investigate phenomena using map analysis in a GIS environment.

In the case of web maps, the designer is guided by both cartographic design principles and web page design. Three basic components should be considered before starting to design a map (Bernhardsen, 2002). First, the map information to be conveyed to the user is identified. This defines the basic motives and goals of the map. Second, the environment in which the map will be used should be clearly defined, with references to its properties. An example of this is when the map is used over the web with or without a handheld device. Sometimes maps are designed to be printed and used as hard copy.

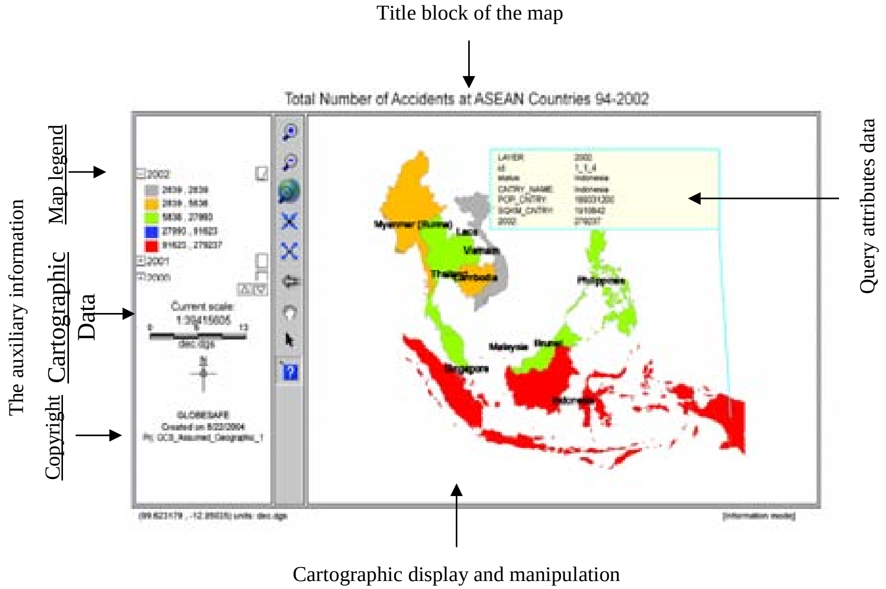


Figure 29. Applying information architecture principles to web maps.

Map elements

An effective and good map has both basic elements and secondary (optional) elements, which are required to capture the message and abstract the reality (Kenneth and Shannon, 1995). These elements should be ordered in accordance to the user's visual perception abilities and a logical hierarchy.

- Basic elements: provide the reader with critical and sensitive information about the map. Examples like map title, legend, north arrow, maps scale, etc.
- Secondary elements: assist effective communication and are always optional. Examples include neat line, overview map, border line, etc.

Visual hierarchy

How should these elements be organized and managed in a way that complies with the visual abilities of all users? The principles of a visual hierarchy should be used in the case of web mapping.
There are different definitions of visual hierarchies that can guide web page designers as well as web map designers to design effective maps.

Visual hierarchy is defined by Clarke (2001), as "*the perceptual organization of page elements such that they appear visually to lie in a set of layers of increasing importance as they approach the viewer*."

This is also defined by Graham (1999) as an "*arrangement of elements according to importance and emphasis. Typically, this involves emphasizing certain elements in order to influence the user to look at and interact with a certain item first, another item second, yet another item third, and so on.*"

The first definition emphasizes the perceptual organization of information layers, according to the level or degree of importance. The layer with the higher degree of importance will strike the viewer quickly. The second definition considers the importance of the map items themselves rather than information layers. Since each element is positioned in the correct order, the user will interact with them in the same way as they were ordered by the map designer. Both definitions can be translated into Figure 30. Figure 30 demonstrates the balance between element order and importance, and how these elements should be drawn within the map frame.

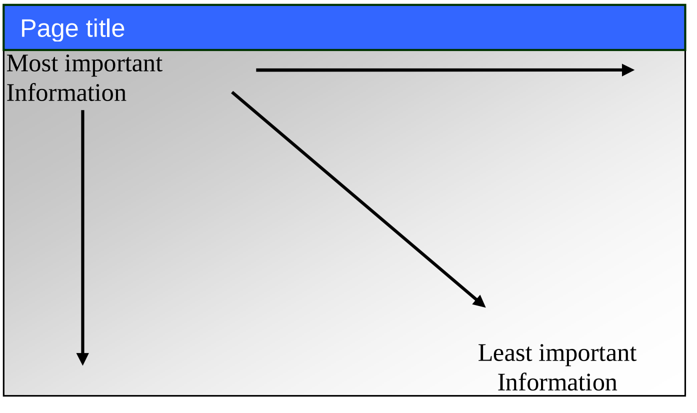


Figure 30. Balance between map frame and visually hierarchy
Source: http://www.colorado.edu.

5.7. Evaluation of GLOBESAFE

With respect to user evaluation, different methods are available to measure user satisfaction and acceptance of the solution. In the case of a web-based system, it is recommended that it should conform to methods of international usability (Nielsen, 2000). Usability and acceptance tests help to measure the extent to which the system complies with the usability design procedures, and also gives the designers feedback that help them to maintain a usable and acceptable system — in this case, to test the system and its acceptance by the users. For this a commercial usability questionnaire, provided by usability experts, was used. For accessibility, user scenario–based testing was performed in a trial session. Each scenario included a group of tasks following the guidelines given by Dumas and Redish (1993). General and demographic information was collected from the users before starting the test. This was used to classify the users into different groups and classes according to certain criteria. Respondents were asked three types of questions, which were:

Pre-test questionnaire

This questionnaire aimed to understand the users' background, knowledge, age and category.

Post-task questions

Tasks should be grouped together in scenarios. Scenarios describe the tasks in a way that takes some of the artificiality out of the test. Further guidelines and a description of a good scenario are explained in Dumas and Redish (1993).

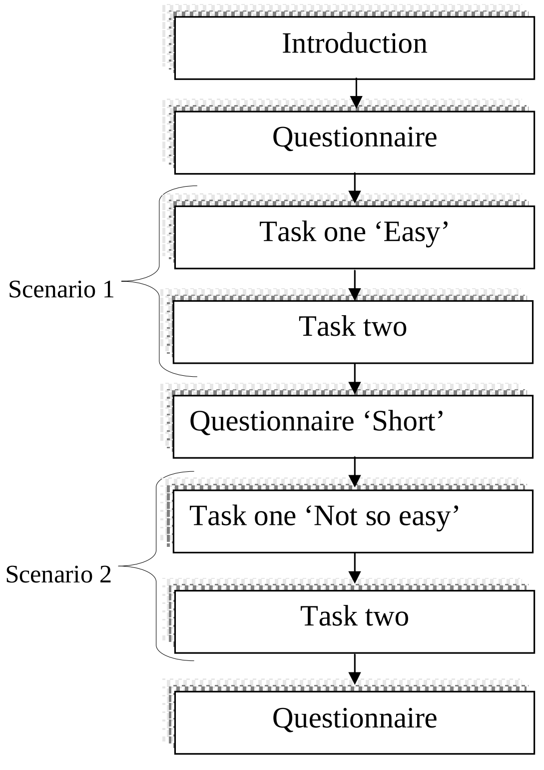


Figure 31. Scenario and questionnaire flow.

Further, a scenario consists of groups of tasks that should be performed by the users within a specified time. At the beginning of the session and before starting the test, short instructions were given to the respondents. The process flow shown in Figure 31.

Post-test questionnaire

This part of the questionnaire helped the tester understand how the respondent(s) felt about the system. The post-test questionnaire helped to evaluate system performance issues. Some questions from the Questionnaire for User Interaction Satisfaction (QUIS) were used (Harper et al., 1997). QUIS rates responses using a Likert scale (Chin et al., 1998). Examples of such questions are shown in Table 3.

Respondents were classified into three groups. Those working in road safety research, GeoInformatics, and those working in the field of user-centered design. Respondents gave their answers with respect to the map elements' visibility. They could select one of four alternatives: very good, good, poor, and very poor.
The second group of questions related to cartographic operations on the map. Here the level of acceptance of a certain operation was rated. The last group of questions, about

the navigation between different maps and home pages, rated the navigation offered from difficult to easy. The rest of the questions helped to identify whether the terminologies used were consistent or inconsistent. After all the questions were answered, the respondents had the opportunity to note anything that was not included in the questionnaire, but was observed during testing.

Table 3

Sample questions from QUIS.

Quality or feature	Rating scale	
The system is reliable	never always 1 2 3 4 5 6 7 8 9	NA
Operations are	undependable dependable 1 2 3 4 5 6 7 8 9	NA
System failures occur	frequently seldom 1 2 3 4 5 6 7 8 9	NA

Analysis of the questionnaire

Analysis of the questionnaires was based on the theories of qualitative analysis that emphasize issues, rather than just describing aspects of the analysis (Lantz, 1993). The general model that was followed in the analysis is described in Figure 32.

Data reduction is a continuous process, closely linked to the critical inspection of the study. The patterns found in the questionnaire were summarized and turned into diagrams. The questionnaire was analyzed using frequency tables and charting techniques and the results are presented in Paper III.

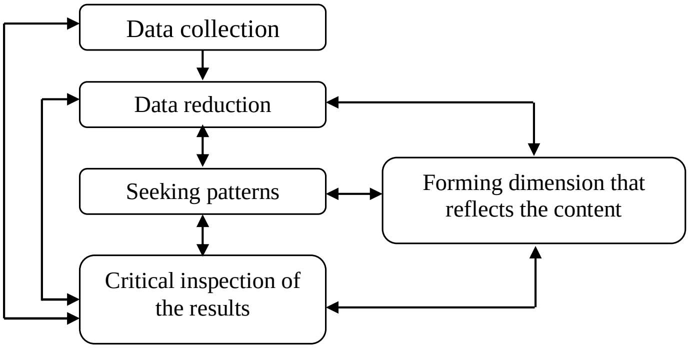


Figure 32. Model for qualitative data analysis (Lantz, 1993).

6. Results

The main contributions of this research mentioned in section 1.4. In this chapter we are going through these contributions as published in different articles. This thesis contributes to the design and implementation of a platform for information sharing among road safety organizations to support decision-making and the development of road safety plans. Moreover, visual data mining methods, for the discovery of hidden knowledge within the road safety database, have been implemented. An additional contribution is made by the application of the system thinking approach for the study of the road accident data acquisition systems; this approach offered good understanding of the process of reporting and disseminating accident data. We introduced different models to handle the changes and the weakness in the current system.

6.1. The conceptual model

During the early stages of this research a conceptual model was developed. The three-layer conceptual model addresses the main issues that were discussed in Chapter 1. That model is presented here.

During the research process, the conceptual model helps to define the problem domains and raise the main issues that should be addressed and investigated by the researchers (Gottesdiener, 2002); see Paper I.

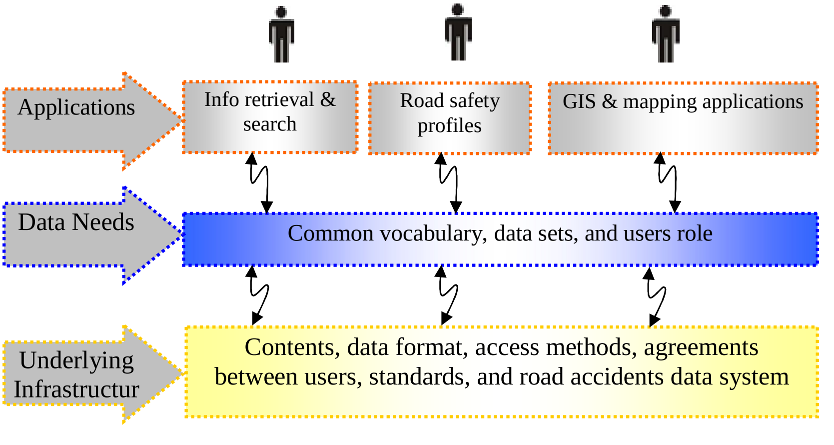


Figure 33 Suggested conceptual model.

During the systems development, conceptual modeling is a phase concerned with different system views. Difficulties may be caused by the use of traditional conceptual modeling based on descriptive properties of information systems (Holm, 1996).

The role of the conceptual model in the information systems development can be summarized as follows:

- Provides means for communication between model developers and the end users of the system.
- Increases the system analyst understanding of the problem domain and all system components.
- Serves as the basis for the design phase.

A conceptual model, Figure 33, for information sharing among road safety organizations was proposed. The model served as a conceptual design to develop and implement (GLOBESAFE) a web-based system for information sharing in road safety.

In the conceptual model, which is three-tiered, users, as external entities, interact with the model. The bottom layer comprises the underlying infrastructure (contents of the databases, data input/output formats, access control and methods, agreement and willingness between users to 'share', standards and the road accidents data system). The other two layers (applications and data needs) operate on top of this layer. This layer determines whether the model can lead to effective sharing of information among the road safety organizations.

On top of this layer is the data needs layer (common vocabulary, data sets, and users' different roles and privileges to operate and maintain the system).

Finally there is the third level, i.e. the application layer, in which a set of applications developed on top of the other two layers is defined. The interaction modes, user interfaces and access methods that exist between the users (external entities) and the three-layer model are defined as well.

Willingness of organizations to share information

The willingness of organizations to share the information and work in partnership is the cornerstone of a successful and effective information-sharing environment. In addition, the *data set* should be defined and agreed upon. Technical and administration procedures should be part of the agreement between all parties (stakeholders of the system).

The common vocabulary

The domain analysis stage was initially used as a starting point to develop the common vocabulary. The purpose of this stage is to learn more about the information needed for performing any kind of road safety analysis by the experts in the UoD. This helps to identify different aggregation levels of the data. Interviews and meetings with domain experts were organized to do that. The intention and main goal of this process was to determine basic terminologies, scope and use of the common vocabulary (ontology). The information was classified on four levels as shown in Table 4.

Table 4.

Knowledge levels in the road safety domain.

Knowledge level	Knowledge Properties
Base	Priorities, trends, policy progress & countermeasures
Intermediate	Pre-accident events
In-depth	Accident and injury causation
Specialist	Specific research questions

The basic entities that exist and, which form the information domains were identified as:

- **Human**

Persons involved in an accident (injured or killed): drivers, passengers, pedestrians, occupants, etc.

- **Road**

Different categories with different characteristics.

- **Vehicle**

Motor (motorized or non-motorized), number of wheels, and number and kind of occupants.

There are important issues in determining different and appropriate road safety measures for particular problems, as described below:

- **Factual information about accidents**

This includes, but is not limited to, data of accident-related information (time, day, month, year), the nature of the day (day before public holiday, public holiday, or popular vacation departure or return dates), location of the accident (urban area or outside urban area.), type of area where the accident occurred (city centre, industrial or business park) and the type of accident. The type of accident describes the conflict which results in the accident, for example, a driving accident, an accident caused by a vehicle turning off the road or an accident caused by a pedestrian crossing the road.

- **Interpreted information about accidents**

Accident severity was determined to find the number of fatalities, the number of injured and their types (occupants or pedestrians).

- **Factual information about non-accidents, exposure information, socioeconomics, and energy**

  This information was obtained from other resources such as population censuses and the country’s statistics.

After specifying the main entities and level of knowledge, domain experts drew the concepts in hierarchical form using CmapTools, a knowledge modeling kit (Cañas et al., 2004). Different models of conceptual maps were developed throughout the meeting with the experts. Analysis of the international database contents and reports also helped the process of developing the maps. In concept mapping processing, the concept in the domain is represented as a class and subclass in the hierarchy (see Figure 34. The relationships between the main and sub-classes were identified. The figures show conceptual maps produced for the main entities (classes). Figure 34 represents the classification of transport accidents. In general, there are at least four kind of transport accidents: aircraft accidents, road accidents, watercraft accidents, and railroad accidents. This is the top level of the concept map. Further on, road accidents were classified into three types according to the outcome of the accident, in terms of injuries: accidents with injured persons, fatal accidents and non-fatal accidents.

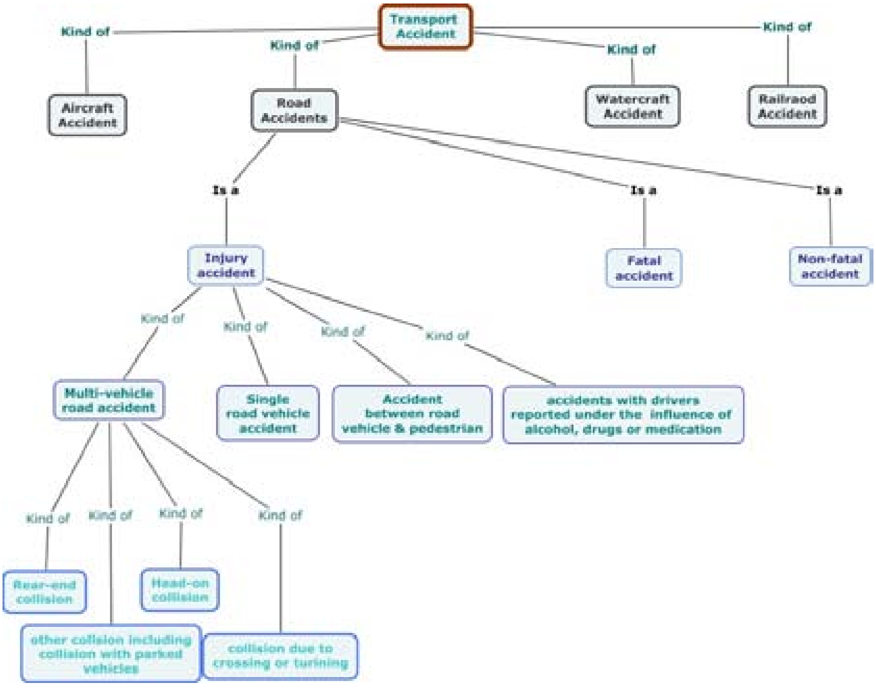


Figure 34. Kinds of transport accidents and the three types of road accidents that can be associated with them.

Figure 35 represents a concept map for the road categories. In this map four categories were represented, including motorways, urban roads, carriageways, and tramways. This is considered a kind of transport network. More maps were developed for each of the four categories.

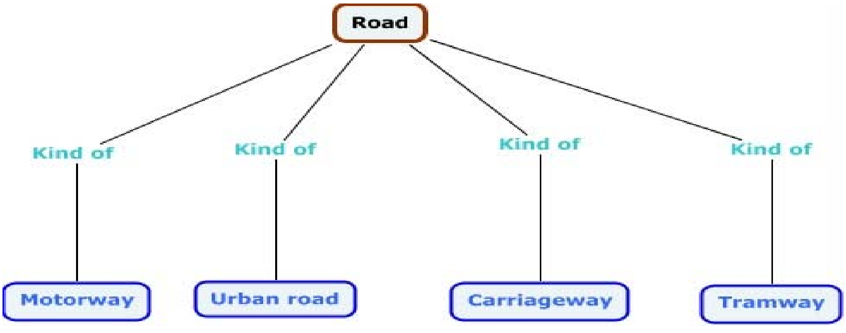


Figure 35. Concept map for road types in transport networks.

The concept map for the third entity is shown in Figure 36. That is, the Person class, which is divided into six categories:

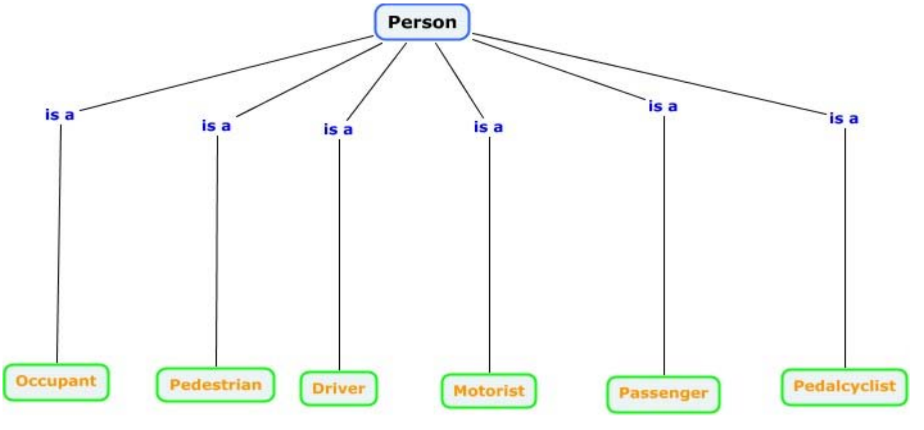


Figure 36. Conceptual map for person.

Coding the Knowledge

The last step is to encode all of the concept maps into knowledge modeling tools. This was done by creating a new project in the Protégé ontology editor. The user interface of Protégé facilitated the creation of the classes, slots, facets, instances, and frames, visually. A THING is a system class in Protégé that is automatically created when starting a new project. The model developed consists of 74 classes as shown in Figure 37.

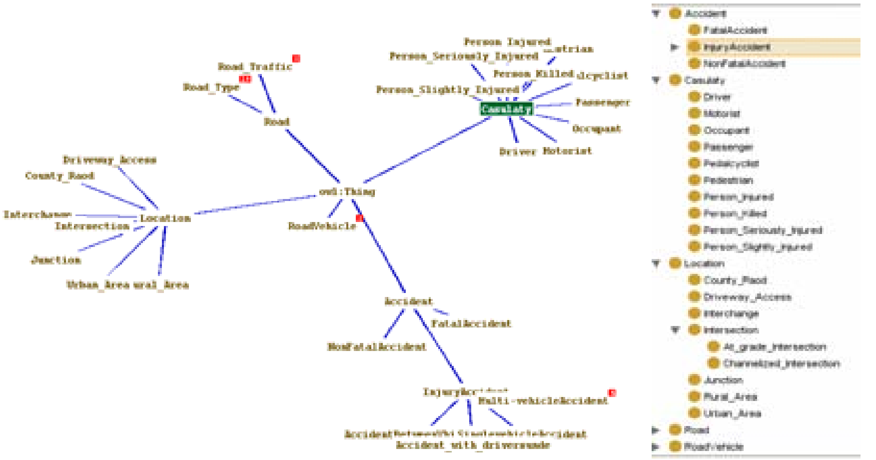
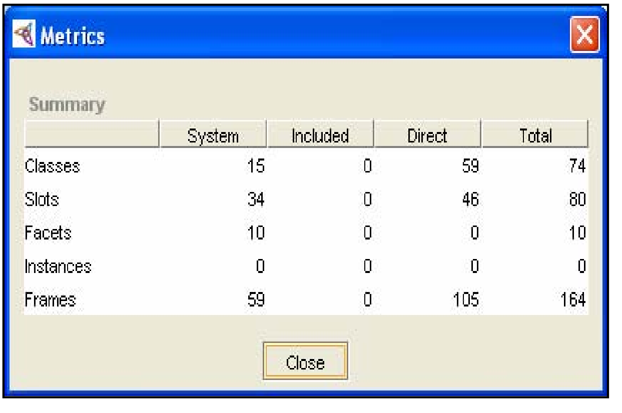


Figure 37. Class graph obtained from TGVizTab and the class hierarchy.

The result ontology is summarized by the report tools of Protégé as in Figure 38.



Metrics

Summary

	System	Included	Direct	Total
Classes	15	0	59	74
Slots	34	0	46	80
Facets	10	0	0	10
Instances	0	0	0	0
Frames	59	0	105	164

Close

Figure 38. Ontology metrics, showing classes, slots, facets, instances, and frames.

6.2. GLOBESAFE

The main purpose of the GLOBESAFE platform is to provide a powerful environment that makes it possible to share road safety information between the users. The basic requirements of such a platform are the ability to evaluate road safety programs and the facility to share knowledge and experience among the members of the road safety community. The front-end structure of GLOBESAFE shows the basic components and different applications developed, Figure 39.

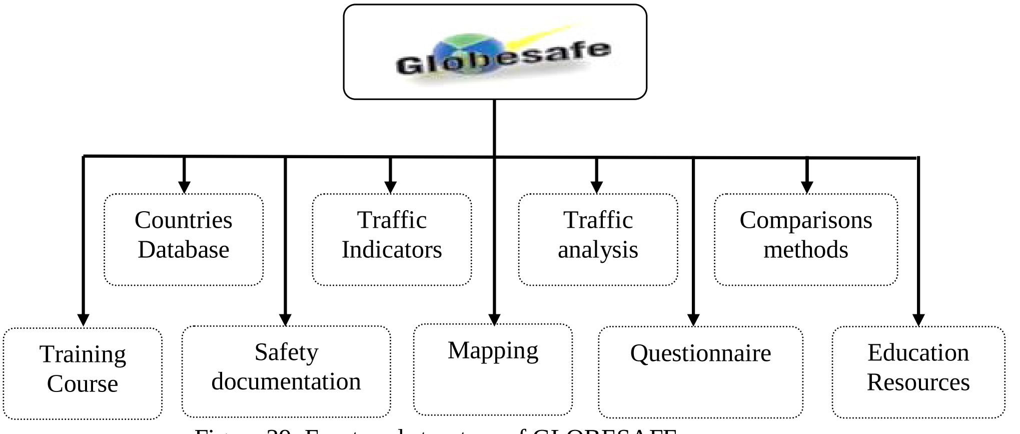


Figure 39. Front-end structure of GLOBESAFE.

Early in Chapter 1 the different methods currently used by the road safety community to share their information were discussed. A system with special functionalities should be developed to handle and meet the user's needs. In Papers IV and III, it was noted that the road safety community searched and retrieved data from different sources, and that different strategies were used to access that information (e.g. annual subscription, purchase). Prototyping methods were applied in the development of GLOBESAFE.

Special properties of GLOBESAFE

- **Harmonization of the data**

Road accidents and injury data are collected and stored by police and health institutes (e.g. hospital inpatient records, emergency room records, ambulance and emergency records).

The data from the police and the health institutes need to be accessed and matched. This ensured that there was no redundancy in the data. Both police and health institutes store disaggregated data at the single accident level.

In the case of aggregated data sources, where there are a number of data sources available, it is important that the data sources should be matched to see the similarities and differences between these sources. In GLOBESAFE, four international databases have been matched. These databases host road safety data from all continents. The databases are as follows:

- Asia-Pacific Road Accident Database (APRAD)
- The Asian Development Bank (ADB)
- Organization for Economic Co-operation and Development (OECD)
- World Health Organization (WHO)

Each source has unique features and covers specific geographical regions. During the design of GLOBESAFE all of these differences were considered. The direct result of this is to reduce the users' efforts in accessing different sources. Instead the search will be of a single source; this, of course, will reduce both time and effort.

- **Definitions and standardization of data**

During this work, it was noticed that different road accident database providers have different definitions of similar attributes in their databases. One regional database and two national databases were studied. One term was used for comparison to see how it was defined.

- Community Road Accident Database

CARE is a European Union database, a disaggregated road accident database, which stores data from EU member countries.

- Traffic Bureau, National Police Agency, Japan

This is a national database for Japan in which details of accidents that occur in the country are stored. The traffic bureau in Japan is responsible for its accident statistics.

- American National Standard — Manual on Classification of Motor Vehicle Traffic Accidents, "FARS".

This is the ANSI classification of road accidents. It defines and classifies all types of accidents that may occur and are reported in the USA.

As an example, the term **accident** has three slightly different definitions in the above organizations, as outlined in Table 5.

Table 5.

Different definitions of the term "accident"

CARE	Traffic Bureau, National Police Agency, Japan	FARS
Occurs on a public road or on a private road to which the public has right of access. Involves at least one moving vehicle. Involves at least one injured or killed person. Is reported by the police.	An accident resulting in death and/or injury, which is caused by the traffic of vehicles or streetcars running on a road.	A transport accident that (1) involves a motor vehicle in transport, (2) is not an aircraft accident or watercraft accident, and (3) does not include any harmful event involving a railway train in transport prior to involvement of a motor vehicle in transport.

Another example, the definition of "killed" in a certain accident is shown in Table 6.

Table 6.

The definition of "killed"

CARE	Traffic Bureau, National Police Agency, Japan
Death within 30 days of a road accident	Died within 24 hours as a result of an accident

The previous examples demonstrate the need to standardize accident registrations and definitions of different terms in the databases.

The approach to overcome these issues in GLOBESAFE was to use a mathematical solution to handle the differences. In the case of “number killed” GLOBESAFE used a correction factor to adapt country data to the 30-day definition of CARE. Also, a common vocabulary using frame-ontology was built to serve as a standard for future development and for database design.

Data accuracy and consistency

Accurate data are essential for prioritizing public health issues, monitoring trends and assessing intervention programs. Many countries have inadequate information systems on road traffic injuries, making it difficult to realize the accuracy of the data they offer. The sources of the data are not well defined in some databases. Attention was paid to accuracy issues at an early stage of this study.

Data accuracy refers to whether the data correctly represent the reality or the occurrence of certain events (O'Leary, 1993).

The study of the international databases showed that the variables collected differ in terms of availability. This investigation resulted in the list of attributes in GLOBESAFE database schema.

A categorization of the variables was made and the variables were classified into different groups. Emphasis was given to the variables that contributed to the process of safety planning and auditing. The following were the main groups:

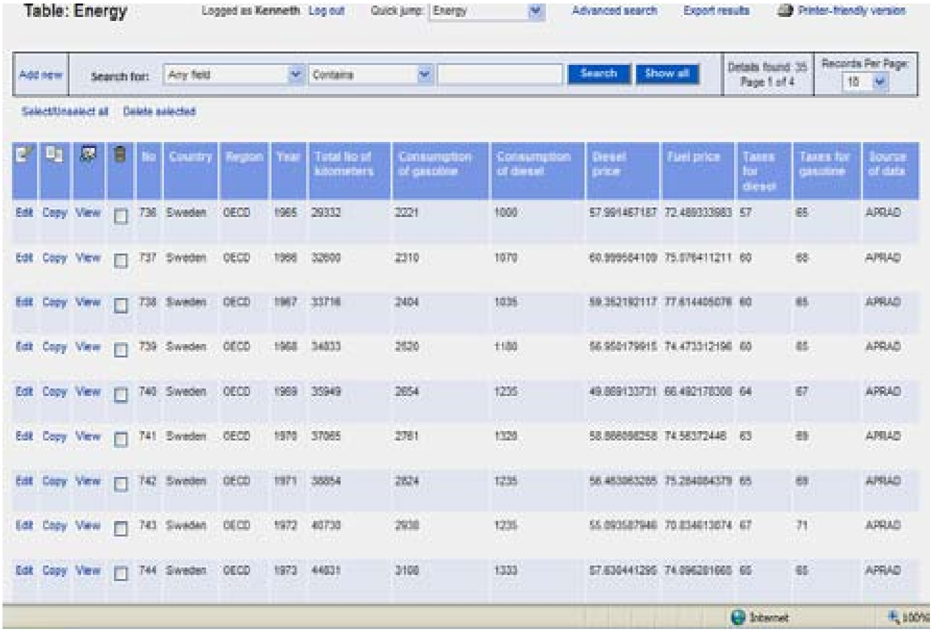
- Accident-related variables
- Demographic variables, divided into age and type groups
- Infrastructure, economic and energy-related variables

Data accuracy is a major challenge to make data analysis easier for comparison of data from different countries and to contribute to the general data quality. Integrity constraints in the database design (primary and secondary keys) were enforced to ensure data accuracy.

User interface formats (input and output)

The results of the study and the comparison of the different international databases led to the design of new input formats that can be recognized by users from different organizations and countries.

The structure of the data set should handle the different formats currently available in different countries (see Chapter 2). A defined format based on Graphical User Interface and InfoViz made it easier to handle the data from the different sources. This format is required to handle all operations a user might need to perform. Figure 40 shows one of the data input formats for energy related variables.



				No	Country	Region	Year	Total No of kilometers	Consumption of gasoline	Consumption of diesel	Diesel price	Fuel price	Taxes for diesel	Taxes for gasoline	Source of data
Edit	Copy	View	☐	736	Sweden	OECD	1965	29332	2221	1000	57.991467187	72.489333983	57	65	APRAD
Edit	Copy	View	☐	737	Sweden	OECD	1966	32600	2310	1070	60.999584109	75.076411211	60	68	APRAD
Edit	Copy	View	☐	738	Sweden	OECD	1967	33716	2404	1035	59.352192117	77.614405076	60	65	APRAD
Edit	Copy	View	☐	739	Sweden	OECD	1968	34833	2520	1180	56.950179915	74.473312196	60	65	APRAD
Edit	Copy	View	☐	740	Sweden	OECD	1969	35949	2654	1235	49.869133731	66.492178308	64	67	APRAD
Edit	Copy	View	☐	741	Sweden	OECD	1970	37065	2781	1328	58.866898258	74.58372446	63	69	APRAD
Edit	Copy	View	☐	742	Sweden	OECD	1971	38854	2824	1235	56.463063265	75.284084379	65	69	APRAD
Edit	Copy	View	☐	743	Sweden	OECD	1972	40730	2938	1235	55.093587946	70.834613074	67	71	APRAD
Edit	Copy	View	☐	744	Sweden	OECD	1973	44631	3108	1333	57.630441295	74.096281665	65	65	APRAD

Figure 40. An example of some data input formats.

Access methods: Role-Based Access Control (RBAC)

A critical principle to be considered is the protection of accident and injury statistics against unauthorized access and malicious attacks. To achieve this, the Role-Based Access Control (RBAC) was used (Park et al., 2001).

Since GLOBESAFE will be web-based, sensitive information needs to be protected against any damage caused by unauthorized access. RBAC technology was originally introduced to manage and enforce the security in large-scale, enterprise-wide systems (Coyne, 1996). The basic idea behind RBAC is that the permissions of access are associated with specific user's role, and then the specific users are assigned appropriate roles. In GLOBESAFE, RBAC was applied for strong protection of the shared information (see Figure 41).

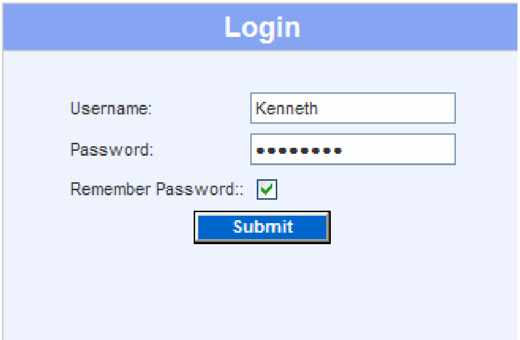


Figure 41. Appling role-based access control.

Applications

A set of applications was suggested and implemented in GLOBESAFE, which aimed to support the following tasks:

- Raising political awareness about road safety issues
- Enabling a detailed analysis
- Providing a basis for countermeasures
- International and regional benchmarking
- The results will be stored as new values in the database for further analysis

Information search and retrieval

Searching the content of the database to retrieve a set of data is an important user requirement. Different search criteria are supported in GLOBESAFE, using any of the database fields as search parameters. The final output to the users can be handled in different formats. Among those formats is XML. XML supports exchange of data between different systems and is a worldwide standard. GLOBESAFE provides two types of search. The first one is simple, using a single key with multiple criteria for the search; Figures 42 & 43.

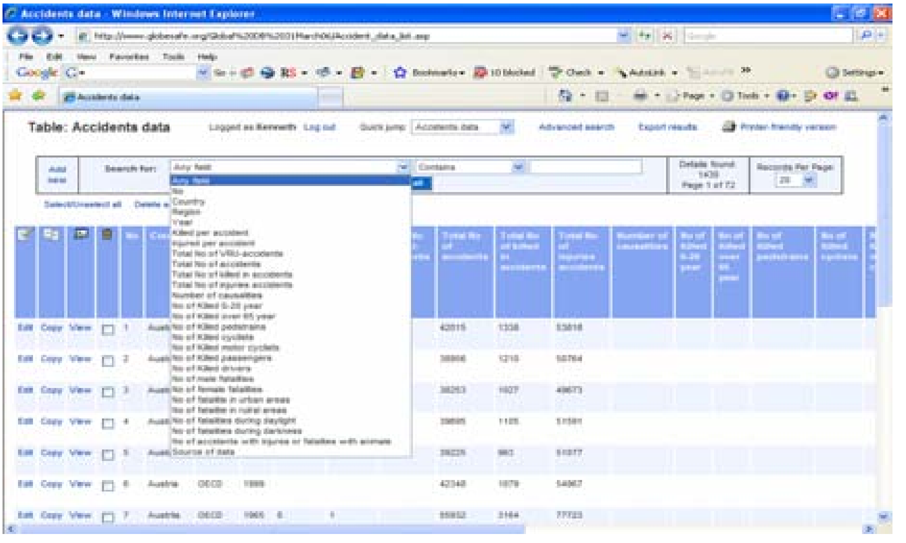


Figure 42. Single key search using different criteria.

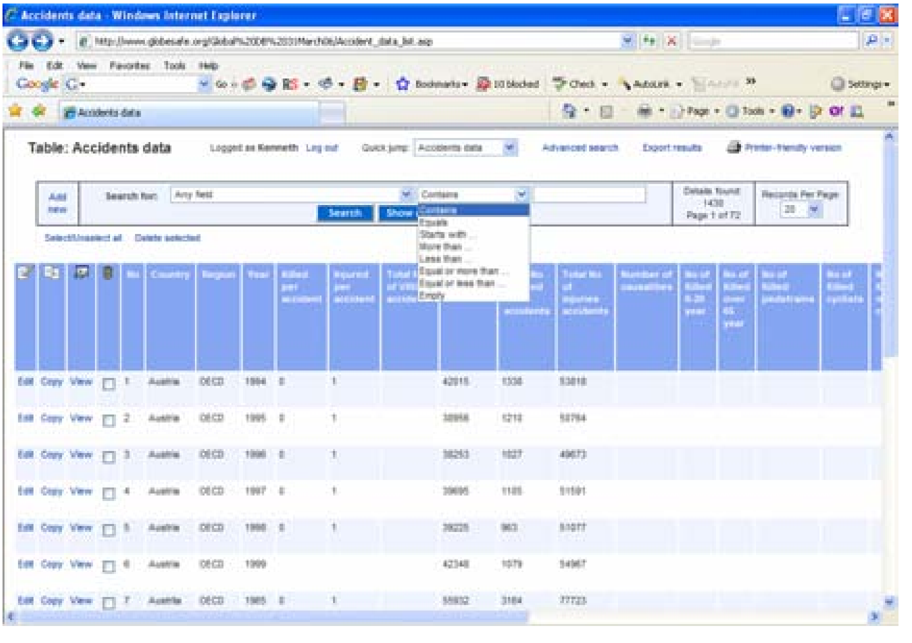


Figure 43. Single key search using different criteria.

An advanced search gives users the ability to combine more than one search parameter with logical operators (see Figure 44).

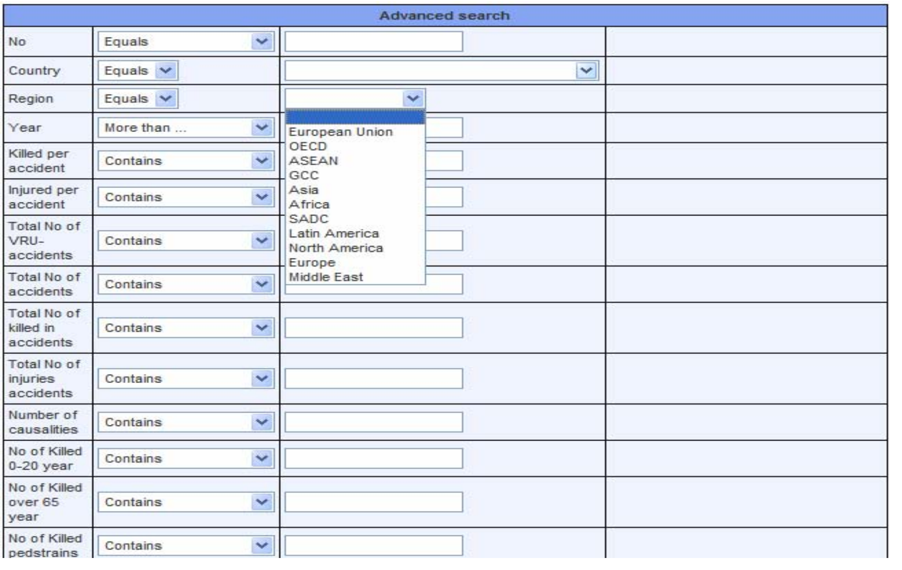


Figure 44. Advanced search with multiple keys.

6.3. Road Safety Profile: a diagnostic tool for road safety analysis

Road Safety Profile (RSP) is an example of an application built on top of GLOBESAFE. This application was required by the Global Road Safety Partnership (GRSP), an international organization working in the field of road safety, following a partnership approach. The organization has a group of focus countries on different continents. RSP measures the effect and implementation of the programs funded by GRSP. The main purpose of RSP is to:

- Monitor road safety status and development
- Analyze the road safety situation
- Benchmark road safety
- Identify weak and strong sectors
- Provide a basis for suggestions of countermeasures

RSP is based on two groups of variables: direct measurement (statistical data) and indirect measurement (performance indicators).

RSP uses a group of performance indicators such as road user behavior, organizational structure, traffic legislation, road quality, health care and rescue services on the road. To create the RSP, the user needs to specify the motorization level (Figure 45). Three motorization classes used in RSP are:

- Low motorization (LM): countries with 100 or less motor vehicles per 1,000 inhabitants.
- Middle motorization (MM): countries with more than 100 but less than 301 motor vehicles per 1,000 inhabitants.
- High motorization (HM): countries with more than 301 or more motor vehicles per 1,000 inhabitants.

A number of traffic indicators are used in the RSP and also appear under the traffic analysis part of GLOBESAFE. Table 7 lists these indicators with their formula in Table 7.

Table 7.

Country traffic indicators

	Indicator	Formula
1	Number of killed / 100,000 inhabitants	Total number of killed divided by the sum of total population number over 100,000
2	Number of motor vehicles/ 1,000 inhabitants.	Total number of motor vehicles divided by the sum of total number of country's population over 1,000
3	Number of road users killed /motor vehicles	a + b + c + d + e/ total number of registered vehicles a. Number killed – pedestrians b. Number killed – cyclists c. Number killed – motor cyclists d. Number killed – passengers e. Number killed – drivers
4	Number VRU killed /motorcycles	a + b + c + d + e/ total Number of motorcycles a. Number killed – pedestrians b. Number killed – cyclists c. Number killed – motor cyclists d. Number killed – passengers e. Number killed – drivers
5	Number of pedestrians killed per motor vehicle	a / b a. Number killed – pedestrians b. Total number of registered vehicles
6	Number of pedestrians killed per motorcycle	a / b a. Number of killed – pedestrians b. Total number of motorcycles
7	Number of physicians per 1000 inhabitants	a / b *1000. a. Total number of physicians b. Total population
8	Personal Risk	a / b a. Fatalities b. Population
9	Traffic Risk	a *10,000 / b a. Fatalities b. Vehicles
10	Severity Index	a * 100 / (b + a) a. Fatalities b. Injuries
11	Number of vehicles/killed	Number Vehicle/ Fatalities

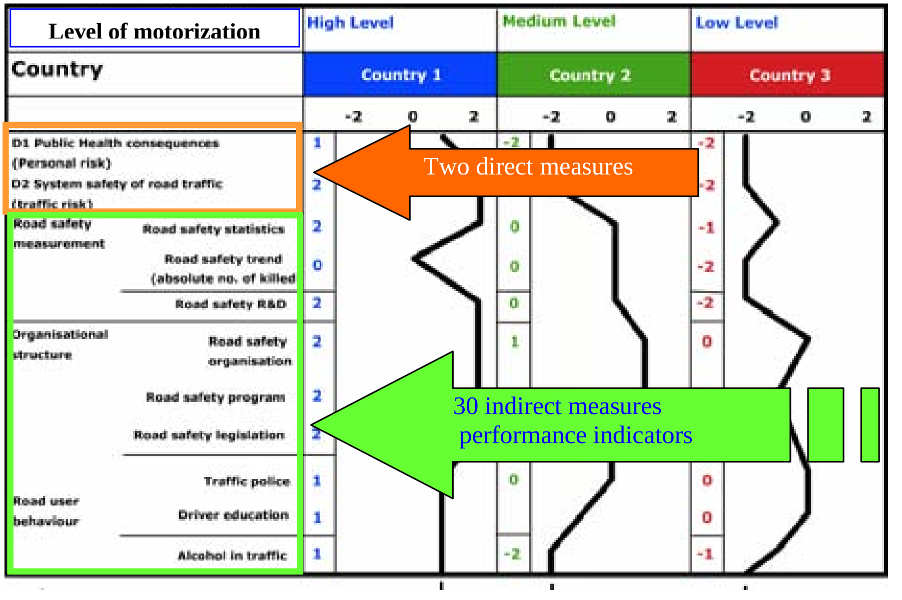


Figure 45. Elements of the Road Safety Profile.

The RSP presents an overall view of the specific indicators. This helps to identify the size of the problem in each sector using scaling (2, 1, 0, -1, -2). An example of the output of the RSP is shown in Figure 46.

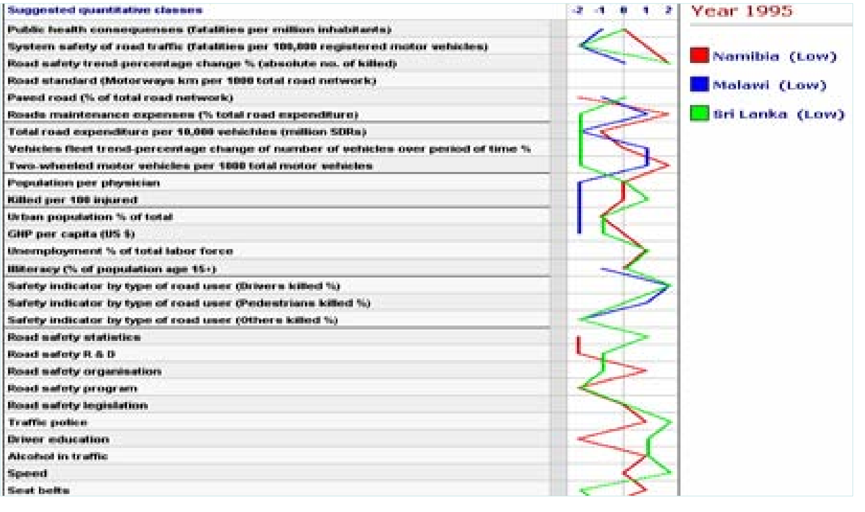


Figure 46. An example of RSP for three countries.

6.4. Web mapping application: Map as an interface for shared information

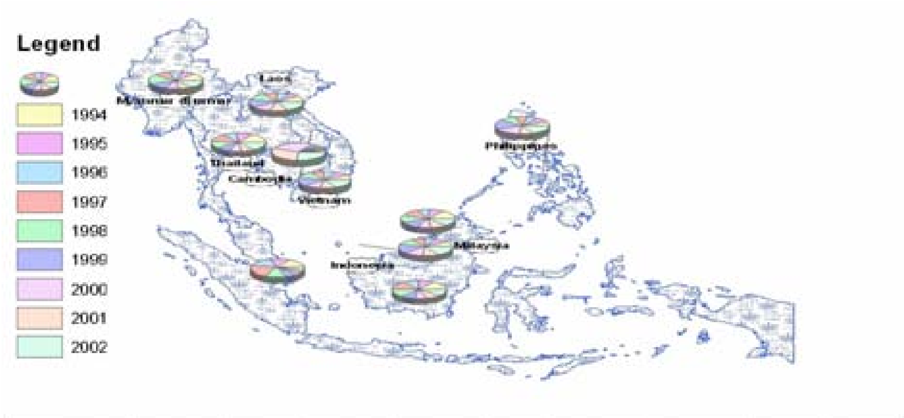
Internet GIS methods were investigated. This was to use a map as an interface for shared information in GLOBESAFE. The web-based mapping tool, as introduced here, represents geographical features from vector maps. Attribute data was linked to those features. The attribute data from different countries were collected and stored in the database (see Papers II and III).

Design principles for web-based mapping were explored as a user interface to access road safety databases. This application helps to improve access to the data stored in database tables. Moreover, it helps the user visualize road safety situations using the map rather than conventional reports.

The implementation of this tool was based on a portrayal model for interactive mapping developed by Open Geospatial Consortium (OGC), a Special Interest Group (SIG) in web mapping. The final output map is given in Figures 47 and 48.

Brunei	Cambodia	Indonesia	Laos	Malaysia	Burma	Philippines	Singapore	Thailand	Vietnam
461	2056	94723	3022	52786	10470	8570	1986	121174	65124

Figure 47. Number of persons killed in ASEAN countries 1994-2002.



Brunei	Cambodia	Indonesia	Laos	Malaysia	Burma	Philippines	Singapore	Thailand	Vietnam
2553	22100	131944	32341	1948677	42328	80687	444397	752365	182956

Figure 48. Total number of accidents in ASEAN countries 1994-2002.

The mapping tool offered different cartographic operations such as zoom in, zoom out and panning.

The main feature of the mapping tool is that key processing is performed by the client and no special software is required. The output maps are in Scalable Vector Graphic (SVG) format. SVG is based on XML and it produces maps with high cartographic quality and map visibility. The tool is appreciated by the user, as seen in the user study and evaluation in section 6.5

6.5. System evaluation: User interaction study

The main objective of the usability and acceptance test in Paper III is to measure user satisfaction and acceptance towards the user interface. Two scenario-based testing sessions were conducted with a group of nine respondents. Each scenario represented a group of tasks designed following the guidelines given by Dumas and Redish (1993), as discussed in Chapter 5. Following the guidelines of Nielsen (2000), nine respondents participated in this study. They had different study and research backgrounds.
The two sessions were conducted in different locations. One of the sessions was held at the GIS Lab at the Department of Computer and Information Science (IDA), Linköping University. The respondents were introduced to the operations provided by the map module. Any questions they had were addressed. The respondents were asked to perform specific tasks individually and then answer the questionnaire. The second session was organized at the Department of Science and Technology (ITN), Linköping University. A group of Traffic Safety and Environment experts participated. Similar procedures as in

the first session were followed. The participants from the second session were familiar with GLOBESAFE, as they had been interviewed during the prototyping process.

Questionnaire Design

The questionnaire was designed in a way that made it easier for the users to select appropriate choices and suitable answers from the different options. The first part of the questionnaire aimed at collecting demographic information about the respondents and their background.
As shown in Figure 49, only four of the respondents were working with any road safety database, either for research purposes or as a job assignment.

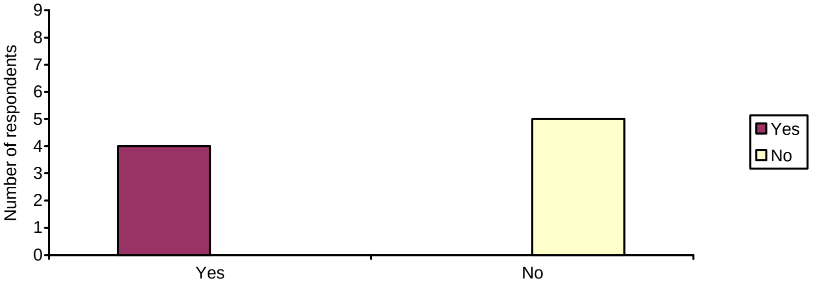


Figure 49. Respondents experience with road safety database.

In the second part of the questionnaire the visibility of the map elements was tested. The respondents were asked to rate the visibility of each element. The visibility of the map elements was considered acceptable by eight of the respondents, rating map elements visibility as good. Also, the appearance and position of the map were rated as good and very good by several of the users (Paper III and Fig. 50).

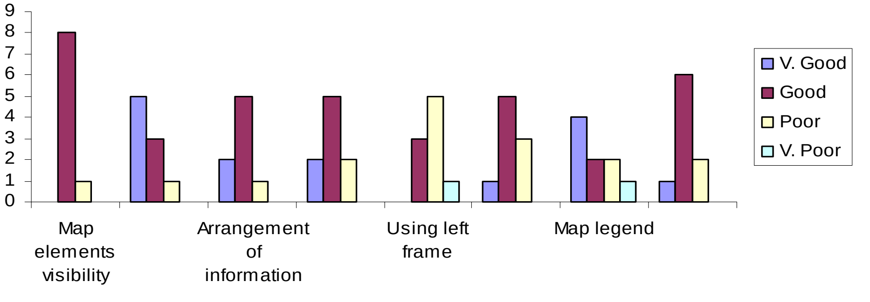


Figure 50. Map elements visibility.

Cartographic operations

This part of the questionnaire was designed to measure the respondents' feedback concerning cartographic operations (zoom in, zoom out and panning). Most of the respondents were occasional users of web maps. Five of them appreciated the response time of the cartographic operations while the other four were not satisfied (Figure 51). It should be noted that during one of the sessions, one of the respondents compared the response to other web mapping approaches and found that the response time of GLOBESAFE was much faster.

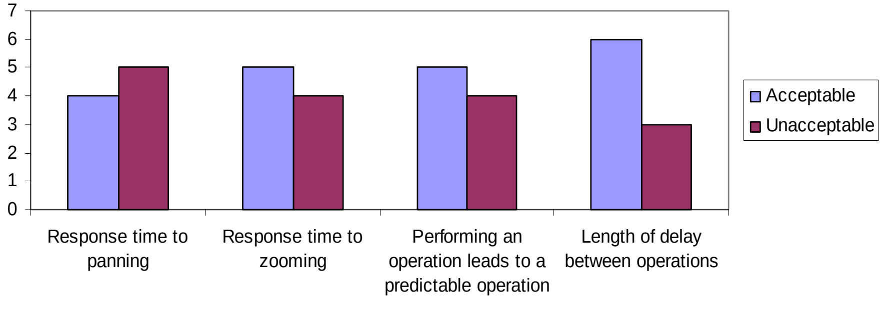


Figure 51. Response to cartographic operations.

Navigation between the maps and other web pages

The fourth part of the questionnaire was designed to measure the navigation capabilities and how the respondent felt when navigating between the maps and the different pages in GLOBESAFE. As part of the navigation, the users needed to switch between the map layers. Figure 52 shows that six of the respondents rated this property as an "easy" task.

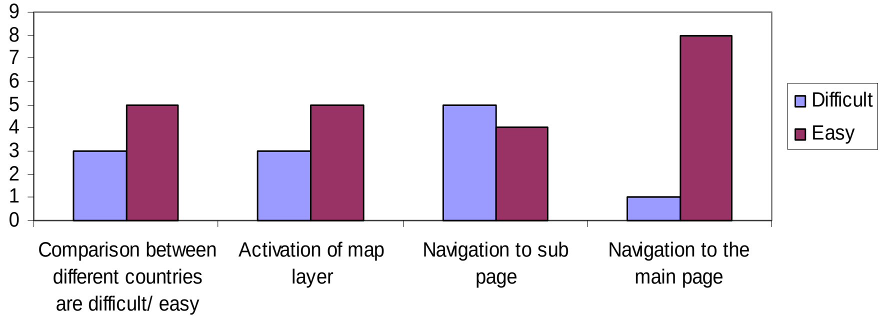


Figure 52. Working with map layers.

These results confirm the acceptance of the GLOBESAFE mapping module. Information Architecture (IA) and Visual Hierarchy (VH) principles, added an explicit design model to the web mapping. A special property of GLOBESAFE is using a map as an interface for shared information. Feedback from users showed that the visibility of the different road safety statistics and indicators are easy to interpret. This is one of the user requirements met during the study (Figure 53).

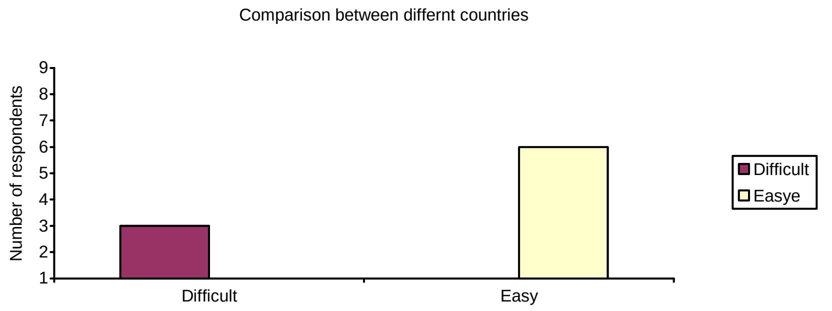


Figure 53. Comparison of countries.

6.6. Results of the STRADA study

In Paper VI, two approaches were used to study STRADA. This system is used as a road accident reporting system. The study focused on the acceptance of the system and investigated under-reporting on STRADA.

Under-reporting refers to situations in which the database holds a total number of accidents less than the actual number of accidents. In many road accident databases there are under-reporting of road accident of deaths and injuries (Hvoslef, 2001). Transport Research Laboratory (TRL) has estimated that only 50% of the road injuries were reported in road databases.
The study of the STRADA database for the period 2002-2005 shows that 60.7% of the accidents are reported by the hospitals which reported to STRADA. On the other hand, 20.5% of accidents are reported by the police authorities. This leads to significant variations in the total number of accidents reported (Figure 21).
In STRADA, similar accidents have a high matching ratio and appear on the same date, time and location. This implies that the accidents were reported by the police and the hospital at the same time and it involves injuries; and also that in that region both the hospital and the police accept the use of the system. These types of accidents are shown in Figure 19 as the intersected area. Matched accidents represent only 18.7% of the total number of accidents in the system, whereas some accidents are neither reported by the

police nor by the hospitals. These accidents are represented by the outer circle of Figure 54.

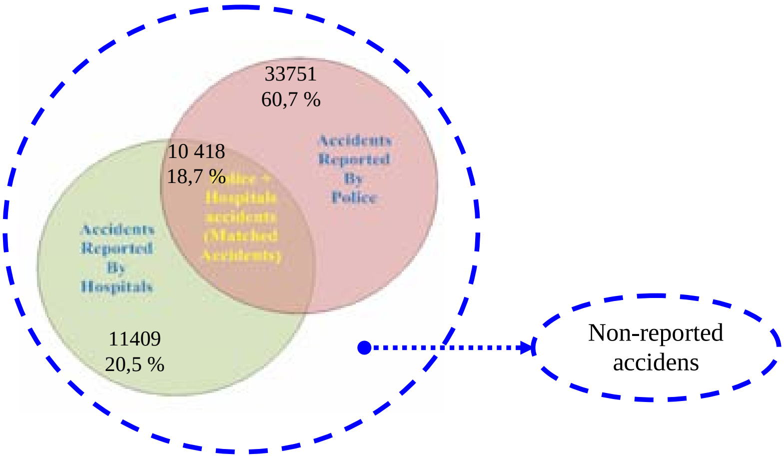


Figure 54. Reported and non-reported accidents.

System acceptance

At the time of the study of STRADA, in early 2007, not all hospitals in Sweden agreed to use and report accidents to STRADA. Hospitals in 10 of the 21 counties in Sweden agreed to report accidents to STRADA (less than 50%); see Paper IV. This has resulted in under-reporting of the total number of accidents. We used Technology Acceptance Model (TAM) (Davis, 1989) to understand the reason behind the low acceptance of the system (see Figure 55).

As was discussed in Chapter 5, perceived usefulness (PU) reflects the extent to which the users believe that the system will help to improve the performance of the organization. In the case of STRADA, PU related to the operability of the system with the existing system used at the emergency unit to report emergency cases.

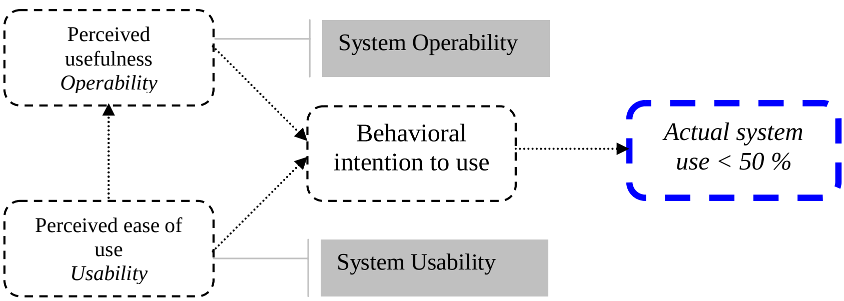


Figure 55. Appling TAM to STRADA, incorporating system usability and operability.

The aim of the study of the system was to identify the reasons that make the actual system acceptance less than 50%. The respondents during the focus group meeting and interviews commented that STRADA has a complexity that makes it difficult to use, requiring extra resources and training to allow them to use it.

The system is composed of different components and interfaces and the users need to be well trained to be able to use and perform the basic operations. This implies that the system is not simple to use or easy to learn. The consequences of poor usability are linked to the "Perceived Ease of Use" in TAM, which reduces the actual system use at the hospitals. The reason behind the poor usability of the system is the fact that STRADA developers did not pay enough attention to the usability of the system, regarding it as a non-functional requirement (NFR; Sommerville and Sawyer, 1997).

Telling the system story by link, loop and delay

Changes in the system usability and the trade-off between usability and complexity issues can be interpreted by using concepts from system thinking theory. Links, loops and delay help to see the whole picture and the patterns of changes. The guidelines in Senge (1994) were used to construct the model in Figure 56.

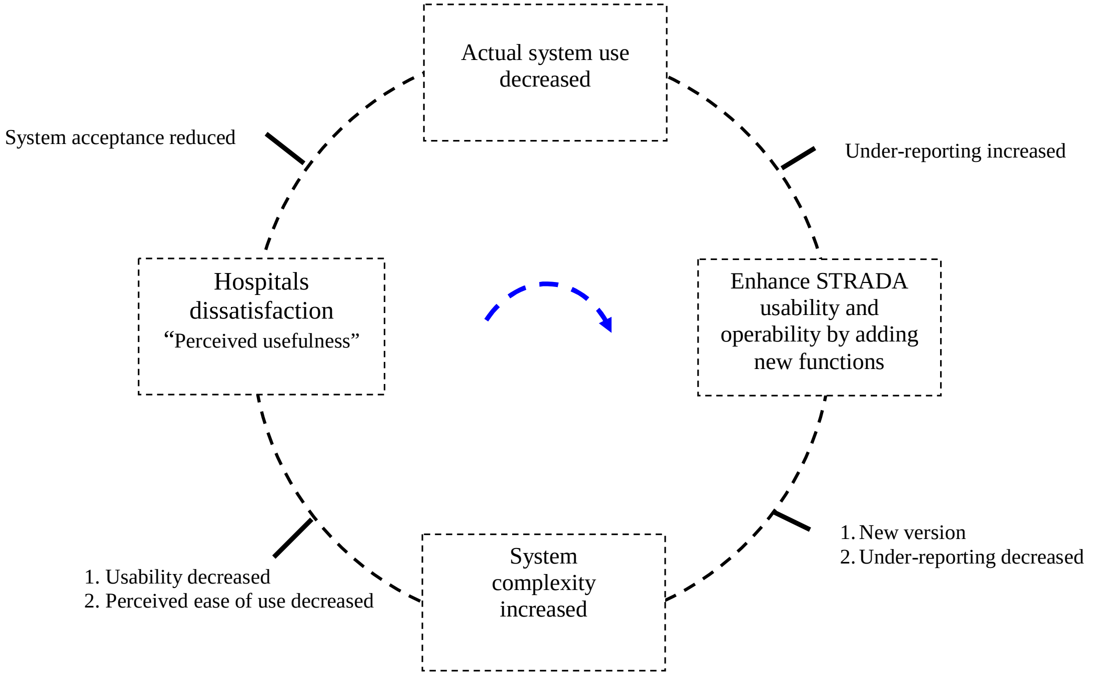


Figure 56. Telling the system story by link, loop and delay model.

Figure 56 shows that usability decreases as the complexity of a system increases (Sikora and Swan, 1998). The main source of complexity in the system came from attempts to increase the perceived usefulness, by adding more functions. These additional features resulted in a more complex system in terms of its interface and additional subsystems.
The model shows that increases in the usability (perceived ease of use) and perceived usefulness will encourage more hospitals to use the system, which in turn will increase the system complexity. This, of course, will affect the service provider's choice of balance between usability and complexity. Figure 56 shows the relationship between system acceptance, under-reporting (as a result of rejection of the use of the system by hospitals) and usability.

Recommendations for future development of STRADA

The following solutions are suggested to improve STRADA and make it usable for the rest of the hospitals in the country.

- **Get stakeholders involvement: Specification changes**

The commitment of the hospitals to accept the system is the first step to increasing actual system use. This could be achieved by letting the hospital authorities and real users of the system (nurses) become involved in the development of system requirements. This will guarantee that the system will satisfy user needs and get acceptance of the system.

- **Life cycle change: Stages execution mode**

During the concepts stage (life cycle starts with the concept stage — to identify the system specification and user's needs — and ends with the utilization stage), the needs of the stakeholders should be converted to specifications.

The developer(s) could introduce an *enabling system* (contribute to the achievement of the goals of various stages as the system moves through its life cycle). A system such as QUIS (Questionnaire for User Interaction Satisfaction) would be of good use to the developers and to the user groups.

- **Interactive and iterative development**

The main advantage of using an iterative model is that it helps to assess system requirements. Models such as the Spiral Model from software engineering feature verification and validation processes through each cycle (Blum, 1992), which ensures that the requirements are verified and validated.

6.7. Application of visual data mining

Chapter 3 introduced automatic and visual data mining methods and techniques. In Paper VI, the methods and techniques of Visual Data Mining (VDM) were investigated. VDM techniques provide a potential approach to explore the hidden knowledge in two data sets, IRTAD and GLOBESAFE. The interaction between the users and the VDM tools enables the users to visually extract more knowledge about the relationships between the data elements. Another feature of the visualization techniques is that the users can classify and identify the problems; the VDM paradigm described in Chapter 5 was used for this purpose. In this section results obtained by using Orange and Tanagra as tools to support VDM are presented the rest of the results are illustrated and presented in Paper VI.

Classification and decision trees

The knowledge flow was constructed in the Orange environment to apply decision and classification tree algorithms. A knowledge flow through different components is represented in Figure 57. We selected a group of fields to apply decision tree (DT) algorithms (the total populations of the country, the total number of killed in accidents, the total number of vehicles and number of injuries).

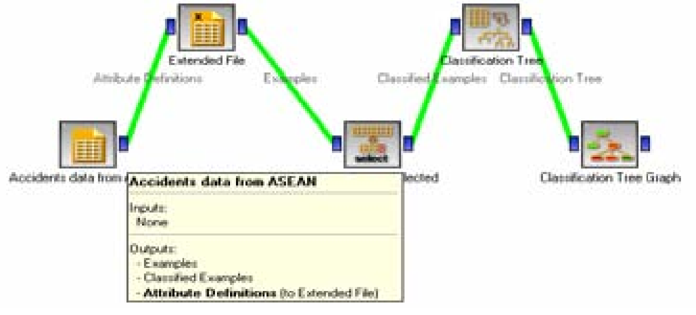


Figure 57. Knowledge flow to apply classification and decision tree algorithms.

The DT is shown in Figures 58 and 59. The DT graph is composed of 17 nodes, 9 of which are leaves, and the depth of the tree is 5 levels. The tree shows the majority class probability and the target class in this tree is Malaysia. A pie chart in each mode represents the number of instances in each sub-tree. At leaf nodes it can be observed that the pie chart has a single color. Each node is colored according to the majority class probability. A classification tree in Figure 26 represents the output of the classification in tabular form.

The right hand sub-tree consists of four countries sharing the same features; those are the Philippines, Thailand, Vietnam, and Indonesia, i.e., the safety situations in these countries are closer to each other than to the countries in the left hand sub-tree (Cambodia, Malaysia, Myanmar, Laos, and Singapore).

For this data set, however, the DT obtained is smaller compared to other data sets, which in some experiments gave trees ten times the size of this one. The larger the data set is, the larger the DT.

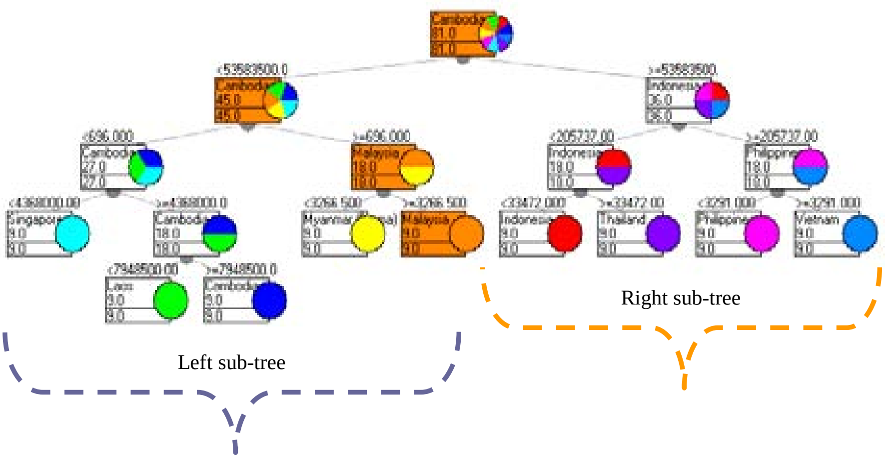


Figure 58. DT graph with Malaysia as target variable.

Classification Tree	Class	P(Class)	P(Target)	#Inst
<root>	Cambodia	11	11	81
Population <53583500.000	Cambodia	20	20	45
killedintraffic <696.000	Cambodia	33	33	27
Population <4368000.000	Singapore	100	0	9
Population >=4368000.000	Cambodia	50	50	18
Population <7948500.000	Laos	100	0	9
Population >=7948500.000	Cambodia	100	100	9
killedintraffic >=696.000	Malaysia	50	0	18
killedintraffic <3266.500	Myanmar (Burma)	100	0	9
killedintraffic >=3266.500	Malaysia	100	0	9
Population >=53583500.000	Indonesia	25	0	36
NoofVehicles <205737.000	Indonesia	50	0	18
Injuriesinaccidents <33472.000	Indonesia	100	0	9
Injuriesinaccidents >=33472.000	Thailand	100	0	9
NoofVehicles >=205737.000	Philippines	50	0	18
killedintraffic <3291.000	Philippines	100	0	9
killedintraffic >=3291.000	Vietnam	100	0	9

Figure 59. Classification tree with Malaysia as target variable.

Linear projection

The linear projection in Figure 60 shows the projection of three attributes from the energy database table: (1) consumption of gasoline; (2) consumption of diesel; (3) total number of kilometers. The way the linear projection is interpreted, such that: All of the countries that have approximately the same or equal values for all three attributes lie close to the

origin point of the linear projection. In cases where one attribute value is larger than the values of the other attributes, the point will be far from the origin and towards the end point of the line.
The linear projection graph shows that the USA has higher consumption of gasoline. Japan and France have the highest values for consumption of diesel fuel. Interaction methods such as brushing and linking also help to expand the knowledge about a specific item or data set.

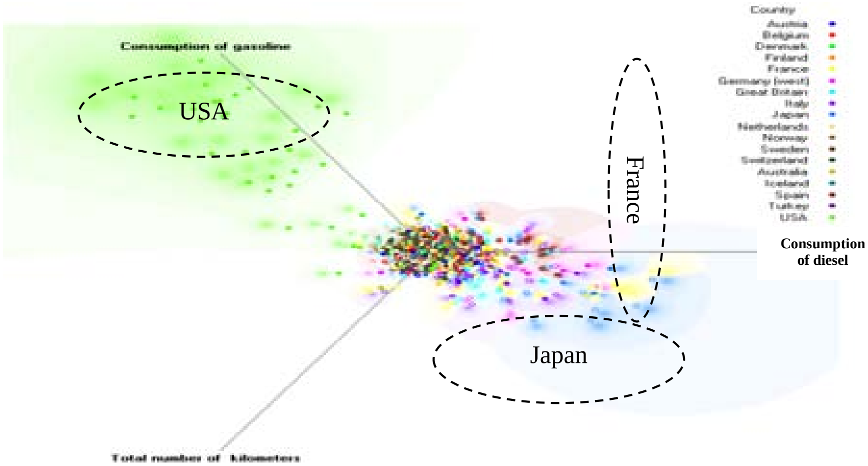


Figure 60. Linear projection of energy related variables.

This could be related to the fact that USA has the highest number of cars per capita among OCED countries (see Figure 61).

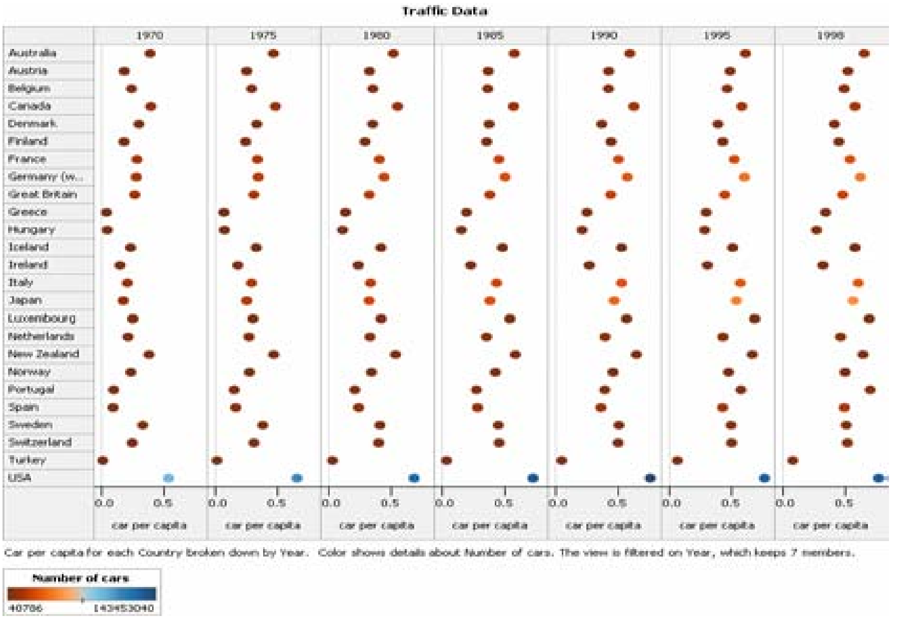


Figure 61. Car per capita for each country broken down by year 1970-1998.

Another example of applying linear projection is shown in Figure 62. Attributes from accident statistics include number killed in traffic, injuries in accidents, total number of vehicles, number of motorcycles, and total number of accidents. Linear projection can be summarized as shown in Table 8.

Table 8.

Interpretation of the results obtained by linear projection.

Country	Number killed in traffic	Injuries in accidents	Number of vehicles	Number of motorcycles	Number of accidents
Thailand		High			
Singapore					High
Indonesia	High			High	
Vietnam			High		

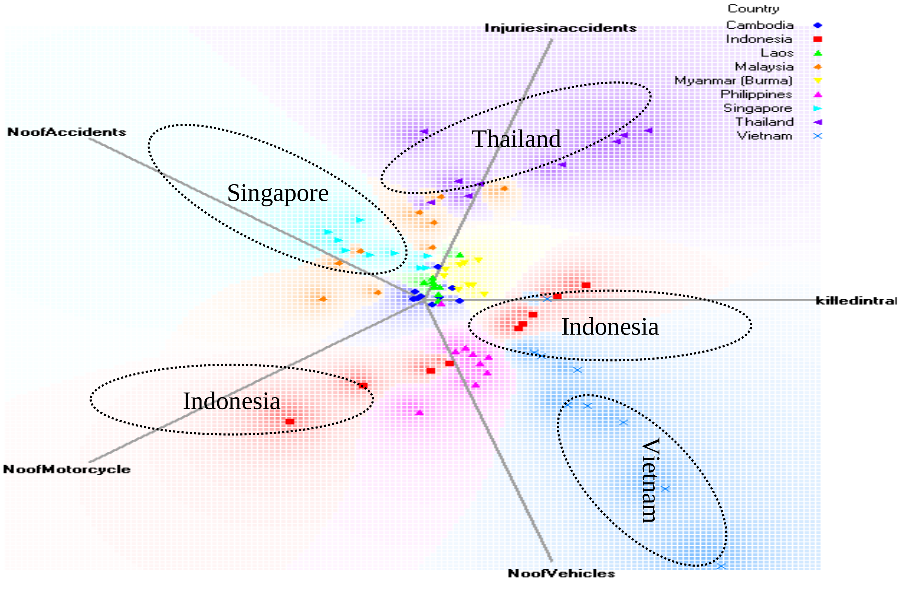


Figure 62. Linear projection with five axes for accident-related attributes.

K-means

Table 9 shows the results from the K-means clustering applied to the data set from ASEAN countries. The same variables as used in the previous example were selected. The mining task explored the main clusters in the data set. After five trials, the result for K-means is five clusters.

Table 9.

K-Means clustering results

K-Means: clustering results, Number of clusters and trail = 5			
Cluster	Countries	Size	Trial
cluster n°1	Malaysia	9	1
cluster n°2	Thailand	9	2
cluster n°3	Cambodia, Laos, Myanmar, Philippines, Singapore	45	3
cluster n°4	Indonesia	9	4
cluster n°5	Vietnam	9	5

The main idea behind VDM is to integrate automatic clustering methods such as K-means in the VDM process. The scatterplot in Figure 63 represents the five clusters found in the data set: yellow triangles represent the cluster with the largest number of countries. The user also can change the parameters of the scatterplot to examine the relationships between other variables. In Figure 63 two variables were selected, population and number killed in traffic. Other relations can be investigated by changing the variables in the scatterplot, In Figure 63 the number of motorcycles is compared to the number of killed in traffic. It is very obvious that Thailand (cluster number 2) has highest number of motorcycles and also the highest number of people killed in traffic.

In Figure 32, X-axes represent the total number of motorcycle and the y-axes, the total number of killed in traffic. All the data items are plotted in the display area according to their attribute values. The color and shape of the data points in the figures are assigned according to the cluster number. The scatterplot indicates the relationship between number of motorcycles / numbers of people killed in traffic. According to the scatterplot, C_Kmeans_2 contains a single country, Thailand (represented by a green cross), which has a highest number of people killed in traffic in the ASEAN region. C_Kmeans_4 (blue cross) has the second highest number of people killed. This cluster contains a single country, Indonesia.

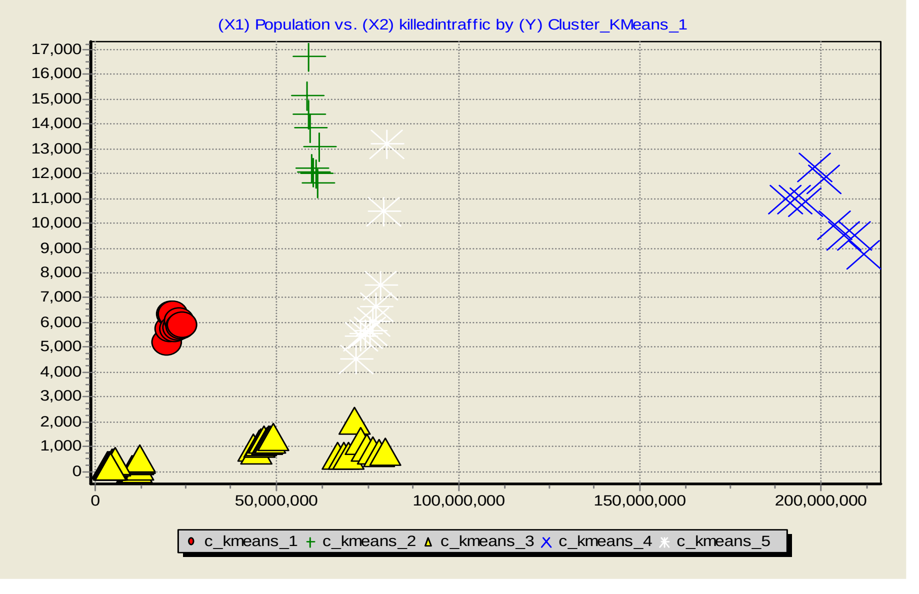


Figure 63. Scatterplot, five clusters were found in the data set; yellow triangles show the cluster with largest number of countries.

7. Discussion

This chapter discusses and evaluates the results presented in Chapter 6. The aim is to assess the research methods used and examine the implications of the results for information sharing practices. The results are analyzed with respect to the main issues identified at the beginning of the research.

7.1. Experience from research methods

At the beginning of this research, the methods used to collect and understand the problems in a systematic way were considered. It is obvious that the methods and techniques of data collection must be reliable and consistent to answer the questions posed by the researchers. People (experts) are the main source of the data; face-to-face interviews with individuals helped to understand the area and the direction of the research. A series of interviews were conducted, based on open questions and sometimes pre-planned questions. During the interviews, answers were clarified through further questioning and, in some cases, questions from previous interviews were repeated.

In Paper V, interviews were combined with focus group discussions (FGD) (Morgan, 1997). The decision to consider FGD in the STRADA study was taken after approaching some of the key persons working with STRADA through questionnaires and emails. The response rates were very low and did not address the questions. This is one of the disadvantages of using questionnaires to target certain groups of people (Germonprez & Mathiassen, 2004).

The FGD meeting was held and organized by the Swedish Transport Authority (Vägverket) in Jönköping, Sweden. The session included the staff responsible for STRADA, and the regional officer in Jönköping who also guided the meeting. During the meeting, group members talked freely and spontaneously about their experiences with STRADA, the information flow, the system structure, the quality procedures and also the cooperation issues between RTA, the police and the hospitals in the region. FGD helped to give an in-depth knowledge and understanding of STRADA from the perspective of its operations. In the same study, interviews were used to complement the FGD, which gave more understanding and helped to verify the outcome of FGD.

7.2. Requirement elicitation, the conceptual model and prototyping

Defining the requirements of an inter-organizational information-sharing platform presented many difficulties. There were problems in getting users to express their requirements and converting these requirements into specifications and system structure. The technical challenges were to develop a system that supports information sharing and makes the organizations involved share and utilize the system and its contents. Through the requirements elicitation process, system requirements were identified by considering all of the requirements of the individual ‘stakeholders’. The approach followed to identify the requirements specifications considered both functional and non-functional requirements; this ensured the completeness and correctness of the requirements (Chung & Nixon, 1995).

Interviewing different experts and researchers from various organizations and backgrounds gave a broad perspective on how traffic and accident data can be collected, analyzed, and shared. The process was investigated from the bottom, where data about individual accidents were reported by the traffic police and then shared with the emergency units at the hospitals. The basic accident reporting form used by the police and the hospital staff has been studied and compared from country to country. It was found that paper-based forms are used in many countries for reporting at the scene of the accident, and as a complement, as in Sweden, police investigators are equipped with GPS receivers for location information and PDAs for filing the basic information concerning the accidents. The information flow, starting with collecting, storing, processing and disseminating accident information was studied following the method described by Shedroff (2001); see also Model Paper V. All of the different international road databases either include aggregated information about accidents or disaggregated information. For international comparability, aggregated data are always needed (Lassarre, 2001). GLOBESAFE was designed to store aggregated data on the country level. It is believed that the majority of experts working in the area is more interested in comparing road safety situations between different countries or regions, and expert’s requirements, in this case, are the aggregated data sets
Studying four international road databases gives an understanding of the issues as definitional problems on these databases, their structures and user interfaces.

The advantages of building a conceptual model were that it defined the system architecture and guided the research of the empirical part of the study and in the implementation (Greenspan, 1985). The model helped to explain the different views and parts of the system. From a research point of view, the model helped to define the different problems and the necessary questions to address. The model also helped in dividing the system into subsystems, which were implemented and tested individually and then integrated to form GLOBESAFE.

During requirement elicitation and development of the system, some of the requirements were unclear and ambiguous; the prototyping helped to clarify these requirements. User feedback obtained from the use of the prototypes was valuable, especially when combined with the outcome of the interviews.

Together with the prototyping approach, an iteration process was used each time in response to the user feedback. In some cases the prototypes were discarded before starting the implementation of the final production 'system' (Scott Gordon & Bieman, 1995). The prototyping was utilized and the first prototype was the first version of GLOBESAFE. This, of course, saved implementation time and the working prototype also provided different stakeholders with a tangible means of evaluating the user interface and the subsystems.

Defining the concepts and terminologies of traffic safety is a prerequisite for the design and development of any accident reporting system (Baker, 1982). It has been mentioned that the field of traffic safety is multidisciplinary. There is a need for agreed definitions, vocabulary and terminologies to be used in a system that offers shared information.
During the early stages of GLOBESAFE ontology was developed that represented the common vocabulary in the domain. This approach provided the possibility of defining all concepts and terminologies that exist in the field of traffic safety. Additionally, the relationships between the concepts were defined.

Different ontology library systems were explored and investigated to find an ontology, which could be reused for traffic safety, instead of developing a new one. The frame-ontology that was developed in this research was converted to the database schema. The main advantage of this part of the research was that it should be possible to reuse the developed ontology for future research. The ontology will minimize the confusion due to differences in the interpretations of the terminology such as 'Accidents, killed'. Also, agreed definitions and a common vocabulary will increase the use of the shared information and even the communication between traffic safety organizations (Wilson & Corlett, 1999).

7.3. Implementation aspects

The development and implementation of GLOBESAFE was based on an iterative prototyping process (Mason & Carey, 1983). This helped to produce a usable system and shortened the implementation time. To meet the requirement specifications, a balance between system performance and the design of the database schema, the user interface, and the access methods must be considered.
Considering that web technologies are in rapid development, at the beginning of this research the Active Server Page (ASP) was the best option, as it supports the building of large, data-driven web-based systems.

The use of ASP supports and enables rapid development and prototyping in a simple way. Another advantage is that ASPs are 'server' pages. That is, ASP scripts are run on the server side, which enables the pages (User Interface) to be browser independent.

GLOBESAFE must be able to handle a variety of queries. This highlights the importance of using ASP technology to build a system that will be shared by a wide range of organizations with different needs and expectations of the systems.

Information architecture principles and visual hierarchy, when applied, lead to an acceptable user interface, as shown in the result of the user study, (Paper III). The design focuses on improving the sharing of traffic safety information through a web interface. The system is designed for different applications to make the user interactions easier regardless of the applications. In addition to the ease of use of the interface, the functionality was in focus as well, and as discussed initially this is a main issue of the research.

GLOBESAFE can handle data sets from different sources, direct input from multiple clients at different locations, and has search facilities supporting the search of the database.

GLOBESAFE is accessible at the URL < http://www.globesafe.org >. The main characteristics and aspects of GLOBESAFE that distinguishes it from other traffic safety databases, are its conventional approaches to accommodate accident statistics, socioeconomic information about countries, energy related information, and spatial data. Moreover, a number and variety of applications were implemented that support benchmarking and comparison of traffic safety situations. The reporting system helps the users to create custom reports on request, and export the output by using different data formats, thus allowing transfer between different systems.

In the analysis of traffic safety, GLOBESAFE offers different types of analytical tools based on mathematical models and correction factors to adapt the data to certain definitions. Also, it presents important indicators when analyzing the traffic safety situation in a country. GLOBESAFE contains methods that can calculate performance indicators, which can be used as diagnostic tools when comparing the traffic safety situation between different countries. All these aspects make GLOBESAFE a promising model for addressing the need to support inter-organizational information sharing in traffic safety**.**

- **Role of information in promoting road safety**

Improvements in traffic safety are the outcome of developments in different sectors and areas, including political awareness and decisions, implementation and enforcement of safety programs (Morsink et al, 2007). Paper II introduced a new method to help decision-makers to improve their understanding of road safety. The method is based on the idea of using a map as an interface for shared information and also to present non-geographical information on the map. The development of this method was combined with other applications, such as Road Safety Profiles (RSP), which provide benchmarking tools for road safety and also help to identify the weak and strong areas of the safety programs. These tools support the benchmarking process that enables comparison between different countries and also helps to share good practices and trends (Harvery, 2004).

The map, as an interface for shared information, was based on an Open Geospatial Consortium (OGC) model and SVG (Papers II & III), which helped to present traffic safety information when using the map as a user interface. It was found that SVG provided adequate and straightforward technology for a 'thick' client –thin server connection using an OGC portal workflow.

Current Internet browsers support SVG with the plug-in SVGViewer. In the near future, SVGViewer will be embedded in all browsers, since it has gained acceptance as a W3C Open Internet standard.

Currently, the process of designing web maps is based around Internet map servers, but this limits the quality of the service, and requires management efforts to keep the server operational and to maintain the clients' requests. The approach used in Papers II and III avoided this limitation by using a client-side approach. In this approach, no map server is required to handle the Geodatabase and the map generation. All operations are executed on the client side, which reduced the connections required between the client and the server.

Current mapping applications offer different cartographic operations that Internet map users are looking for. All operations are performed by sending requests and the response comes from the client. All interaction functions are obtained by adding scripted XML files, with each file relating to one element.

7.4. Quality of the shared information

The accuracy and the validity of the decisions regarding traffic safety depend on how well the information is collected and shared, and the quality of the compiled general data. In general, high-quality information should match the users' needs (Wang & Strong, 1996; Strong et al., 1997). In this research, consideration was given to the five dimensions of data quality as discussed in Erlander (1974) and Yang & Diane (2003).

- **Accuracy and completeness**

The extent to which the information was accurate was a concern for the traffic and accident data that were to be shared. The data must be complete, where completeness refers to the need for the system to contain all the required data to evaluate the road safety situations, analyze and evaluate safety measures, and to support research and development in the field. These factors were considered when building the database schema. Organizations will not be able to share other organizations' data, unless they can ensure that the data set is accurate and complete. Incompleteness refers to the missing data, resulting in data that is not sufficient for that analysis or comparison tasks.
Accurate data are essential for prioritizing public health issues, monitoring trends and assessing intervention programs (Peden et al., 2002). Many countries have inadequate

information systems on road traffic injuries, making it difficult to identify the full nature of the problem and thus gain the attention that is required from the policy and the decision-makers (Proctor et al., 2001). There are a number of areas where road traffic injury data are often problematic. An example of such a situation is the identification of the sources of the data, when the data are from a number of different sources.

- **Relevancy of information**

Relevancy, as with information quality, measures the extent to which the data in the system are applicable to and useful for different tasks required by traffic safety experts or analysts. However, relevancy was nevertheless maintained in GLOBESAFE. Much of the information shared was gained by combining traffic data, accident data, economic and demographic data. The use of the information determines its relevancy to the task. In the case of GLOBESAFE, where data is aggregated at country level, the relevancy of the information is that it can be used for analyses at a country or regional level. Relevancy is linked to the coverage of the information in terms of time and region.

- **Timeliness**

Timelines refer to the extent to which the information is up-to-date for the required task. In the case of traffic safety analysis, timelines could be linked to time series of the available data. This quality was needed to measure and observe safety trends in specific geographical regions (comparison over time). Currently, GLOBESAFE accommodates data with different time series on population data (five years) and traffic and accident data (one year).

- **Accessibility and availability**

Accessibility refers to the extent to which the data are available and retrievable. Availability also reflects the reliability of the system. In GLOBESAFE; users can get into the required data sets as long as they have the access means and authorization to go through that system. Accessible data must be usable in its format and also easy to use and manipulate. Different applications offered by GLOBESAFE are discussed in Paper IV and the applications of visual data mining to increase the usability of the data in Paper VI.

7.5. Increase trust among organizations

Traffic safety organizations are dealing with sensitive and private information (personal information). Security protocols are needed to give the traffic safety organizations enough incentives for sharing information with each other. In some countries, regulations such as the Personal Data Act in Sweden (PuL/SFS 1998:204) regulate the way the personal information is treated and handled by the authorities.

A secure information-sharing platform is required for effective information sharing (Chu et al, 1997; Opplinger, 1999; Hess et al., 2002). One of the primary issues that was considered in the design of GLOBESAFE was the control of user access to the system.

Role-based access control (RBAC) granted access rights to the organizations and to the individual users. Access rights management will increase the partner trust and the user's confidence in sharing the information.

7.6. System thinking approach

Systems like STRADA are prerequisites for organizations such as RTA. STRADA helps to gather the required knowledge to improve traffic safety in Sweden. In Paper V, a system thinking approach was used to analyze and study the results from the FGD and the interviews. This approach was introduced to study the operational system and its life cycle. The intention was to investigate, in depth, the low acceptance of STRADA. The importance of STRADA was discussed within the context of change management and by viewing the road safety management from a system perspective. Shedroff (2001), chain from data to wisdom, was applied and gave good explanations for the different conversion processes.

STRADA has been operating since 2002 and has undergone several development steps. Based o n experience a number of lessons learned from the utilization of the system have been gained. The major problem in the early utilization of STRADA was the lack of interest from the health authorities in accepting and using the system. This resulted in the under-reporting of the total number of accidents. The consequence of this was explained by using a technology acceptance model (Paper V).

The results showed that there is a need for further developments (improvements) of the system, and that the lack of utilization of the system might be due to issues within the user organization. There were not only questions about the technology, but also about the organizations' behavior relative to the system. More effort should be dedicated to motivate the health authorities (Landstingen) in the different counties to demonstrate the benefits in using STRADA at the emergency units. More efforts needed to make Landstingen understand how this will contribute to the country's safety management.

7.7. Implications of visual data mining

Due to the wide availability of traffic data, accident statistics and the need for turning such rich data sets into information and knowledge, data mining was attractive for this research to enable information discovery in the traffic and accident database (Paper VI).

The approach used in Paper VI integrates visual and automatic data mining techniques and methods. It proved to be of great importance and value to discover complex information relations in the road accidents database. VDM allows the discovery of useful and hidden knowledge from the traffic safety data, especially by using visual data mining techniques. Visual data mining (VDM) combines automatic data mining algorithms and information visualization, 'InfoViz' methods, making VDM a highly attractive and effective tool for determination and understanding distributions, patterns and clusters in the data (Han and Kamber, 2006). The results obtained helped to understand the safety-related situations in regions or specific countries. Clustering techniques, used to partition the data set into smaller groups, countries within clusters, are similar in traffic safety situations and other counter measures.

Cluster analysis is performed by detecting similar clusters and patterns in the data set. By using visual ability, a road safety expert might be able to estimate, determine and observe relationships that could otherwise be hidden. From such patterns and observations hypotheses about future trends in the safety situations may easily be developed. The methods used prove the capabilities of the knowledge discovery tool. However, the low dimensionality of the current data set limits the exploration task. The VDM methods used proved to be useful in discovering different relationships in the data set. The main goal was to identify countries that had a high frequency of accidents, killed in accidents, injuries in accidents, a high level of motorization, and where the consumption of diesel and gasoline are high. Detecting countries with high values of these variables will help to relate aspects of safety management to specific reasons, for example high populations, weak safety regulations, poor road design and so forth.

Another advantage of this approach is that hypotheses can easily be formulated to predict future trends. The visualization of the different attributes allowed the descriptions of the characteristics in a single country as well as in groups of countries and regions that share similar characteristics and, therefore, share the same cluster. Returning to the main objective, the discovered knowledge was expressed in an understandable way by using InfoViz techniques, involving humans in the exploration tasks.

8. General conclusions and future work

In this chapter, the general conclusions are presented, reviewing the most important results and their implications for road safety. Ideas for possible future research in this field are also discussed.

8.1. Conclusions

The main theme of this thesis was to support information sharing between organizations and individual researchers/experts working in the field of road safety to develop safety programs and promote road safety in general.

In the previous chapters and in the attached papers the results obtained from the research were presented and discussed. The main findings demonstrate a new approach for sharing traffic and road accident data based on a three-layer conceptual model. This model defined the required infrastructure for the system and other applications.

In the second part a set of requirement specifications for GLOBESAFE were presented. GLOBESAFE was developed based on a client/sever technology, and considers user interface issues. The schema structure of the database was built around the five aspects of information quality. This was achieved by development of a common vocabulary and a set of concepts agreed to in traffic safety when viewed as a knowledge domain.

The behavior of the organizations in the sharing and in the pooling of their information is another important factor utilized in the system. This could be referred to as an infrastructure for information sharing. In the specifications of GLOBESAFE the quality of the information and the security issues related to this information were considered, as these two factors have a direct impact on the organizations' trust in the system.

Special emphasis was put on the accident data acquisition system. Furthermore, a study of the Swedish Traffic Road Data Acquisition system 'STRADA' was conducted, which showed low acceptance of the system and pointed to the under-reporting of accidents as a consequence. Recommendations for future developments and improvements of the system were proposed as well. STRADA was studied in two different contexts. The first was through the Swedish Road Authority (Vägeverket) at the Jönköping office. This office is responsible for operating the system in southeast Sweden. The second context was concerned with VTI, the Swedish National Road and Transport Research Institute in

Linköping, where the research views on STRADA were obtained. The two studies complement each other and gave useful results.

Introducing the visual and automatic data mining techniques to the analysis and knowledge discovery processes to the road traffic and accident databases was of great benefit to the experts. Exploration data analysis also allowed experts to be involved in the exploration tasks. Information visualization methods and techniques improved the understanding of the different data patterns and clusters in the data sets. This method is needed for benchmarking and international comparability of information from different countries.

8.2. Evaluation of GLOBESAFE

Evaluation of an information system refers to the process of measuring or exploring attributes of a specific information system (Ammenwerth et al, 2004). The mapping subsystem of GLOBESAFE has been evaluated using usability and acceptance test as in Paper III. This helps to measure user satisfaction and acceptance of the web map. Heuristic and usability evaluation methods can be used to measure user acceptance and also to understand system performance. Evaluation can serve the purpose of detecting and determining unpredictable characteristics of the system (Kaplan, 2001). A system such as GLOBEASFE should be evaluated with respect to its organizational impact (will a new organization structure be introduced?), its impact on users' roles in accident data collection and analysis (whether the system will improve the way analysis has been done and the consequence of that on road safety).
The most important aspect of evaluation is to avoid system failure in the long run when the system is put into operation.

8.3. Information quality on the scene

Future work may focus on improving the information quality in the traffic and accident data. More research should be conducted at the accident scene level to assure that quality is maintained when data are collected at a single-accident level. Research on information quality processing will decrease under-reporting on the database and will increase the accuracy and the availability of the data. With an efficient accident data system, it is possible to obtain high-quality accident data. The effects of poor quality data will decrease the ability of road safety authorities and organizations to function effectively and promote a higher level of safety. In the case of developing countries, ICT methods need to be introduced as under-reporting in these countries frequently reach the 50% level.

8.4. Geospatial data mining of road accidents database

Visual data mining has been applied to attribute data only. Still, space information is a considerable part of the concern when analyzing accident information. Geospatial data mining ought to be investigated for mining location information related to road traffic accidents. Spatial analysis and modeling is widely used and are core issues in the area of environmental science for modeling spatial phenomena. Spatial data mining can also be used to find relationships between different factors that may lead to traffic accidents.

List of references

Abugessaisa, I., Sivertun, Å., & Le Duc, M. (2007). GLOBESAFE: A platform for information sharing among road safety organizations. In *Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries*, 28-30 May 2007, CD-ROM ISSN 1981-3945. São Paulo, Brazil: IFIP-W.G.

Aeron-Thomas, A. (2000). Under-reporting of road traffic casualties in low income countries, TRL Report PR/INT/199/00. TRL Ltd, Crowthorne. U.K.

Albert, D. & Steiner, C.M. (2005). Empirical validation of concept maps: preliminary methodological considerations. In *Proceedings of the IEEE Fifth International conference on advanced learning technologies*, ICALT'5, 5-8 July 2005 (pp. 952- 953). Los Alamitos, CA: IEEE Computer Society

Ammenwerth, E., Brender, J., Nykanen, P, Prokosch, H.-U., Rigby, M., & Talmon, J., HIS-EVAL Workshop Participants. (2004). Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck International Journal of Medical Informatics, 73, , 479-491.

Andriole, S. J. (1994, March). Fast, cheap requirements: prototype, or else!. *IEEE Software*, 11(2), 85. Retrieved Jan 23, 2008, from IEEE Computer Society Press.

Aptel, I., Salmi, L., Masson, F., Bourde, A., Henrion, G., & Erny, P. (1999). Road accident statistics: discrepancies between police and hospital data in a French island. *International Journal of Accident Analysis and Prevention*, 31,101–108.

Arnold, S. & Lawson, H. (2004). Viewing systems from a business management perspective: The ISO/IEC 15288 standard. *Journal of Systems Engineering*, 7(3), 229-242.

Asian Development Bank, ADB. (2000). *Road Safety guidelines, road accident data system*. Retrieved Jan 23, 2008, From Linköping University, ITN Blackboard site: http://blackboard.liu.se

Baguley, CJ. (2001). The importance of a road accident data system and its utilization. In *Proceedings of the International Symposium on Traffic Safety Strengthening and Accident Prevention*, 28–30 November 2001 (PA3807/02). Nanjing, China: Transport Research Laboratory.

Baker, B.J. (1982). Injury classification and the international classification of diseases codes. *International Journal of Accidents Analysis and Prevention*, 14, 199-201.

Beer, S. (1985). *Diagnosing the system for organization*. Chichester: John Wiley & Sons.

Bernhardsen, T. (2002). *Geographic information systems: an introduction*. New York: John Wiley & Sons.

Berns, S. & Brühning, E. (1998). Definitions and data availability, special report, Retrieved January 28, 2008, from IRTAD: International Road Traffic and Accident Database special report Web site: http://cemt.org/IRTAD/IRTADPublic/publications/abstdeutschland.html

Björketun, U.(2005). *STRADA, koppling mellan Polis- och Sjkvårdsrapporterade data*. Linköping: VTI press.

Blum, I. (1992). *Software engineering: a holistic view*. New York: Oxford University Press.

Breunig, M., Kriegel, H., Kröuger, P., & Sander, J. (2001). Data bubbles: quality preserving performance boosting for hierarchical clustering. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, 21-24 May 2001 (pp. 79–90). Santa Barbara, California USA. ACM Press.

Brouwer, M. (1997). Road safety information system: key information supporting traffic safety policy in the Netherlands. In *Contribution to: Traffic Safety on Two Continents Conference*, 22-24 September 1997 (Report D-97-14). Lisbon, Portugal: Stichting wetenschappelijk onderzoek verkeersveiligheid.

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M., Carvajal, R. (2004). CmapTools: A knowledge modeling and sharing environment. In *Proceedings of the First International Conference on Concept Mapping*, 17-19 September 2004 (pp. 125-133). Pamplona, Spain: Universidad Pública de Navarra.

Carter, R. (1992). Perspectives on sharing data in geographic information systems. *Photogrammetric Engineering and Remote Sensing*, 58 (11), 1557-1560.

Castro, AG., Rocca-Serra, P., Stevens, R., Taylor, C., Nashar, K., Ragan, MA., & Sansone, SA. (2006). The use of concept maps during knowledge elicitation in ontology development processes-the nutrigenomics use case. *BMC Bioinformatics*, 5, 267.

Checkland P. (1999). *Systems Thinking, Systems Practice: includes a 30-year retrospective*. Chichester: John Wiley & Sons.

Chin, J.P., Diehl, V.A., & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of SIGCHI* *conference on human factors in computing systems*, 15-19 May 1988 (pp.213-218). Washington, D.C., USA: ACM Press.

Chu, Y., Feigenbaum, J., LaMacchia, B., Resnick, P. & Strauss, M. (1997). REFEREE: Trust management for Web applications. *The World Wide Web Journal*, 2, 706-734.

Chung, L. & Nixon, B. (1995). Dealing with non-functional requirements: three experimental studies of a process-oriented approach, In *Proceedings of the 17th International Conference on Software Engineering***,** 24 – 28 April 1995 (25-37). Seattle, Washington, USA: ACM Press.

Chung, K., Lisa Davis, C., Fee, C., Grammer, R., Brown, D., & Dalton, J. (2004). Facilitating public safety analysis through enhanced information sharing: improving the geospatial repository for analysis and safety planning. In *Proceedings of the 2004 Systems and Information Engineering Design Symposium*, 16 April 2004 (pp. 87- 93). Charlottesville, VA, USA: Department of Systems and Information Engineering, University of Virginia.

Clarke, K. C. (2001). *Getting started with geographic information systems*. New Jersey: Prentice Hall.

Cobb, D.A. & Olivero, A. (1997). Online GIS service. *Journal of Academic Librarianship*, 23(6), 484-497.

David, S., & Branche, C. (2004). Road safety is no accident. *Journal of Safety Research*, 35 (2), 173-174.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-339.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. (Doctoral dissertation, Sloan School of Management, Massachusetts: MIT, 1986).

Demšar, U. (2006). Data mining of geospatial data: combining visual and automatic methods. (Doctoral Dissertation in Geoinformatics, Royal Institute of Technology at Stockholm, 2006).

Demsar, J., Zupan, B., & Leban, G. (2004). Orange: from Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

Dillon, A. & Morris, G. (1996). User acceptance of new information technology: theories and models. In M. Williams (Ed.), *Annual Review of Information Science and Technology* (pp.3-32). Medford, NJ: Information Today.

DLTR. (2002). *The 2002 Quality Review of Road Accident Injury Statistics*. London, UK: The Department of Transport, Local Government, and the Regions.

Dong, A. & Agogino, A. M. (2001). Design principles for the information architecture of a SMET education digital library. In *Proceedings of the 1st ACM/IEEE CS Joint Conference on Digital Libraries*, 24-28 June 2001 (pp.314-321). Roanoke, VA., USA: ACM Press.

Douglas, D.N. (2000). Developing spatial data infrastructures: The SDI Cookbook Version 1.0, available at http://www.gsdi.org/pubs/cookbook/.

Dumas, S.J. & Redish, C. J. (1993). *A practical guide to usability testing*. Norwood: Ablex Publishing Corporation.

Easterby-Smith, M., Thorpe, R. & Lowe, A. (2002). *Management Research: An Introduction*. London: Sage Publications.

Elvick, R., & Vaa, T. (2004). *Handbook of Road Safety Measures*. Oxford: Elsevier Science Ltd.

Erlander, S. (1974). Suggestions for an improved system for statistical data about road accidents and road traffic. *Accident Analysis & Prevention*, 6(2), 125-139.

European Transport Safety Council. (2001). *Transport safety performance indicators*. Retrieved January 25, 2008, from European Transport Safety Council Web site: http://www.etsc.be/oldsite/perfindic.pdf

European Transport Safety Council. (2003). *Cost effective EU transport Safety measures*. Retrieved January 25, 2008, from European Transport Safety Council Web site: http://www.etsc.be/oldsite/costeff.pdf

Fayyad, U., Grinstein, G., & Wierse, A. (2002). *Information visualization in data mining and knowledge discovery*. San Francisco: Morgan Kaufmann Publishers Inc.

Forslund, G. (1997). Science, computer science and the scientific value of systems built. In S. Hägglund (Ed.), *Selected term papers on methodology of research in computer science* (pp.1-12).Linköping, OS: LiU-Tryck.

Forward S. & Samuelsson, P. (2007). *STRADA- did it turn out the way it was planned?* Linköping: VTI Press.

Gennari, J., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubezy, M., Eriksson, H., Noy, N.F., & Tu, S.W. (2002). The evolution of Protégé: an environment for knowledge-based systems development, *International Journal of Human-Computer Studies*, 58(1), 89-123.

Germonprez, M. & Mathiassen, L. (2004). *The role of conventional research methods in information systems action research*. Boston: Springer.

Gillavry, E. M. (2000). *Cartographic aspects of Web GIS-software*, Department of Cartography, Utrecht University, Netherlands.

Gomaa, H. & Scott, D. B. (1981). Prototyping as a tool in the specification of user requirements. In *Proceedings of the 5th International Conference on Software Engineering*, 9- 12 March 1981 (pp. 333-342). San Diego, CA, USA: IEEE Press.

Gomez-Perez, A., Fernandez Lopez, M., & Corcho, O. (2004). *Ontological engineering: With examples from the areas of knowledge management, e-Commerce and semantic web*. London: Springer-Verlag.

Gottesdiener, E. (2002). *Requirements by collaboration: Workshops for defining needs*. New York: Addison Wesley.

Graham, L. (1999). *The Principles of Interactive Design*. New York: Delmar Publishers.

Greenspan, S. (1984). Requirements modeling: A knowledge representation approach to software requirements definition, (Doctoral Dissertation, University of Toronto, at Toronto, 1984).

Grinstein, G., Trutschl M., Cvek, U. (2001). High-dimensional visualizations. In *Proceedings of the Visual Data Mining Workshop*, 26 August 2001 (pp. 6-19). San Francisco, CA, USA: ACM Press.

GRSP Global Road Safety Partnership. (2003). *Road crash and injury data systems*. Retrieved December 10, 2006, From GRSP Website http://www.grsproadsafety.org/ .

Hanakawa, K. (2004). *Traffic accident analysis system using GIS*. Retrived January 27, 2008, From Environmental System Institute's library Web site: http://www.esri.com

Hand, D., Heikki, M., & Padhraic, S. (2001). *Principles of Data Mining*, Cambridge, MA: MIT Press.

Harper, B., Slaughter, L., & Norman, K. (1997). Questionnaire administration via the WWW: A validation and reliability study for a user satisfaction questionnaire. In *Proceedings of WebNet 97 World Conference of the WWW, Internet, and Intranet*, October 31- November 05, 1997: Toronto, AN, Canada: Association for the Advancement of Computing in Education.

Harris, R.L. (1999). *Information Graphics: A Comprehensive Illustrated Reference*. New York: Oxford University Press.

Harvey, L. (2004). Analytic quality glossary. Retrieved February 3, 2008, from Quality Research International Web site: http://www.qualityresearchinternational.com/glossary/

Heinrich, J. & Mikulik, J. (2005). IRTAD–Reliable past and challenging future. In *Proceedings of Vti Road safety on four continents*, 5 - 7 October 2005. Warsaw, Poland: Vti, Linköping Sweden.

Hendler, J. (1995). Experimental AI systems, *Journal of Experimental and Theoretical Artificial Intelligence*, 7(1), 1-5.

Hess, A., Jacobson, J., Jarvis, R., Seamons, K.E., Smith, B., Yu, L. Yu, T. & Winslett, M. (2002, November – December). Negotiating Trust on the Web. IEEE Internet Computing, p. 30

Hills, B.L. & Baguley, C.J. (1994). Accident data collection and analysis: Progress in the use of the Microcomputer Package MAAP in the Asian Region. In *Proceedings of the 8th REAAA Conference*, 17-21 April 1994 (PA1356/95) Taipei, Taiwan: Transport Research Laboratory.

Hirji K. (2001). Exploring data mining implementation, Communications of the ACM 44(7).87-93.

Hoffman, P.E., Grinstein, G., Marx K., Grosse, I., & Stanley, E. (1997). DNA visual and analytic data mining, In *Proceedings of IEEE Visualization*'97, 19-24 October 2001(pp. 437-441). Phoenix, AZ, USA. IEEE Computer Society Press.

Holm, P. (1996). *On the design and usage of information technology and the structuring of communication and work*. Stockholm: Stockholm University Press.

Hvoslef, H. (1994). Under-reporting of road traffic accidents recorded by the police, at the international level. Retrieved January 28, 2008, from IRTAD: International Road Traffic and Accident Database special report Web site: http://cemt.org/IRTAD/IRTADPublic/publications/abstnorwegen.html

IEEE Std. 1233. (1998). IEEE Guide for Developing System Requirements Specifications (SyRS).

IEEE Std. 830. (1998). IEEE Recommended Practice for Software Requirements Specifications (SRS)

IEEE 1220. (1999). IEEE standard for application and management of systems engineering process, IEEE New York, USA.

Inselberg, A. (2002). Visualization and data mining of high-dimensional data. *Chemometrics and Intelligent Laboratory Systems*, 60, 147–159.

Inselberg, A. (1981). N-dimensional graphics, part I-lines and hyper planes. *Technical Report G320-2711*, IBM Scientific Center, Los Angeles. CA. USA.

Jacobs, I., Gunderson, J., & Hansen, E. (2002). User Agent Accessibility Guidelines 1.0. Retrieved January 28, 2008, From W3C Recommendation Web site: http://www.w3.org/TR/UAAG10/

Jacobs, G., Aeron-Thomas, A., & Astrop, A. (2000). Estimating global road fatalities (TRL Report 445). Crow Thorne: Transport Research Laboratory.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Survey*, 31(3), 264–323.

Kaplan, B., (2001) Evaluating informatics applications-some alternative approaches: theory, social interactionism, and call for methodological pluralism, Int. J. Medical. Informatics 64 , 39–56.

Keim, D. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8,1-8.

Kenneth, E. F. & Shannon, C. (1995). *Cartographic communication*, The Geographer's craft project, Texas: University of Texas.

Koster, J.H. & Langen, M. (2000). Low cost mobility in African cities. In *Proceedings of the Expert Group Meeting*, 21-23 June 2000 Delft, The Netherlands. World Bank.

Larose, D. (2006). Data mining methods and models. Hoboken : John Wiley & Sons Inc.

Larsen, L. (2004). Methods of multidisciplinary in-depth analyses of road traffic accidents. *Journal of Hazardous Materials*, 111,115-122.

Lassarre, S. (2001). Analysis of progress in road safety in ten European countries, *Accident Analysis & Prevention*, 33(6), 743-751.

Lawson, H. (2006). A journey through the systems landscape. Unpublished.

Liu, P. and Chetal, A. (2005). Trust-based secure information sharing between federal government agencies. *Journal Journal of the American Society for Information Science and Technology*, 56(3), 283-298.

Livari, J. & Karjalainen, M. (1989). Impact of prototyping on user information satisfaction during the IS specification phase. *Journal of Information and Management*, 17(1) 31-45.

Luoma, J. & Sivak, M. (2007). Characteristics and availability of fatal road-crash databases in 20 countries worldwide. *Journal of Safety Research*, 38(3), 323-327.

Maher, M. J. (1991). The evaluation of road safety schemes: problems with statistical appraisal. In *Proceedings of County Surveyors' Society/University Transport Studies Group Joint Seminar on the Evaluation of Road Safety Schemes*. London.UK: Transport Research Laboratory.

Makedon, F., Sudborough, C., Baiter, B., Pantziou, G., & Conalis-Kontos, M. (2003). *A safe information sharing framework for e-government communication*. Retrieved January 29, 2008, from Dartmouth College: Institute for Security Technology Studies Dartmouth Computer Science Department Web site: http://www.ists.dartmouth.edu/library/sis0903.pdf

Mason, R. E. & Carey, T. T. (1983). Prototyping interactive information systems. *ACM Communication,* 26(5), 347-354.

Mikulik, J. (1996). Road safety measures: A new way to evaluate them. In *Proceedings of Road Safety in 3 Continents*, Vti P7000:04A:1, Linköping, Sweden.

Mitchell, K.D. (2002). Collaboration and information sharing: an ROI perspective?. *The Public Manager*, 31(1), 59-61.

Moon, J. (2003). APRAD and road safety in ASEAN. *In Proceedings of: The 3rd GRSP ASEAN Seminar*, UNESCAP, Bangkok, Thailand.

Morgan, D. (1997). *Focus groups as qualitative research methods*. Thousand Oaks, CA: Sage Publications, Inc.

Morsink, P., Oppe, S., Reurings, M., & Wegman, F. (2007). Development of Footprint Methodology for Road Safety, *Highway Safety, School Transportation, and Emergency Evacuation*, 104-112.

Murch, R. (2000). *Project Management: Best Practices for IT Professionals*. Indianapolis: Prentice Hall

Murray, C.J.L. & Lopez, A.D. (1996). *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge: Harvard University Press.

Myers, M. D. (1997). Qualitative Research in Information Systems, *MIS Quarterly* 21(2), 241-242.

Nielsen, J. (2000). *Designing Web Usability*. Indianapolis: New Riders Publishing.

Novak, J.D. (1998). Learning, creating, and using knowledge: Concept maps as facilitative tools for schools and corporations. Mahwah, N.J.,Lawrence Erlbaum & Assoc.

O'Leary, D. E. (1993). The impact of data accuracy on system learning. *Journal of Management Information System*, 9 (4) 83-98.

Opplinger, R. (1999). Security technologies for the World Wide Web. Boston: Artech House

Pang-Ning, T., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining,* Boston: Pearson Addison Wesley.

Park, J.S., Sandhu, R. & Ahn, G. (2001). Role-based access control on the web, *ACM Transactions on Information and System Security*, 4(1), 37-71.
Pearlman, E. & House, L. (2002). *SVG for web developers*. New Jersey: Prentice Hall.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A., Jarawan, E. & Mathers, C. (2004). *World report on road traffic injury prevention*. Geneva: WHO Press.

Peden, M., Krug, E., Mohan, D., Hyder, A., Norton, R., MacKay, M. & Dora, C. (2002). *A 5-years WHO strategy for road traffic injury prevention*. Geneva: WHO Press.

Peng, Z. & Tsou, M. (2003). *Internet GIS: Distributed Geographic Information services for the Internet and Wireless Networks*. New Jersey: John Wiley and Son.

Persaud B. & Lyon, C. (2007). Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. *Int. Journal Accident analysis and prevention*, 39, 546–555.

Plewe, B. (1997). *GIS Online: Information Retrieval, Mapping, and the Internet.* Santa Fe: OnWord Press.

Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. New York: Oxford University Press.

Proctor, S., Cook, P., & Belcher, M. (2001). *Practical road safety auditing*. London: Thomas Telford Limited.

Rakotomalala, R. (2005). TANAGRA: a free software for research and academic purposes, In *Proceedings of the EGC'2005*, 19-21 January 2005 (pp.697-702). Paris, France: Association Extraction et Gestion des Connaissances

Reeves, N. (2007). Zing'Em: A Web-based Likert-scale student-team peer evaluation tool, In *Proceedings of Seventh IEEE International Conference on Advanced Learning Technologies ICALT 2007*, 18-20 July 2007 (pp.887-889). Niigata, Japan: IEEE Computer Society Press.

Reichwein, S. (2006). The IRTAD Database: Content, Coverage. Retrieved January 30, 2008 from 3rd IRTAD Conference Web site: http://www.cdv.cz/text/seminare/irtad2006/presentations.htm

Richmond, A. (2002). Web Browsers, available at http://www.wdvl.com/

Roget, P.M. (1987). *Roget's Thesaurus of English Words and Phrases*. Harlow: Longman.

Rosenfeld, L. & Morville, P. (2002). *Information architecture for the World Wide Web*. Sebastopol: O'Reilly.

Road Traffic Accidents. (1999). Road traffic accidents in New South Wales 1997: Statistical Statement. Retrieved January 30, 2008, From RTA Web site: http://www.rta.nsw.gov.au/

Safe traffic – Vision Zero on the move. (2000). Retrieved January 30, 2008, From Vägverket Web site: http://www.vv.se /filer/35555/vision_zero_0604.pdf

Sandhu, R., Coyne, W., Feinstein, H., & Youman, C. (1996, February). Role-Based Access Control Models, *IEEE Computer*, 29(2), 38-47. Retrieved January 24, 2008, From IEEE.

Scott Gordon, V. & Bieman, J. 995). Rapid Prototyping: Lessons Learned. *IEEE Software*, 12(1) 85-95.

Senge, P.M. (1990). The fifth discipline: The art & practice of the learning organization. New York: Doubleday.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *1996 IEEE Symposium on Visual Languages*, 3-6 September 1996 (pp.336- 343). Boulder, Colorado, USA: IEEE Computer Society: IEEE Computer Society Press.

Sluis, J. (2001). *Traffic safety information in South Africa – how to improve the national accident register*. Amsterdam: SWOV institute for road safety research.

Sommerville, I. & Sawyer, P. (1997). *Requirements engineering a good practice guide*. New York: John Wiley & Sons Inc.

Soukup, T. & Davidson, I. (2002). *Visual data mining [Elektronisk resurs] techniques and tools for data visualization and mining*, New York: John Wiley & Sons Inc.

Strand, E. (1998). What's the right way to web map data, Synergetic Inc., at http://ww.geoplace.com/gw/1998/1298/1298nab.asp.

Strong, D.M., Lee, Y.W. & Wang, R. Y. (1997, May). Data quality in context: Communications of ACM, 40(5),103. Retrieved January 25, 2008, from ACM.

Stuart, K., Mackinlay, D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*, San Francisco: Academic press

Svedjemo, G. & Jungert, E. (2006). Ontology as conceptual Schema in database modelling. In *Proceedings of the International Conference On Signal-Image Technology & Internet–Based Systems (*Sitis'2006), 17-21 December 2006 (pp. 186- 197) Hammamet, Tunisia: IEEE SITIS 2006.

SWOV. (2001). Research activities report, Institute for Road Safety Research, Leidschendam.

Taylor, K., Ackland, R., Walker, G., & Jackson, D. (1997). WWW delivery of spatio-temporal modeling results for catchments management: A case study. In *Proceeding of MODISH 97, International Congress on Modeling and Simulation*, 8-11 December 1997 (pp. 1574-1579). Tasmania, Hobart, Australia: Modeling and Simulation Society of Australia and New Zealand Inc.

Toon, M. (1997). The World by your window. *GIS Europe*, 6(11) 38-41.

Tosta, N. (1992). Data sharing. *Geo Info Systems*, 32-34.

Transport Research Laboratory (TRL). (1994). Microcomputer Accident Analysis Package, Version 5 (MAAPfive), User Guide. Crowthorne, UK.

Trawen, A., Maraste, P. & Persson, U. (2002). International comparison of costs of a fatal casualty of road accidents in 1990 and 1999, In: *Accident analysis and Prevention*, 34(3), 323–332.

Tufte, E. (1986). *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.

Wang, R.Y. & Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information System*, 12, 5-34.

Wieringa, R.J. (1996). *Requirements Engineering: Frameworks for Understanding*. New York: John Wiley & Sons.

Wilson, J. & Corlett, J. (1999). Evaluation of human work: *A practical ergonomics methodology. London*: Taylor and Francis.

Wegman, F. (2001). A road safety information system: from concept to implementation, contribution to the road safety training course of the World Bank, Washington D.C. USA.

WHO. (1998). New approaches to improve road safety, Technical report series 781, Geneva Switzerland.

Yang, W. L. & Strong, D. M. (2003). Process knowledge and data quality outcomes, In *Proceedings of the 8th International Conference on Information Quality (MIT IQ Conference),* 7-9 November 2003 (pp.96-106). Cambridge, MA, USA: MIT.

Yannis, G., Golias, J. & Kanellaidis, G. (1998). A comparative analysis of the potential of international road accident data files, *IATSS, Special Feature ITS and Safety*, 22(2), 111-120.

Zheng, Y. (2007). A preliminary evaluation of the impact of local accident information on the public perception of road safety. *Reliability Engineering and System Safety*, 92 1170–1182.