TURUN KAUPPAKORKEAKOULUN JULKAISUJA

PUBLICATIONS OF THE TURKU SCHOOL
OF ECONOMICS AND BUSINESS ADMINISTRATION

Juha Kontio

DIFFUSION OF DATABASE INNOVATIONS
A Multiple Case Study in Six Finnish Organizations

Sarja / Series A-16:2004
ISBN: 951-564-234-5 (PDF) 951-564-233-7 (nid.)
ISSN: 1459-4870 (PDF) 0357-4652 (nid.)

ACKNOWLEDGEMENTS

I started this project at the beginning of 1997 when I signed in as a postgraduate student in Turku School of Economics and Business Administration. At that time professor Mikko Ruohonen was appointed as my supervisor. Years went by and the research didn’t really proceed at all because of my work. After Mikko moved to University of Tampere professor Reima Suomi replaced Mikko as my supervisor. I still remember the first reply from Reima to my email when he wrote something like “nice to hear from you, but who are you?”. Today Reima probably knows me better and I have a lot to thank him for. He helped me kick this research off and assisted in numerous ways during the process. He helped in finding the case studies and has given valuable comments to many papers and earlier versions of this dissertation. Thank you!

Professor Airi Salminen was one of the first persons that commented the idea of this research. She gave me hints and suggestions to the draft of the research. Therefore I appreciate that she agreed for being my opponent and I certainly hope that I have avoided the traps you warned at the beginning.

Professors Ulrich Frank and Matti Linna acted as reviewers of this dissertation. I want to thank you for pointing out the shortcomings of the dissertation, for encouraging me and for asking difficult questions. Your feedback has guided me through this process and showed me the right way to go. In addition, thank you Ulrich for the hospitality during my visit in Koblenz.

My employer Turku Polytechnic and especially my former boss Olli Mertanen have supported my research both mentally and financially. In particular, the financial support that Olli arranged was a key factor in ensuring that the study was completed. Without the chance to concentrate on doing research, I would have never succeeded. I will remember this!

During this research I worked with six interesting organizations, all of which were committed to the project and whose staff open and co-operative. Hopefully you have also gained something from the exercise. It was certainly a real pleasure to do the research with you. My special thanks go to the contact persons of the organizations, who were valuable sources of information and provided essential help with the local arrangements. Thank you Jaakko, Raimo, Leena, Jouni, Juhani and Ari!

This research also showed me the value of a well-organized and well-functioning library. The Polytechnic library in Salo has helped me with my information needs in numerous ways. Thank you!

Foundation of Economic Education, Salon Liikesivistyksen tuki ry and Turku School of Economics Association have supported this research financially. This support enabled fluent progress of the project, whilst also giving me an indication that somebody else is also interested in my work and that it might be important.

I also want to thank professor Hannu Salmela and people in information systems science department in Turku together with rector Tapio Reponen. Thank you Birgit Haanmäki for the help at the end of the process.

One of the final steps in this process was to polish the language. Peter Eckersley had the unenviable work to improve my English. He did a great work with the language as well as pointed me some unclear ideas in the thesis.

My biggest thanks go to my family. This has been a very hard time for me, but it must have been even harder for you. I know that I haven't been the perfect husband and the father during this process. I'm a very lucky man that I have a wife like you! I have seen you tired, but still with sufficient strength to keep the house running and you have given me the time and support to finish this work! And boys, I promise that daddy now has time to play with you more often and you may even be allowed to touch the computer. My warmest thanks and hugs to you Elina, Juuso and Joonas! We have finally made it and it's now time to live a normal life – if we still remember what that means – and enjoy it for a while…

Halikko, 24th of October 2004

Juha Kontio

TABLE OF CONTENTS

FIGURES

TABLES

1 INTRODUCTION

This chapter explains the motivation for this research, defines the main concepts and introduces the reader to the structure of the thesis.

1.1 Motivation for the research

Organizations process large amounts of data in their daily processes. The proper management, effective processing and use of this data is essential for all types of organizations (Walsham 1993). Poor data management will result in ineffective processes and poor decision-making. Altogether, data management has direct influence – either positive or negative – on the business result of the organization. Effective use of data is also consistently rated very highly as a corporate resource (Shanks 1997).

However a recent study shows that more than 25 percent of critical data is incomplete and inaccurate (Gaudin 2004). This poorly managed data presents real problems also from the managerial point of view. The management of large organizations and strategic decision-making typically requires information from several functional areas within an organization (Shanks 1997). These information-related issues have been ranked very highly in several studies. For example, in a Norwegian study IS managers were asked to rank today’s key IS management challenges in order of importance and developing and implementing information architecture came fourth (Christensen et al. 2000). The same paper studied other similar research results and concluded that making effective use of the data resource was in second place. Another study came up with similar results when ranking critical IS management issues: developing and implementing an information architecture was ranked fourth and making effective use of the data resource was in seventh place (Brancheau et al. 1996). Yet another study ranked organizing and utilizing data third out of top IS management issues (CSC 2002a). The same survey also studied the important technology trends and in the Europe the top two issues were 1) Real-time information access through innovations in data storage and management and 2) Open data exchange extending the reach of information across the extended enterprise.

A database plays a central role in modern data processing and it is an essential resource of the information system. It is fair to say that databases play a critical role in almost all areas where computers are used (Elmasri et al. 2000). However, the gap between what the legacy information systems can deliver and the strategic vision of the organization widens when the legacy information systems are unable to adapt to meet new requirements (Gibson et al. 1999).

Relational database management systems have been at the core of data management in all organizations. These have basically offered reliable and powerful solutions for all the data management needs of organizations. Nowadays organizations are confronted with new data management requirements that relational database management systems no longer take care of easily. Vendors have also identified these new requirements and database management systems and other technologies relating to databases have developed significantly in recent years. In databases the development has been both functional and contextual. At first, traditional relational database management systems were extended with object features, which allowed users to define their own data types with methods, i.e. abstract data types. Recently, database management systems have been extended with XML (eXtensible Markup Language) data management capabilities.

All these developments in database management systems are called ‘innovations’, database management system innovations. In this report a shorter version *database innovations* is used when we refer to these developments. Lyytinen and Rose (2003) use the term IT innovations and all of these developments in database management systems can be called ‘IT base innovations’ and located in the Base technology innovation subcategory. Changes in base technology, like functionality, speed, reliability and architectural principle, belong to this subcategory. Rogers (1995) defines innovation as “an idea, practice, or object that is perceived as new by an individual or other unit of adoption”. As this technological development shows no signs of slowing, organizations will face these challenges all the time and the processes for adoption and diffusion of technology remain strong (Chin et al. 2001). All the time we are continuously confronted by new technology and pressured to find ideas that will help generate the most effective and efficient uses of information systems (Gallupe 2000). At the same time, research has shown that the effects of information technology confirm the significance of performance and capability (Hartono et al. 2003). Most IT innovations are also inherently linked to a continuously improving computing capability-cost ratio (Lyytinen et al. 2003). Research also shows that greater investment in IT is associated with greater productivity growth (Dedrick et al. 2003). However, as Rogers (2003) explains: *Getting a new idea adopted, even when it has obvious advantages, is difficult. Many innovations* *require a lengthy period of many years from the time when they become available to the time they are widely adopted.*

The essential role of data management and the developments occurred and occurring in database management systems led to a multiple case study research project called RETRO (Relational Database Development Trends in Organizations). A case study methodology was selected since it is a well-suited approach when the interactions between information technology-related innovations and organizational factors are studied (Broadbent et al. 1998). At the beginning of the research two main goals were defined. Firstly, the project will investigate how different organizations currently exploit databases and secondly how they plan to exploit databases in the future to improve their business. To be exact, the goal was to describe data management solutions in different organizations. From these goals the main research question was defined as *How do database innovations diffuse in organizations?* The main research question is further divided in two dimensions: empirical and theoretical.

In the empirical part two subquestions are defined:

- *How have database innovations been diffused in the cases?*
- *How will database innovations diffuse in the cases in the future?*

In theoretical part three subquestions are defined:

- *How do database innovations influence data management?*
- *How can diffusion of innovations theory predict the adoption of database innovations?*
- *What are the critical organizational success factors for the adoption of database innovations?*

This research is concerned with the field of information systems. This field covers more than just the technological system, or the social system, or even the two together. It also investigates the phenomena that emerge when the two interact. (Lee 2001) Research interest in the IS field has shifted from technological to managerial and organizational questions and now more interest is placed on how context and innovations interact (Benbasat et al. 1987). Like Walsham (1993) says, IS discipline is concerned not only with the development of new information technologies but also with questions such as: how can they best be applied, how should they be managed, and what are their wider implications?

This multiple case study research looked at the main information systems and databases of six Finnish organizations. The organizations and the companies were selected according to specific rules, e.g. companies developing mainly information systems were not selected. Multiple data collection methods were employed during the research. Semi-structured interviews were the main data collection method and altogether 54 interviews were arranged. Other data collection methods were document analysis and observation.

This research also has some roots in my previous work as a principal lecturer at Turku Polytechnic. While teaching database related topics I have constantly had a suspicion as to whether we are teaching the right things or not. A question raised was how do the topics we teach correspond to the methods and tools used in real-life in organizations and companies? A further question was, in which direction should the curriculum be developed?

1.2 Main concepts of the research

As the research focuses on the innovation and adoption of database innovations in organizations, an organization is naturally one main concept of the research. There are many definitions of ‘organization’. For example, Zaltman et al. (1973) define an organization as a social system created for attaining some specific goals through the collective effort of its members. Its most salient characteristic is its structure, which specifies its operation. Similarly, Rogers (2003) defines an organization as a stable system of individuals who work together to achieve common goals through a hierarchy of ranks and a division of labour. Walsham (1993) defines an organization through two metaphors: a cultural and a political metaphor. In his cultural metaphor an organization is viewed as patterns of symbolic discourse and action. The organization’s culture is an active, living phenomenon through which people create the world in which they live. Subcultures are also important since they maintain a distinctive character and ascribe different meanings to the same events. Finally, the culture cannot be controlled, but the management can influence the evolution of organizational culture. In his political metaphor an organization is a loose network of people with divergent interests, but who gather together for the sake of expediency. In all human actions power is intrinsic and it is exercised in a continuous process with subtle local properties. Morality is inevitably involved in the exercise of power. Here, management has a role of balancing autonomy and control. Basically all these definitions have almost the same elements: individuals working together for some reason. Altogether, the organization is the structure where the adoption and diffusion process happens and therefore the concept is very essential to this thesis.

From technical point of view the main concept of the research is the database. It is divided in two parts: DBMS and data. DBMS represents the database management system, i.e. the actual software taking care of the data management. A DBMS can be described with properties such as a) the data model it supports, b) the data types it supports and c) the query languages it supports. An implemented DBMS can further be described by its properties like the name, the version and the year of implementation. Data represents the stored data that is managed by the DBMS. Stored data can be described by its properties like number of rows, space quota used, logical structures used, type of data and maintenance procedures. This research concentrates on databases and database management systems that are based on the relational model or that are extended further. In practice this means that pure object-databases and extended versions of those are out of the scope of the study.

Another important concept in this research is the information system. It is actually a larger concept and it contains the database (see Figure 1.1). However, this research usually presents information systems and databases as two separate concepts, while emphasizing the essential role of the database. When only the information system concept appears it covers also the concept of the database. When only the database concept appears it is always understood as part of an information system. An information system can be described for example by its properties like the type of the IS, implementation year, interfaces, connections to other ISs, number of users, number of transactions, level of documentation and maintenance procedures.

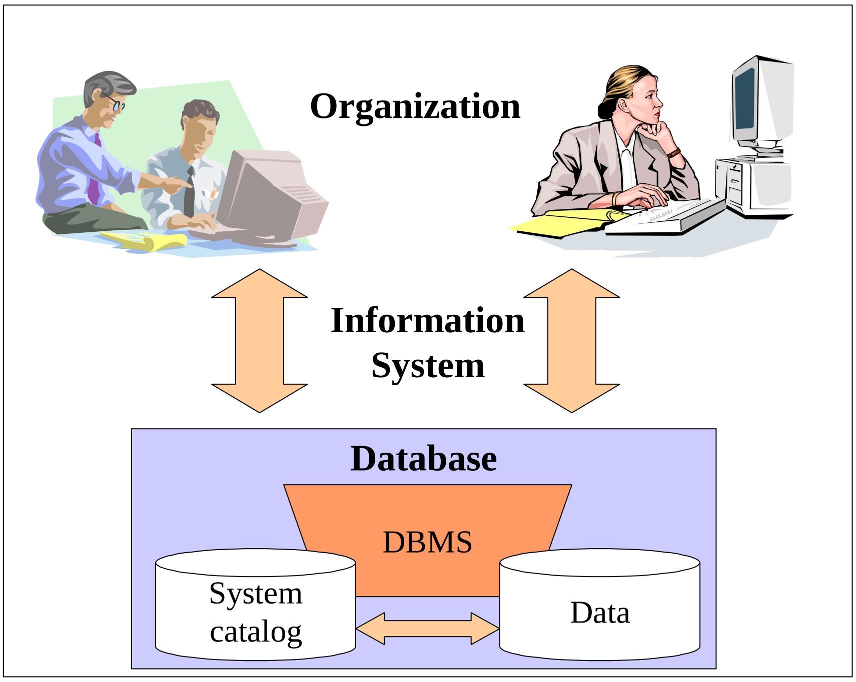


Figure 1.1: Main concepts in this research.

There are also more formal definitions of these concepts, for example in basic database course books. These definitions are presented in Table 1.1. The basic idea of these definitions is the same, but different concepts cover different areas. For example only Begg et al. (2002) and Kroenke (2004) argue that the description of the data is an essential part of the database, in addition to the data itself. Table 1 shows how others have defined ‘database’, in the same way as this thesis defines ‘data’. In this research the concept of database covers data, DBMS and naturally also the description of the data. Thus a database is a collection of related data that is defined in a system catalogue and these are both managed by a DBMS. Date (1995) and Elmasri et al. (2000) use the concept of ‘database system’ in the same way as database is used in this study. DBMS is defined equally in all references presented in Table 1.1 and in this research.

Table 1.1: Other definitions of the main concepts.

	Database	DBMS	Database System
(Date 1995)	A collection of persistent data that is used by the application system of some given enterprise	A layer of software between the stored data and the users of the system	A computerized record-keeping system
(Korth et al. 1996)	A collection of data containing information about one particular enterprise	A collection of interrelated data and a set of programs to access this data	
(Elmasri et al. 2000)	A collection of related data	A collection of programs that enables users to create and maintain a database	DBMS and database together
(Begg et al. 2002)	A shared collection of logically related data and a description of this data, designed to meet the information needs of an organization	A software system the enables users to define, create, maintain, and control access to the database	
(Hoffer et al. 2004)	An organized collection of logically related data	A software system that is used to create, maintain, and provide controlled access to user databases	
(Kroenke 2004)	A self-describing collection of related records	A set of programs used to define, administer, and process the database and its applications	

1.3 Structure of the thesis

This thesis is made up of six chapters (see Figure 1.2). Chapter One is the introduction and it describes the motivation and the main concepts of this research.

The second chapter describes database innovations and their relationship with information systems. It starts by looking at developments in information systems and moves then to integration issues. After that it presents the database innovations relevant for this research and finally it discusses challenges in database design.

Chapter Three introduces the ‘diffusion of innovations’ theory and other diffusion-related topics. These theories were used in interpreting the case organizations’ usage of databases. This chapter is essential to understand database architectures and trends in the case organizations. This chapter also introduces a developed framework of the attributes affecting the adoption of database innovations.

The fourth chapter describes the research: the RETRO-project. It introduces the design of the research, how the data was collected and how the data was analysed.

Chapter Five describes the database architectures and trends in the case organizations. It starts describing the present state of the art in database exploitation from several viewpoints. Finally, the trends in each area are introduced and the case organizations are evaluated with the developed framework.

The last chapter outlines this research’s empirical and theoretical conclusions. This chapter also discusses the reliability and validity of the research. Finally suggestions for future research are presented.

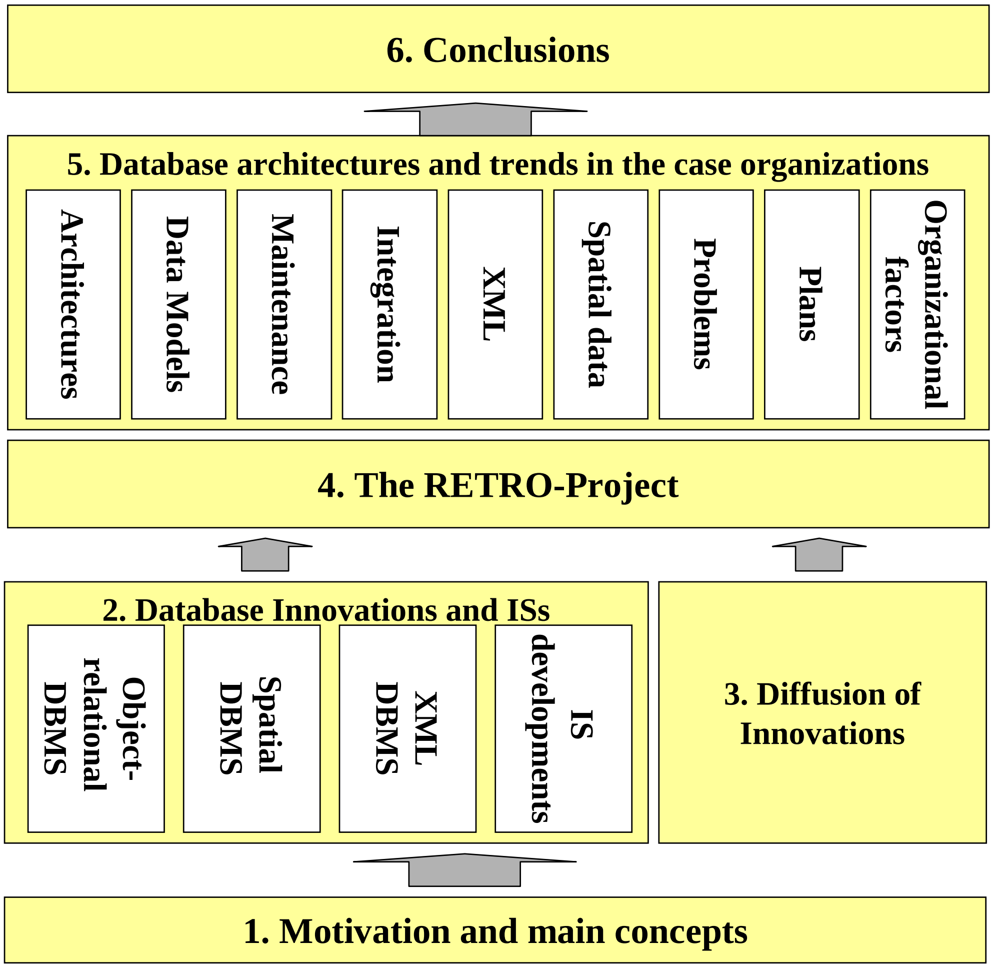


Figure 1.2: Structure of the thesis.

2 DATABASE INNOVATIONS FOR INFORMATION SYSTEMS

This chapter describes the theoretical background of recent developments in information systems and databases. It begins with a discussion of developments in information systems and explains the role of integration. Then it describes the development of database innovations, from relational to object-relational to spatial to XML-enabled database management systems. Next it discusses examples of two important areas of database exploitation: CRM (Customer Relationship Management) and the data warehouse. Both issues have actually been ranked first and second in the four years since 2000 in surveys of top technology-related constraints on organizations' business performance (CSC et al. 2003). Finally the chapter discusses new and emerging database design requirements.

2.1 Developments in information systems

In the first computerized information systems data was stored in files. This file processing typically meant that every purpose had its own specific application files. For example, customer data is needed both in invoicing and in marketing; hence customer data is stored by both of these applications. This led of course to redundant data, which resulted in wasted storage space and redundant efforts to keep common data from different applications up to date (Elmasri et al. 2000). It also meant that the integrity of the data was questionable: which of the customer's addresses was correct, the one in the invoicing file or the other in the marketing file? In addition, the separation of files for specific application programs made it more difficult to access data that should have been available. Finally, the redundant data might even lead to loss of metadata integrity, because the same data items could have different names in different files or same name might be used for different data items in different files (Hoffer et al. 2004).

Typically, the files were dependent on the information system, while the data definition was part of the application program. Thus any changes to the structure of a file also required changes in all application programs using this file. This kind of dependence is called program-data dependence (Elmasri et al. 2000; Begg et al. 2002; Hoffer et al. 2004). A contradictory property is program-data independence, which means immunity of applications to changes in the way the data is stored and accessed (Date 1995). This independence is actually a result of data abstraction. A DBMS provides users with a conceptual representation of data that does not include many of the details of how the data is stored or how the operations are implemented. (Elmasri et al. 2000) Program-data dependence also led to long development times. Another major problem of file storage was that it did not support concurrent users. Of course the data files could have been copied to every user, but this would introduce redundancy and integrity problems once again.

In summary, the file-based systems had two major limitations (Begg et al. 2002): a) the data definition was embedded in the application program and b) there was no control over access to and manipulation of the data.

Little by little, alternatives for files were brought into markets once database management systems based on hierarchical and network data models were introduced. These systems solved many of the problems associated with the files, but still some problems stayed and others were introduced. For example, data access required complex special programs, data independence was minimal and no widely accepted theoretical foundation existed. Finally, relational database management systems were introduced and they have now become the cornerstone of almost every information system in the world. This relational database approach offers many benefits over the older approaches (Date 1995; Elmasri et al. 2000; Begg et al. 2002; Hoffer et al. 2004):

1. Redundancy can be reduced.
2. Inconsistency can be avoided.
3. The data can be shared.
4. Standards can be enforced.
5. Security restrictions can be applied.
6. Integrity can be maintained.
7. Conflicting requirements can be balanced.
8. Data independence can be achieved.
9. Centralized control of the data is provided.

From the very beginning, information systems have been very narrowly concentrated systems, often planned for some special function like order management or invoicing. In an organization many special information systems existed for every important function or process. Unfortunately every information system also had its own database for managing the data once it had been processed. These databases exhibit many of the same problems as old file systems. They were totally separate from each other and no integration was implemented. In practice this meant that the same data was stored in many different data stores and its integrity was poor. For example, a customer could be stored in an order entry database, an invoicing database and a marketing database. When a customer's phone number changed it was updated in one of the databases, but not in the other, separate databases. We can call this phase of databases as 'workgroup databases' (Hoffer et al. 2004). However, when compared to file-based solutions, workgroup databases did offer benefits, for example better data sharing and concurrent multi-user facilities.

In the next stage, information systems started to use common databases. Organizations started to integrate data sources of functions that were related to each other. Data management was organized according to the functional units, typically departments. Therefore we can refer this phase of databases as department databases. For example, the databases of Order entry and Invoicing systems were integrated. This improved the data management situation substantially, but there were still many separate information systems with separate databases. Thus the need to integrate different data sources still existed.

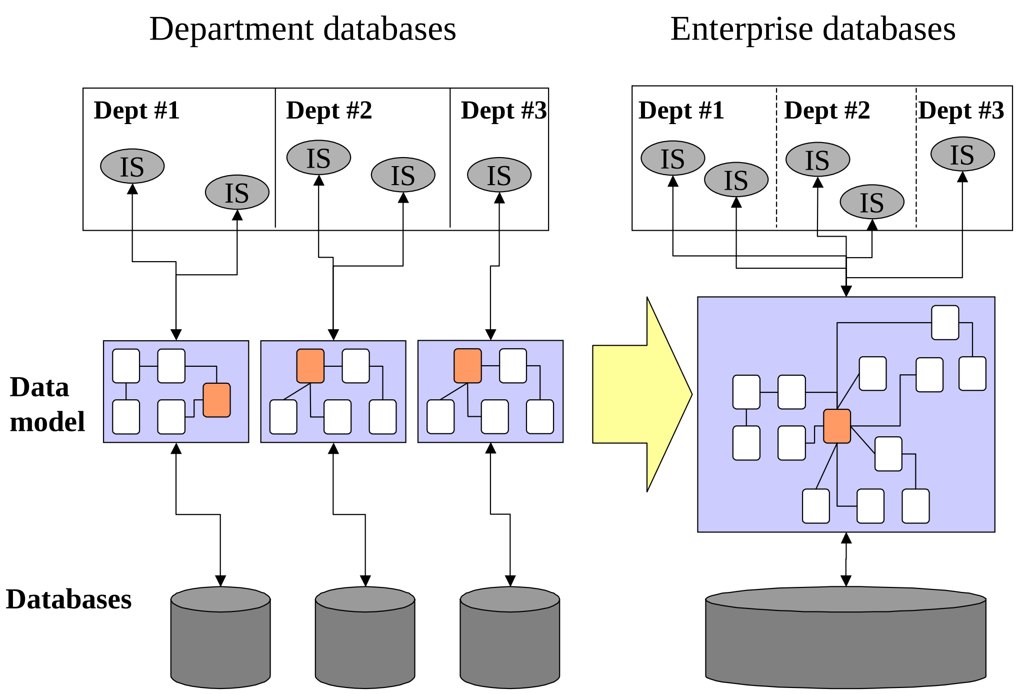


Figure 2.1: Department databases vs. Enterprise database.

An enterprise database is intended to support organization wide operations and decision-making. Typical examples of enterprise databases are Enterprise Resource Planning (ERP) systems and Data Warehouse systems. These databases are based on enterprise level data model. The differences between department and enterprise databases are visualized in Figure 2.1. Enterprise databases require standardized data names, definitions, formats and other related issues. It is also essential to successfully integrate numerous old systems into the new enterprise database. In addition to this internal integration there is an increasing need to integrate with other organizations and offer customers means to access the information systems. Basically, there are requirements for dealing with Business-to-Customer and Business-to-Business relationships. These issues are described in the next section.

The next step for information systems was workflow-oriented information systems. These systems coordinate all entities involved in the execution of a process (Becker et al. 2002) and provide a workflow-based view of the data. There is a new business layer to access all data sources (see Figure 2.2), which could also be called the ‘workflow management system’ (Becker et al. 2002). This business layer takes advantage of the enterprise-level data model and offers technical means to access the different data sources, for example with XML-based solutions. Thus the business layer provides XML to the information systems when and where it is required. Actually, data integration is required to make workflow-relevant data accessible to the business layer (Becker et al. 2002). The enterprise level data model plays a very essential role here, because the data must be described and understood.

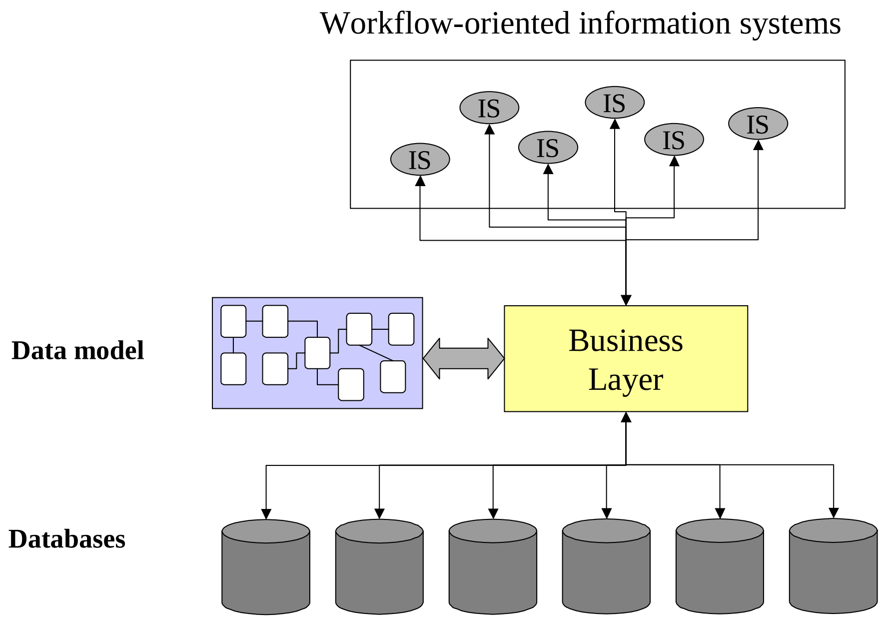


Figure 2.2: Workflow-oriented information systems.

There is a major idealistic difference between the enterprise database solution and workflow-oriented information systems. In a workflow-oriented information system the information system asks for a service from the business layer and the business layer takes care of retrieving and returning the necessary data. This means that the data source is not predetermined. In an enterprise database solution the information systems query the enterprise database and the data source is predetermined. The example in Figure 2.2 describes an intra-organizational structure, but the business layer of the workflow management system might also serve inter-organizational processes. This fosters the development of vendor-independent frameworks that aim to standardize data schemas for common business documents and typically XML is introduced as the solution.

2.2 Integration of information systems

As the previous section shows, the integration of information systems is becoming increasingly important. A number of studies also demonstrate this – for example, 80 percent of respondents to one survey indicated that it was significant for their organization to achieve higher integration between heterogenous systems and applications. Furthermore, 32 percent of those evaluated thought it was extremely significant. (CSC et al. 2003) Difficulties with integrating existing systems are also identified as one top barrier to conducting e-Business (CSC 2002a).

One commonly used term when speaking about integration is Enterprise application integration (EAI). Fundamentally, EAI is about integrating information systems and databases to solve problems in business processes. The term 'Business-to-Business' (B2B) is used, but then the focus becomes integrating different organizations and their information systems. (Linthicum 2001) Integration also supports effective intra-organizational business processes within organizations (Hasselbring 2000). Especially e-Business and business process management are processes that speed up integration efforts: e-Business requires linking the core information systems to the web site and automation of process management requires the integration of core systems (Fremantle et al. 2002). The e-Business infrastructure must integrate numerous information systems to be able to provide necessary service to customers (Nguyen 2000). Typically, organizations maintain many databases that were not designed to interoperate or integrate. Arjansani (2002) has developed a model (see Figure 2.3) describing the levels and steps to take when integrating various data sources (silos) to gain greater competitive advantage. As Arsanjani (2002) writes, these levels should be considered as a roadmap to determine how to map the current state of information and service integration within an organization onto a desired future state.

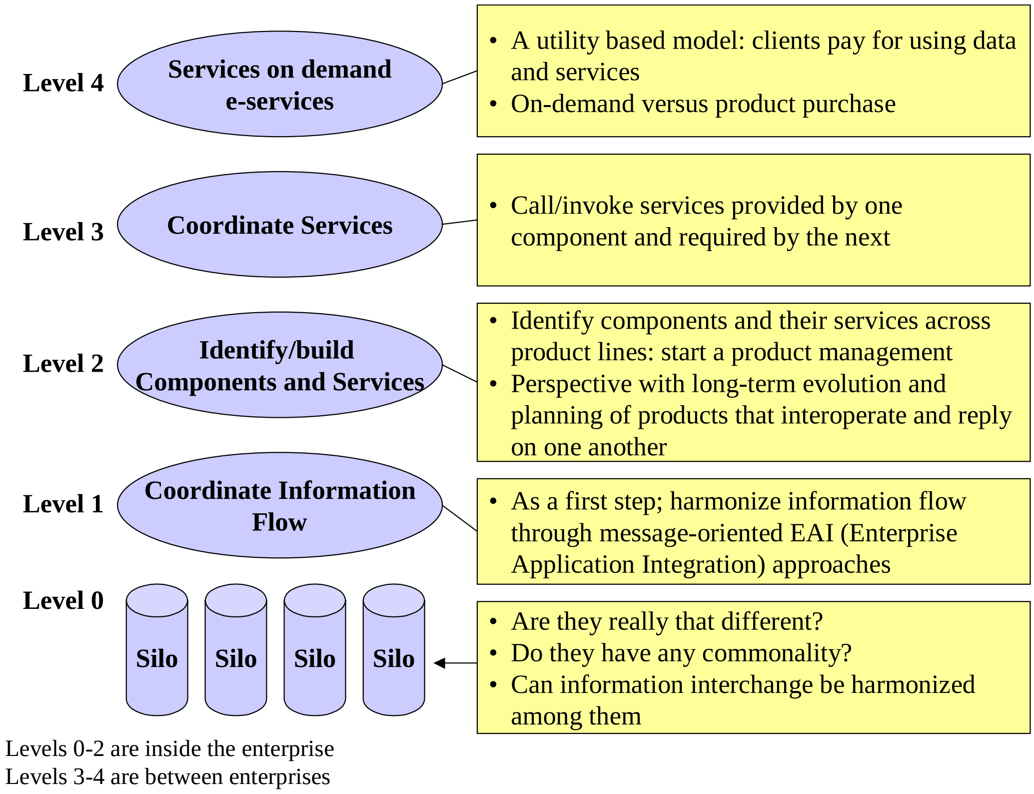


Figure 2.3: Levels of integration (Arsanjani 2002).

The interoperability of information involves the physical environment that is responsible for exchanging the information between the systems, but also the actual information itself (Busichia et al. 1999). Therefore, one basic requirement to be able to integrate information systems is that the data in these systems is understood. On the one hand, the data must be modelled in order to understand it, but on the other, the format of how data is provided to other information systems must be standardized. (Batini et al. (1986) and Barron et al. (1995) recommend integrating databases at the conceptual level. Then the integration process is divided into three stages: 1) Comparing the conceptual models, 2) Standardizing the conceptual models and 3) Combining the conceptual models. The result of this process is an enterprise-level data model (Batini et al. 1986), which describes a minimum information structure that the information systems must meet to be able to interact with each other (Busichia et al. 1999). Integration of databases can be called ‘data-oriented integration’ (Linthicum 2001). This only supports static and functional integration and does not adequately support process-oriented organization. Other types of integration include application interface-oriented, method-oriented and process integration-oriented integration (Linthicum 2001).

Web Services is an example of application interface-oriented integration. The main goal of Web Services is to offer information systems a method to communicate with each other over the web (Treese 2002). This is also a typical example of where the format of the data must be standardized, and quite often XML is introduced as a solution (Hasselbring 2000). Web Services represents an integration technique that is based entirely on XML and web-standards (Stal 2002). The strengths of XML make it very valuable in all types of integration projects, but it should also be remembered that other integration alternatives exists (Linthicum 2001). However, as a system independent language, XML offers tremendous advantages in presenting the data in a format that is understandable both to a human and a computer (Chester 2001). This independence means that every computer that understands XML will recognize any XML document that is offered to it. Of course, the computer might need further information to fully understand the XML document. For example, if the DTD (Document Type Definition) or the XML Schema is accessible, the structure of the XML document can be described and validated easily (Chester 2001). A typical way of using XML is to use the SOAP (Simple Object Access Protocol) –protocol, in which programs request actions from other programs (Treese 2002).

The benefits of XML will not be realized in full until organizations have several information systems capable of managing the language (Lear 1999). However, that time is fast approaching (Linthicum 2001). Thus, the management of XML documents in databases is becoming an important question. The first requirement is that information systems must externalise the information as XML documents. Another requirement is that XML documents or the information in the XML documents must be stored in a database. However, this XML data management is not a straightforward issue. For example, there are several ways to store XML documents in a database (Ha et al. 2001; Khan et al. 2001; Kim et al. 2001; Larson et al. 2001; Schöning 2001; Banerjee 2002; Chen et al. 2002). This combination of XML and databases will be further discussed in section 2.6.

2.3 Evolution of databases

When database management systems are classified into generations, a typical distinguishing factor is the data management capabilities of the DBMS, with its underlying data model as a critical factor. A data model is a collection of concepts that can be used to describe the structure of the database: data types, relationships, and constraints (Elmasri et al. 2000). It defines the available data management possibilities. The relational data model has basically only two constructs to describe the data: its relations and attributes. In addition it offers semantic features like domains and constraints, e.g. referential integrity constraints. The relational data model does not support encapsulating the data and behaviour for example. However, along with relational database management systems it has become a standard of data management. Relational databases have dominated data management markets since the 1980s. They have developed and shaped themselves to a very reliable and efficient solution for most data management needs, becoming almost a standard part of information systems and form the second generation of databases. The step from the first generation of databases (from databases based on the hierarchical and network data models) has been one of the most significant in modern data management.

However, data management needs are changing and new requirements have emerged. For example, relational databases or actually the relational data model are criticized for breaking the natural entity into multiple smaller parts. Along with object-oriented programming, new requirements and criticism toward relational DBMSs, object database management systems were developed. An object DBMS combines the traditional database capabilities (persistence, transactions, concurrent users, recovery, integrity and query capabilities) with capabilities of object-oriented programming (composite objects, encapsulation, inheritance, overloading and late binding). Thus in an object database, as in object-oriented programming, the basic idea is encapsulation, which means combining the object’s data and the management of that data into a single unit (Banerjee et al. 2000). An object DBMS is based on the object data model, which offers a higher level of abstraction than the relational model. The data is represented in an object DBMS more naturally and closer to the real world than in the relational data model.

There are a number of formal definitions of an object DBMS, including:

- Object-Oriented Database System Manifest (Atkinson et al. 1989).
- Third-Generation Database System Manifest (1990).

In addition, standards like ODMG[1] 1.0 (Atwood et al. 1993), ODMG 2.0 (Barry et al. 1997) and ODMG 3.0 (Barry et al. 2000) have been developed.

Object databases never really gained a position in database markets, but the idea of combining objects and databases also encouraged developments in relational database management systems. As a result, object-relational database management systems were introduced at the end of 1990s, at approximately the same time as SQL:1999 was approved. The SQL:1999 standard is actually divided in number of parts (Begg et al. 2004). It offers

[1] ODMG 1.0, 2.0 and 3.0 are all examples of standards that the Object Database Management Group has defined. Object Database Management Group is a consortium of several database vendors that defines standards for object database management systems.

four new data types: large object (LOB) with the variants character LOB (CLOB) and binary LOB (BLOB), Boolean, Array and Row, but maybe the most fundamental facility is its support for user-defined types. Furthermore, it provides syntax for references i.e. REF types. (Eisenberg et al. 1999) With REF types relationships can be implemented without using foreign keys (Cavero et al. 2004).

However, the new object-relational DBMSs had their own interpretations of the SQL:1999 standard and they varied from product to product. The object-relational and the object database management systems belong to the third generation of databases (see Figure 2.4), forming the object-oriented DBMS generation (Cavero et al. 2001).

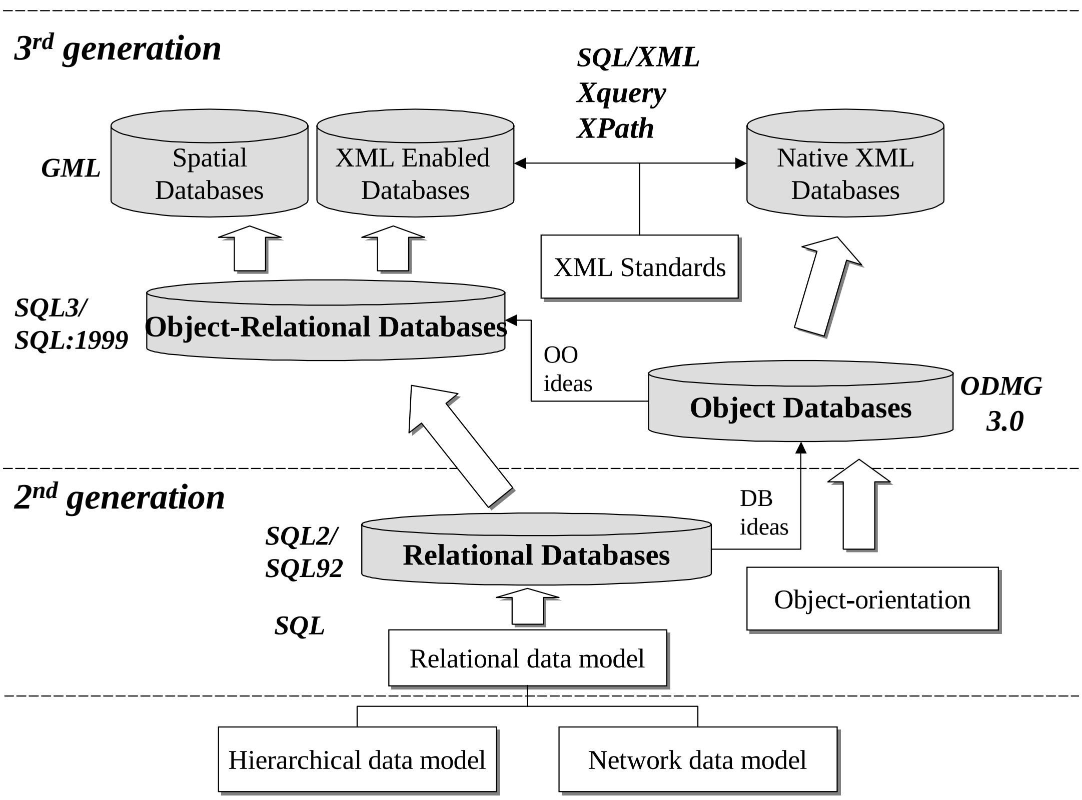


Figure 2.4: Generations of databases.

Recently, developments have concentrated on combining XML and databases. One sign of the importance of XML is that most database management systems manufacturers are already providing or are developing XML-related tools and functions in their products (Cheng et al. 2000; Bertino et al. 2001; Dodge 2001; Leon 2001). Also, studies confirm that the role of XML is becoming important. Research has shown for example that 79 % of North-American companies are already using or are planning to use XML (Hicks 2002). In 2002 Gartner estimated that 70 % of online transactions would utilize XML at the end of 2003 (Oracle 2002), although corresponding real data is not yet available unfortunately. However, databases supporting XML are said to ease the integration efforts of information systems (Scannell et al. 2001). This development has enriched the third generation of databases and as a result both XML-enabled and native XML databases are now available (Dyck 2001; Tillet 2001; Bourret 2002a; Chen et al. 2002). Typically, when object-relational database management systems have been enriched with XML data management capabilities, they are referred as 'XML-enabled databases'. Similarly, enriched object databases are called 'native XML databases'.

Nevertheless, it must be emphasized that both XML enabled and native XML databases belong to the third generation of databases: they are only enriched with XML data management tools and properties. However, this enrichment has also influenced SQL and its new standard SQL:2003 has been ratified. For example, this new standard contains the SQL/XML part, which defines extensions to SQL for XML data management (Begg et al. 2004). As this research has its roots more on the relational database side than the object database side, we will concentrate on issues of XML-enabled databases.

Similarly, spatial databases are only enrichments of object-relational database management systems. Figure 2.5 describes what has actually happened. Relational DBMSs set the standards for database management systems, by supporting data independency, integrity, transactions, concurrency and many more relevant features. An object-relational DBMS offers these services too, but is also capable of managing objects. Similarly, XML-enabled and spatial DBMSs have added new features to object-relational DBMS and are capable of managing even more versatile groups of data.

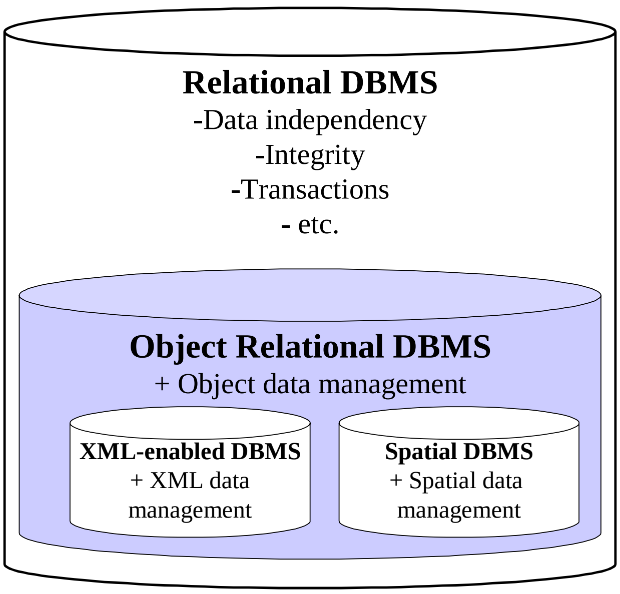


Figure 2.5: Enriching relational DBMS.

When third generation databases are enriched with XML data management capabilities additional features and tools are brought into the database management system, because:

- XML documents are semantically different from any existing data type.
- XML documents are typically composed of several different data items.
- XML documents typically follow some predefined model (DTD, XML Schema) of the application area in question.
- XML documents are self-explanatory, whilst containing also the meaning of the data, which is not typical compared to data in other formats.
- Storing XML documents exploits solutions that already exist, without requiring additional tools.

Table 2.1 outlines the relevant new features and contribution of each database generation to the database world. As the table shows, the first generation laid the basis for databases. The second generation then offered the strengths that we nowadays find in relational DBMSs. In the third generation the user-defined types and large objects are introduced. Finally, in the enriched third generation the focus is on XML data management and spatial data management.

Table 2.1: Main features and contribution of database generations.

Generation	**Relevant new features**	**Main contribution**
1^{st}	The idea of a database	Alternatives to data management
2^{nd}	Nonprocedural data access Data independence	Better usability and performance of data management
3^{rd}	Rich object structures and rules	Object management with SQL Complex data types available; closer to real life
Enriched 3^{rd}	XML data management	Tools for IS integration; Combines management of structured and unstructured data
	Spatial data management	Tools for storage and analysis of geographical objects

In fact, XML data management requires many new features from the database management system. New enrichments include XML storage, support for XPath, XQuery and XML Schema:

- XML storage means that the DBMS supports storing XML documents in the database in their native form, preserving the original structure of the XML document.
- XPath and XQuery are both query languages for searching information within an XML document. They both use regular path expression to navigate the logical structure of the XML documents (Meier 2003).
- XPath provides common syntax and semantics for locating and linking to information contained within an XML document (Chaudhri et al. 2003).
- XQuery is a W3C[2] working draft for an XML query language and XPath is an earlier subset of the draft (W3C 2004b).
- XML Schema is used to define the structure of the XML document: the order of the elements and corresponding data types (Chaudhri et al. 2003).

2.4 From relational to object-relational databases

A traditional classification of database field (see Figure 2.6) is the one by Brown et al. (1999). The classification is based on two dimensions: type of data and query requirements. Both dimensions are divided into only two parts to keep the Figure easy to understand. The type of the data is either simple or complex, although in reality the complexity of an application's data can vary. The situation is the same with queries, although only No Query and Query alternatives are presented here. Typically, information systems have managed simple data and used querying capabilities effectively, thus justifying the use of relational database management systems. A traditional relational DBMS manages numbers, characters, dates and strings, but is incapable of managing complex data like geographic objects, pictures and multimedia (Stonebraker 1998; Brown et al. 1999; Elmasri et al. 2000). To represent complex objects, a relational database decomposes objects into a large number of rows that are linked with each other using foreign and primary keys. When these objects are queried a number of joins are required, resulting in reduced performance. (Cavero et al. 2001) This provides the main rationale for object-relational

---

[2] World Wide Web Consortium (W3C) promotes and develops standard technologies for the Web. See www.w3.org.

database management systems: data is no longer only simple rather it is becoming increasingly complex. In addition, Brown et al. (1999) mention that information systems are becoming universal applications where management of complex data is added to traditional information systems. They also estimate that the relative size of DBMS markets in 2005 will be on a ratio of: Object-Relational DBMS 150: Relational DBMS 100: Persistent Language[3] 1.

Query	Relational DBMS	Object-Relational DBMS
	Universal applications	
No Query	File System	Persistent Language
	Simple Data	Complex Data

Figure 2.6: A matrix for classifying database management systems (Brown et al. 1999).

There is no exact definition for object-relational database management systems, but there have been two quite in-depth discussions on the subject. Brown et al. (1999) define four major features of a DBMS:

1. Support for enlargement of data types via SQL including functions and operators.
2. Support for complex objects at the SQL level (object collections, object references, user defined data types).
3. Support for inheritance at the SQL level.
4. Support for rule systems (e.g. triggers).

The other presentation stems from the relational model and concentrates on combining relations and objects (Date et al. 1998), rather than the main features of an object-relational DBMS. For example, an object-relational DBMS should a) include an enhanced version of SQL for managing both tables and objects, b) support traditional object-oriented features and c) support some non-standard data types within a relation (Hoffer et al. 2004).

The adoption rate of object-relational databases has not been high, partly because they were introduced at the same time as organizations had to deal with Y2K (Year 2000) challenges (Levin 1998). However, through version updates the database management systems have changed into object-relational

[3] Persistent language refers to tight integration with a programming language and high performance for updates to persistent variables referring actually to object-oriented databases that are usually based on C++, Java or Smalltalk.

DBMSs, but the information systems are still only using the features typical to relational databases and the new object-related features of the DBMS are not exploited. Thus an object-relational DBMS is a safe investment, since it can be operated with an earlier relational DBMS and it is also possible to develop information systems exploiting object-relational features (Brookshaw et al. 1998). These object-enlargements enable new application areas, one of which is geographical information systems (OpenGIS 2002).

The object-relational database management systems have data types for large objects (character LOBs and binary LOBs) and they accept user-defined data types, but at the same time these SQL:1999 extensions meant that SQL has become more complicated (Lee 2002). One advantage of user-defined types is that the database structure can be modelled closer to the real world. For example, an order is typically represented in a database as Figure 2.7 shows, but when user-defined types and collections are utilized the definition looks simpler as in Figure 2.8[4].

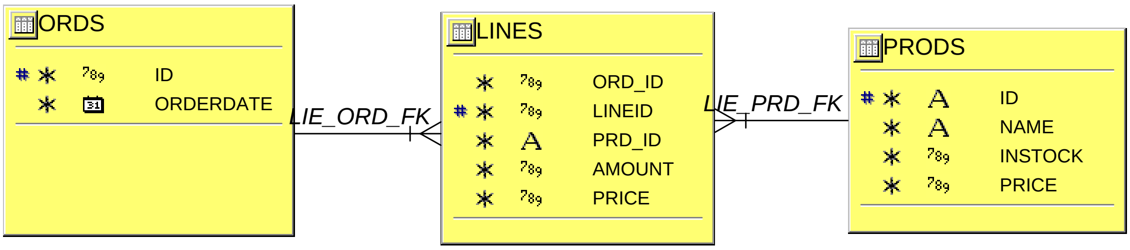


Figure 2.7: Example of an order in a relational DBMS.

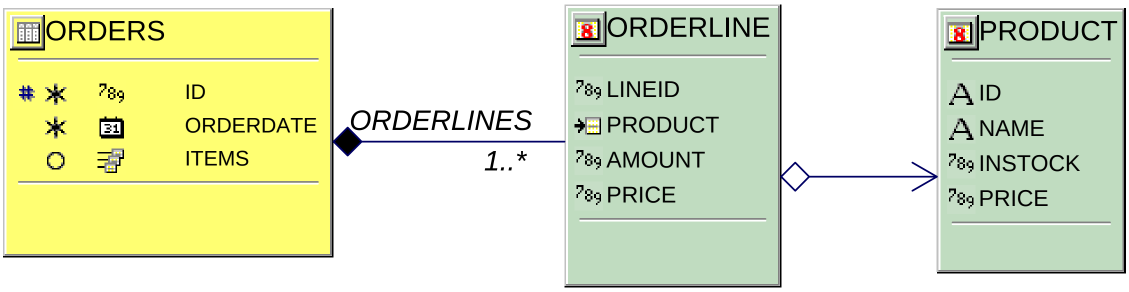


Figure 2.8: Example of an order in an object-relational DBMS.

[4] Both diagrams are drawn with Oracle Designer's Design Editor.

The difference is that in the object-relational solution the order contains order lines that contain the products as references. To query an order in a relational database you need to access and join up three tables to find the relevant information. Basically, the user must be aware of the attributes involved in the relationships. The number of required joins also reduces the performance of the query. However, in an object-relational database you query an orders-table and you get all of the information via that table. For example, the corresponding SQL statements[5] for order 111 are shown in Figure 2.9 for both cases. Furthermore, an object-relational database management system usually offers an additional set of operators and functions to manage these new types of data. For instance, there are special functions (like DEREF[6]) to manage the references in the previous order-example.

RDBMS

```
SELECT o.id, orderdate, name, l.amount amount, l.price price
FROM orders o, lines l, prods p
WHERE o.id=l.ord_id and l.prd_id=p.id and o.id=111
```

Id	Orderdate	Name	Amount	Price
111	18-MAY-04	Portable PC	1	24
111	18-MAY-04	Printer	5	10

ORDBMS

```
SELECT id, orderdate, deref(i.product_id) item, i.amount amount, i.price price
FROM orders o, table(select items from orders where id=o.id) i
WHERE id=111
```

Id	Orderdate	Item (id, Name, Instock)	Amount	Price
111	18-MAY-04	PRODUCT('2', 'Portable PC', 25)	1	24
111	18-MAY-04	PRODUCT('3', 'Printer', 10)	5	10

Figure 2.9: Examples of queries.

Another advantage is that user-defined types also allow for the definition of methods, such as in classes of an object-oriented programming language. For example, the T_Person object type has a method called Age (see Figure 2.10 and see also the example of defining and using a method in Figure 2.11). This functionality implements program-operation independence: Information

[5] The statements are tested in Oracle 9.2 DBMS.
[6] DEREF(parameter) return the object that the parameter refers to.

systems can now use the data by invoking these methods through their names and arguments, regardless how the methods are implemented. For example, the age of a person can be queried using the Age method (see Figure 2.11). User-defined types also offer a way to standardize the database structures. For instance, an address type could be defined and used everywhere address information is needed in the database. This way address is always expressed in the same way and attributes are consistently named everywhere an address is needed. This enforcing of common labels for disparate data sources is very important, for example from the viewpoint of integration and data warehouse design (Kimball 2002). In Figure 2.10 a collection by both Employees and Customers tables uses the T_Address type and also the T_Phone type. Similarly, the Customers table's 'contact' attribute is defined as T_Person, thus standardizing the way a person is expressed in the database.

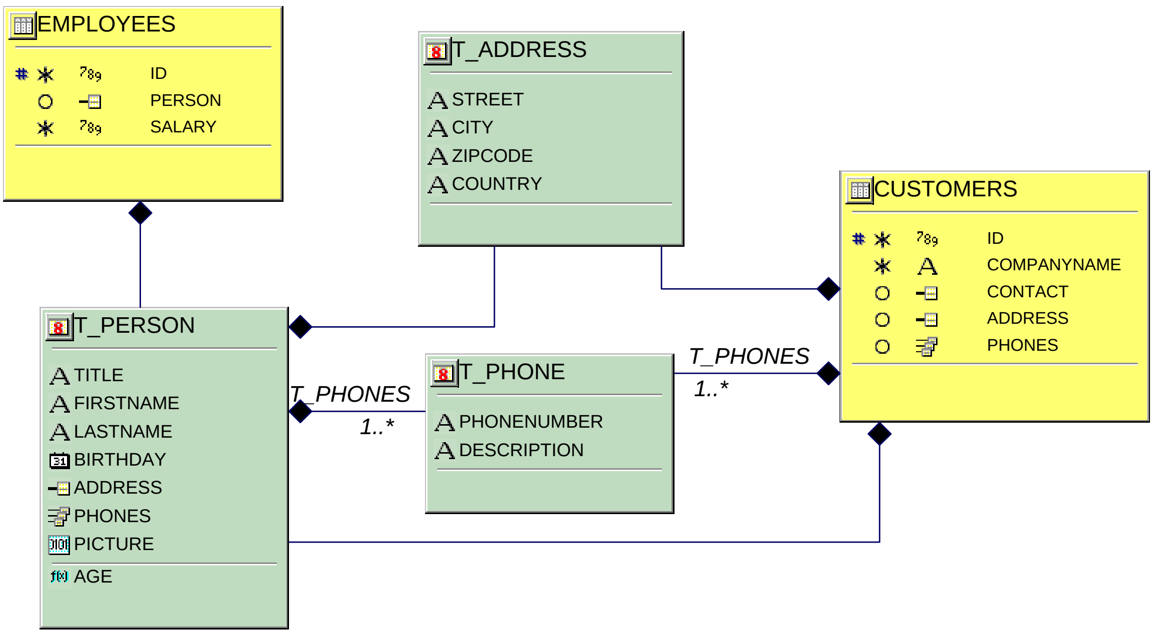


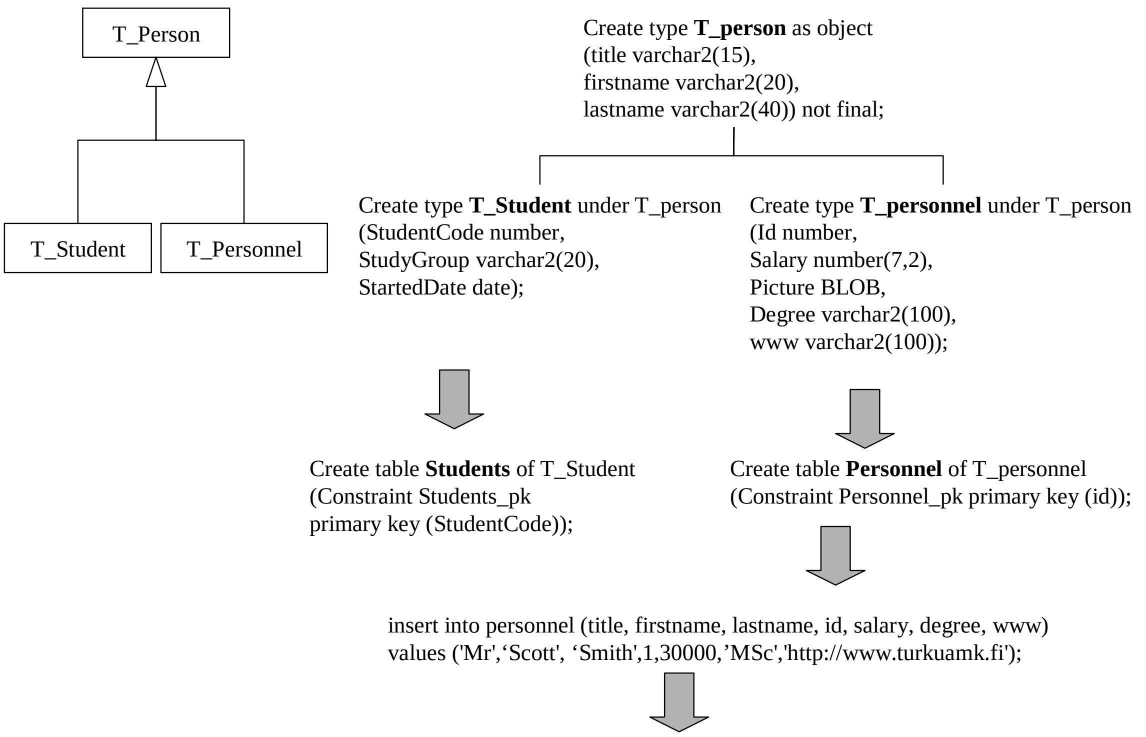
Figure 2.10: Using user-defined types.

```
CREATE OR REPLACE Type Body t_person Is
Member Function Age
    Return Integer
    Is
    Begin
        return months_between(sysdate,birthday)/12;
    End;
End;

INSERT into employees
values (1,
    t_person('Manager', 'Scott', 'Smith','20-JAN-1967',
        t_address('Road 123', 'SALO', '24130','FINLAND‘),
        t_phones(
            t_phone('12202020','Office'),
            t_phone('7282828','Mobile') ),
        empty_blob()
        ),
        4500    )

select e.person.firstname, e.person.lastname, e.person.birthday, sysdate, e.person.age()
from employees e
```

Scott	Smith	20-JAN-67	21-JUL-04	38

Figure 2.11: An example of defining and using Age-method.

As Brown et al. (1999) defined, an object-relational DBMS should also support inheritance at the SQL-level. When the first object-relational DBMS were introduced, inheritance usually was not supported, but the latest object-relational DBMSs have also implemented inheritance in SQL-level. In Figure 2.12, an example of a simple inheritance hierarchy is presented as it is defined in the Oracle DBMS. The T_Person type definition has a ‘not final’ keyword, which means that subtypes can be added to it. The default keyword of a type definition is ‘final’, meaning that no subtype may be defined. An alternative is to use keyword ‘not instantiable’, which implies that there is no constructor (default or user-defined) for the type. Thus, it is not possible to construct instances of this type.



Title	Firstname	Lastname	Id	Salary	Degree	www
Mr.	Scott	Smith	1	30000	MSc	http://www.turkuamk.fi

Figure 2.12: An example of inheritance in SQL-level.

The fourth essential feature of an object-relational DBMS according to Brown et al. (1999) is that is should support a rule system. The rule system:

1) Should support both update and retrieve events.
2) Must be integrated with other object-relational capabilities.
3) Should support execution semantics[7].

Furthermore, the rule system must not loop. A rule system is valuable in most information systems because rules protect the integrity of the data and make maintenance simpler (Brown et al. 1999). The rule system is recognized in the SQL:1999 standard, but typically the concept of trigger is used. Triggers also support the notion of active databases. (Eisenberg et al. 1999) An active database follows the Event-Condition-Action model where events activate actions in the database if conditions are met (Elmasri et al. 2000). A trigger (a rule) is a stored program that is attached to a table or view (Kroenke 2004) and is stored in the database and controlled by the DBMS (Hoffer et al. 2004). Triggers are typically written with a DBMS specific programming language,

[7] {immediate, deferred} and {same transaction, different transaction}.

but in many DBMSs it is possible to use some general programming language as well. For example, Java could be utilized when writing triggers in the Oracle DBMS.

2.5 From object-relational to spatial databases

Geographic information systems are probably the purest and most successful example of exploiting the capabilities of object-relational database management systems. The possibility to exploit object-relational features in this way is important, since a study has estimated that about 80% of data stored in enterprises' databases have geographical features (Gonzales 2000). An object-relational DBMS, together with spatial extensions[8], provides for the management of geographical objects (Goodchild et al. 2001) and allows the user to implement a set of spatial data types and set up operations on those types (Chawla et al. 1999). In these spatial extensions SQL is extended with new functions and procedures that manage spatial data and the DBMS has special features necessary for efficient spatial data management (Rigaux et al. 2002). These functions and procedures allow users to compare geometries and carry out spatial analysis, such as defining distances.

2.5.1 Defining spatial databases

There are many definitions of a spatial database and the basic features it should contain (Guting 1994; Goodchild et al. 2001; Rigaux et al. 2002; Chawla et al. 2003). In general, they emphasise the following features:

1) It can manage spatial data in addition to alphanumeric data.
2) The spatial data types are supported in the data model and in the query language of the spatial database management system.
3) The spatial database management systems offer supporting features like spatial indexing and efficient algorithms for spatial operations.

The first generation of geographic information systems was built on top of proprietary file systems and both spatial and alphanumerical data was stored in the files (Guting 1994). The next phase was to save this alphanumerical data in relational databases, however the spatial data was still stored in files. These solutions did not respect the important data independence principle of databases and led to many problems of security and concurrency control.

[8] For example IBM has Spatial Extender for DB2 and Oracle has Spatial Option for their DBMS.

(Rigaux et al. 2002) Spatial data could also be stored in relational databases, but nowadays it is accepted that they are a poor solution for storing spatial data, because they are only able to handle the basic data types (number, characters) and the storing of whole geometric objects is impossible (Goodchild et al. 2001). In other words, it is not suitable for managing data that is multidimensional with complex structures and behaviours (Sanglu et al. 2000). The representation of spatial data in a relational database requires also a remarkable amount of rows (Rigaux et al. 2002). Moreover, relational databases come off badly with some geometric queries (Goodchild et al. 2001) and they break data independence, since one needs to know the storage structure to perform queries (Rigaux et al. 2002).

For example, it is not possible to store the road presented in Figure 2.13 as one entity in a relational database, because there is not a predefined structure that accepts storage of points. In other words, there is no special data type for storing that kind of an object. Of course, the road with five points could be stored in a relational database using tables and normal data types, but the original idea of storing the road as one object is lost and instead of that six entities are required (Figure 2.14). Another alternative would be the use of a character vector, but here additional predefinition is also required to be able to manage and understand the vector (i.e., the string) correctly. Finally, both of these solutions present problems when querying the roads, because one needs to know for example that there is a points-table, as well where the points of the roads are stored.

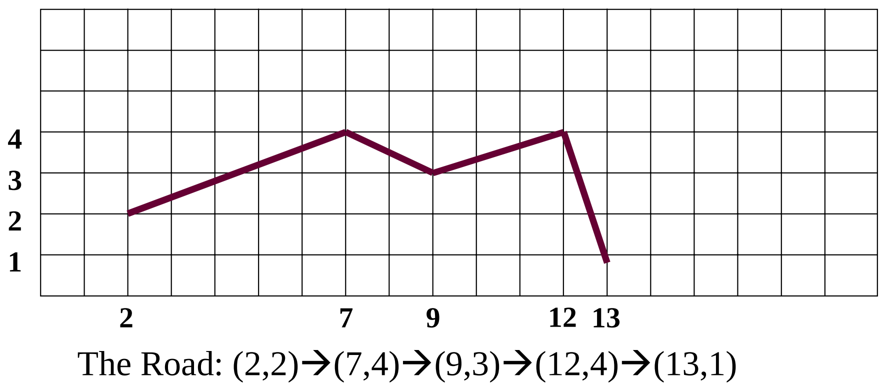
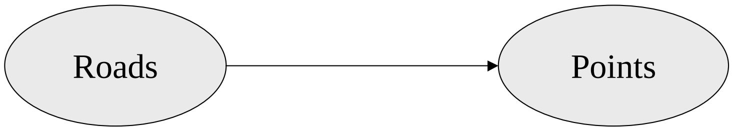


Figure 2.13: Example of a road.



Road ID	Name
E18	Turku-Helsinki

PointID	X	Y	Order	RoadID
11	2	2	1	E18
12	7	4	2	E18
13	9	3	3	E18
14	12	4	4	E18
15	13	1	5	E18

Figure 2.14: Example of storing roads in a relational database.

Object databases were first introduced as a solution to these limitations, but later object-relational DBMSs were extended with spatial extensions. Now these object-relational databases allow for the management of geographical objects with spatial data types and special operations on those types (Goodchild et al. 2001; Chawla et al. 1999). In relational databases the set of data types is fixed, but object-relational databases solve the problems associated with relational databases, since they permit the definition of abstract data types (Chawla et al. 2003). For example, an abstract data type for roads could be defined and the road in Figure 2.13 could be stored as one object.

The spatial DBMSs have their own abstract data types for spatial objects, but although the area is standardized (Carson 2000; ISO 2000; Östensen 2001; OpenGIS 2002; Rigaux et al. 2002), different DBMSs use different constructs for spatial objects (see Table 2.2). These spatial data types are actually comparable to user-defined types, but they are predefined in the DBMS with necessary methods. In recent years XML has been discussed as an alternative for spatial data management and actually a dialect of it has been developed for spatial data. It is called GML (Geometry Markup Language) (see 2.5.4).

Table 2.2: Spatial data types available in four DBMS.

IBM DB2	**MySQL**	**Oracle**	**Postgres**
ST_Point	Point		Point
ST_LineString	LineString		Lseg
ST_Polygon	Polygon		Box
ST_MultiPoint	MultiPoint	Sdo_Geometry	Path
ST_MultiPolygon	MultiPolygon		Polygon
ST_MultiLineString	Geometry		Circle
ST_GeomCollection	Geometrycollection		

2.5.2 Spatial indexing

Additional features for a spatial DBMS are special index structures for indexing spatial data (Guting 1994; Goodchild et al. 2001). The main purpose of spatial indexing is to support retrieving spatial objects that match some query criteria (Guting 1994). The efficient processing of spatial queries requires special access methods relying on a data structure called an index. Indexing reduces the number of rows to be processed, accelerating access to the data. (Rigaux et al. 2002) Database management systems use this index structure to find the relevant data from the database files (Begg et al. 2002). The normal index structures of DBMSs are one-dimensional and poor in indexing spatial data. Consequently, alternative ways of indexing spatial data have been developed and examples of these are Grid, Quadtree and R-tree. (Goodchild et al. 2001) Spatial indexing is usually based on the minimal bounding box-method[9]. There are two different categories of spatial index structures: space-driven and data-driven structures. The space-driven structures partition the embedded space into rectangular cells independently of the objects within them. Grid index and Quadtree are space-driven index structures. The data-driven structure is opposite to space-driven and partitions the space depending on the distribution of the objects. R-tree is a data-driven index structure and it is a natural extension to traditional B-tree. (Rigaux et al. 2002) An example of an R-tree index structure is in Figure 2.15. All spatial DBMSs do not offer all of these index structures; different index structures are supported in different DBMSs (Kontio 2003).

[9] The minimal bounding box–method means that for every indexed object the smallest bounding rectangle is defined and it is used when creating the index structure. The smallest bounding rectangle is used differently depending on the type of the index that is being created. (Rigaux et al. 2002.)

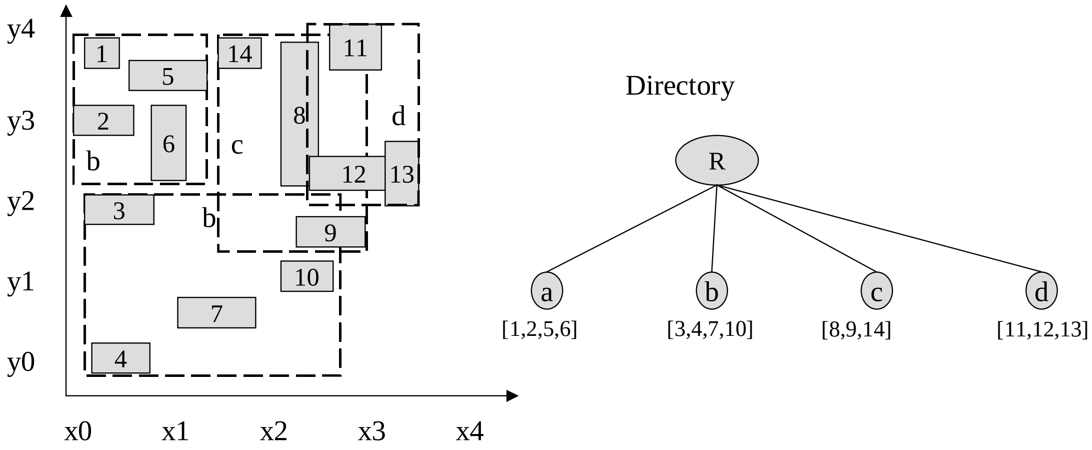


Figure 2.15: Example of a R-tree index.

2.5.3 Standards: ISO/TC 211 and OpenGIS

During the 1980s and at the beginning of 1990s a couple of national and regional communities developed some of their own geographical data management standards, but no internationally agreed standards existed. In 1995 both the ISO/TC 211 and the OpenGIS consortium (OGC) introduced the first internationally-recognized geographical data management standards. (ISO 2000) In 1999 these organizations agreed on co-operation (Östensen 2001). One of the main objectives in these efforts toward standardizing exchange formats and spatial data models was to improve the interoperability of different GISs (Rigaux et al. 2002). There are of course other players in the field of spatial data management like the ISO/IEC JTC1 SC32 organization, which is working on spatial enlargements of SQL and on Geometry-model. (OpenGIS 2003a).

The ISO Technical Committee 211 Geographic Information/Geomatics is responsible for the standardization of geographical data and geomatics (Carson 2000). The TC/211 committee is preparing a standard suite within geographical data in co-operation with the other technical committees of the ISO in this field. The TC/211 committee is divided into many groups, which consider geographical issues from different viewpoints. An interesting one is the sub-project number 19107 (Geographic Information: Spatial Schema), aiming at a data model of geographical entities. (Rigaux et al. 2002.)

The OpenGIS consortium was founded in 1994 (Rigaux et al. 2002). It is a combination of companies aiming at improving the interoperability of geographical techniques (OpenGIS 2002). In general, the OpenGIS consortium is concerned with software specifications, while the ISO TC/211 concentrates more on data standards (ISO 2000). The OGC develops and provides implementation-level technical specifications, which are based on the results of the TC/211 working groups (Östensen 2001). It is also submitting specifications for ISO approval via ISO/TC 211. In addition, the OGC has a specific conformance and testing program for the specifications they develop. (ISO 2000.)

The OGC has adopted the ISO 19107 Spatial schema as the abstract specification for geometry and topology (Östensen 2001). The OGC Simple Features Specification for SQL defines a standard SQL schema, which supports the storage, access and update of simple geographical objects (OpenGIS 1999). This geometry model plays a very central role in these specifications (Chawla et al. 2003). The model is presented in Figure 2.16 with UML notation.

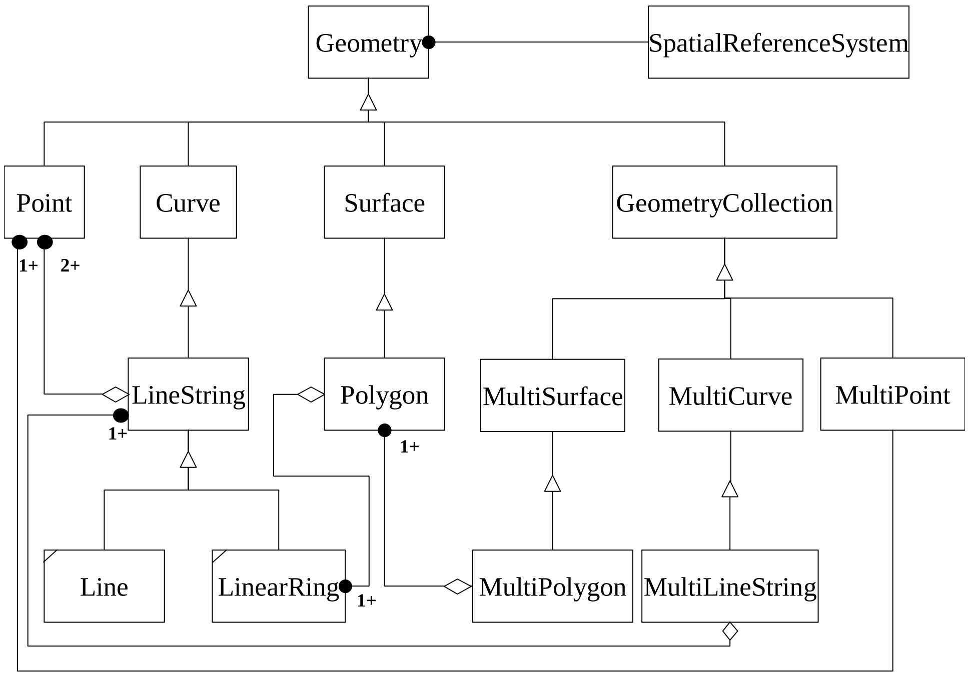


Figure 2.16: Geometry Class Hierarchy (OpenGIS 1999).

The OpenGIS standard also attributes some basic operations to these classes (Chawla et al. 2003) and these operations can be posed in SQL (Chawla et al. 1999). The operations can be categorized in three groups (OpenGIS 1999):

- Basic operations, common to every class.
- Methods for testing spatial relations between geometric objects.
- General operations that support spatial analysis, like the distance of two spatial objects.

The OpenGIS specification is based on SQL92 standard and three alternative implementation models are defined (OpenGIS 1999):

1) SQL92 using numeric SQL types for geometry storage.
2) SQL92 using binary SQL types for geometry storage.

3) SQL92 with geometry types.

The third alternative is easy to implement using SQL:1999 compatible DBMS and user defined Abstract Data Types. The SQL:1999 standard itself defines also spatial types and functions as part of the multimedia enlargement SQL/MM (Goodchild et al. 2001).

2.5.4 GML - Spatial XML

With the development of the Internet and XML it has become possible to define a new common format for spatial data. The OGC has actually developed such a format, calling it GML (Geography Markup Language). GML 3.0 was approved in January 2003 (OpenGIS 2003b). The GML is an XML encoding for the transport and storage of spatial/geographic information, which may include both spatial and non-spatial features (Córcoles et al. 2002). It provides the grammar for encoding geographical content (Lake 2001). The GML definition is based on the geometry model of the OGC (OpenGIS 2003b).

In the Internet XML has helped to distinguish content from representation and now GML is doing the same thing for geographical data (OpenGIS 2003b). This means that the data is in XML format and it can be presented differently just by changing the stylesheet. Similarly, in GML geographical data and its graphical interpretations like maps can be separated. This means GML data can be used for many purposes and in many different devices. (OpenGIS 2003b) Since GML is based on XML, special attention must also be paid as to how databases support XML storage.

2.6 From object-relational to XML-enabled databases

One of the main reasons for the need to combine XML with databases is the fact that information systems are becoming increasingly integrated and XML is considered a respectable alternative to present integration solutions (Lear 1999). XML is seen as a system independent technology for moving information from one information system to another (DeJesus 2000; Linthicum 2001). It also simplifies data transmission (Brown et al. 2001) and offers a standard way to express information, whilst also including necessary metadata (Scardina 2000). Furthermore, XML is replacing HTML in many domains as the next generation standard for data representation and exchange (Cheng et al. 2000; Fong et al. 2003) and becoming the standard for e-Business (Banerjee et al. 2000; Bertino et al. 2001). Therefore, the amount of

XML data exchanged between organizations is expected to grow exponentially, meaning that XML will need to be stored effectively in organizational databases (Champion 2001; Fong et al. 2003). The effective use of XML in electronic business requires systems to store and retrieve XML data (Ha et al. 2001). To summarize, XML is a very relevant technology, because it is well–standardized, widespread and thus also providing economies of scale. However, the fact that information is moving in XML format raises the issue of whether information systems and the underlying database management systems can manage with XML.

Basically, there are two cases when an information system must operate with XML. The first of these is when XML needs to be produced from the data in the database and the second is when information is received in XML format and must be stored in the database (Kontio 2003). There are various ways of achieving both of these outcomes. Different DBMS generations enable different options for managing XML-documents (Figure 2.17). The Figure shows that traditional relational DBMSs allow XML-documents to be both produced and stored in a database, but only structures relevant to the relational model are available.

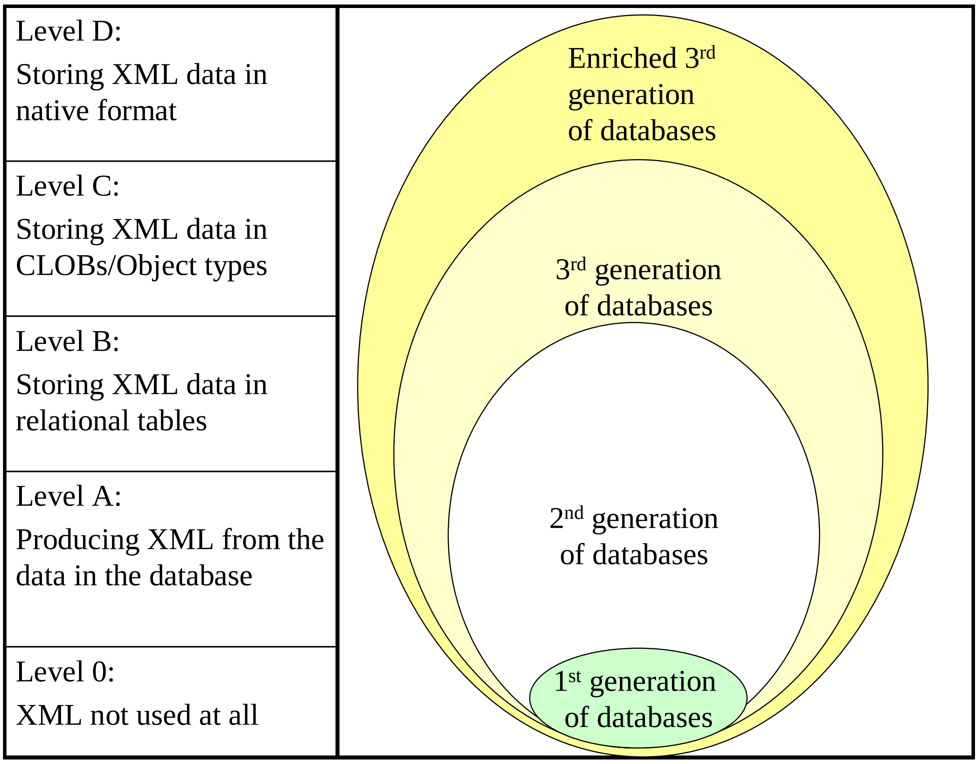


Figure 2.17: How database generations enable XML management.

2.6.1 Different XML documents

XML's specification requires that documents written in the language conform to its basic structural requirements (W3C 2004a). The document must for example start with XML declaration and have only one root element (i.e. all other elements are contained within a single element). A document that conforms to the basic structural and notational rules of XML is considered to be a 'well-formed' XML document. This is an important issue when the documents are managed with computer programs like XML parsers. Another level of XML documents is 'validity', which means that its structure conforms to the definition expressed in a DTD or an XML Schema. An XML document is valid if it has an associated document type declaration (a DTD or an XML Schema) and if it complies with the constraints expressed in it (W3C 2004a).

A DTD defines the valid syntax for an XML document in a particular application area (Begg et al. 2004). It lists, for example, the permitted element names, the combination of elements and the attributes available for each element type. However, with DTD there is no real way to constrain the value of the data in the elements (Needleman 2001). With DTDs an XML parser has no way to validate whether the data is solid. For example <birthdate> element could contain dates of various formats like

- <birthdate>21-AUG-2004</birthdate>.
- <birthdate>2004/08/21</birthdate>.
- <birthdate>21.8.2004</birthdate>.
- <birthdate>2004-08-21</birthdate>.
- <birthdate>Don't want to reveal it</birthdate>.

All these are syntactically correct and the XML parser has no way to prevent any of them from being used. In practice this means that the information system that manages the data has to determine the integrity of the data.

With an XML schema many of the limitations associated with DTDs can be overcome. This defines the data model and establishes data types for the information held in the document (Hoffer et al. 2004). This is a clear improvement on DTDs, because the data types are also denoted. DTDs offer only #PCDATA as a data type, but XML schemas provide many built-in data types (Begg et al. 2004). With an XML schema the above example of <birthdate> element can be solved by limiting the number of date formats that can be approved.

From a data management point of view all XML documents should be well-formed and valid, but before looking at the various ways to work with XML documents and databases it is important to recognize different types of XML-documents. The nature of an XML-document is especially significant when their storage alternatives are discussed. One significant distinguishing factor is the integrity of the data to which the XML document must conform. For example, when an XML document represents an order, it has very high integrity requirements. In these cases it is essential that the structure of the XML document is predetermined with a DTD or preferably with an XML Schema. With an XML schema we can achieve higher integrity using the built-in data types.

XML-documents with high integrity requirements are typically called ‘data-centric’ XML documents since their meaning depends only on the structured data represented inside them (Bertino et al. 2001). Their structure should be regular and the content homogenous, but the order of elements is not significant (Bertino et al. 2001; Renner 2001; Bourret 2002a). An example of a data-centric XML-document could be an order received in XML format (see Figure 2.18). Typically, they are used in expressing product catalogues, invoices, flight schedules and other similar well-structured cases. Data-centric XML-documents occur typically in Business-to-Business applications where information systems communicate with each other (Bertino et al. 2001) and human processing is limited or non-existent (Bourret 2002a).

```
<?xml version='1.0' encoding='windows-1252'?>
<Order ID="82728">
 <From>
  <Company>Firm A</Company>
  <Address>
   <Street>Helsingintie 2</Street>
   <Zipcode>00001</Zipcode>
   <Area>Helsinki</Area>
   <Country>FINLAND</Country>
  </Address>
  <Contact>
   <Person>Scott Allen</Person>
   <Phone>+358 50 484838</Phone>
  </Contact>
 </From>
 <orderdate>18.5.2004</orderdate>
 <Orderlines>
  <Line>
   <Product>82828</Product>
   <Price>34</Price>
   <Amount>2</Amount>
  </Line>
  <Line>
   <Product>28376</Product>
   <Price>249</Price>
   <Amount>1</Amount>
  </Line>
  <Line>
   <Product>11112</Product>
   <Price>1920</Price>
   <Amount>1</Amount>
  </Line>
 </Orderlines>
</Order>
```

A data-centric XML-document/
Data with high requirements of integrity

- Regular
  - The structure of the XML-documents is always similar in certain application areas and, typically, standardized with DTDs or XML schemas
- Homogenous
  - The content is expressed similarly throughout the XML-document
  - For example, the Product-tag always contains a product number, not a product name and price is always expressed in Euros
- Order of elements
  - Order of line-elements could be changed without disturbing the XML-document’s information

Figure 2.18: An example of a XML-document containing data with high integrity requirements.

When an XML document contains data with low integrity requirements we can refer them as ‘document-centric’ XML-documents. At first sight they look quite similar to data-centric XML-documents. However, there are three main differences: the regularity, the structure and the order of the elements. Document-centric XML-documents are common in Business-to-Customer communication, where the documents are designed for human consumption like web pages (Bourret 2002a). A document-centric XML-document is irregular, heterogeneous and the order of elements is significant like Figure 2.19 shows (Bertino et al. 2001; Renner 2001; Rosenthal et al. 2001; Schöning 2001; Bourret 2002a). Typically, document-centric XML-documents are found in magazine articles, books, press releases and XHTML-based web sites. A third type of XML-document combines the two. In this case the XML-document has both a data-centric and a document-centric part.

```
<?xml version='1.0' encoding='windows-1252'?>
<Doc>
<Parag>
Document-centric XML-documents are common in
Business-to-Customer communication, where the
documents are designed for human consumption
like web pages <Reference>(Bourret 2002a)</Reference>.
<Definition>A document-centric XML-document is
irregular, heterogeneous and the order of
elements is significant </Definition> like
<Link>Figure 21</Link> shows
(<Reference>Bertino et al. 2001</Reference>;
<Reference>Renner 2001</Reference>;
<Reference>Rosenthal et al. 2001</Reference>;
<Reference>Schöning 2001</Reference>;
<Reference>Bourret 2002a</Reference>). Typically,
document-centric XML-documents are found in
magazine articles, books, press releases and
XHTML-based web sites.
</Parag>
</Doc>
```

A document-centric XML-document/ Data with low requirements of integrity

- Irregular
  - The structure of the XML-document is not pre-determined
  - A content model might exist, but the content is not defined in detail
- Heterogeneous
  - Different XML-documents do not have similar content
  - The document contains a large amount of mixed content such as pure text and sub-elements
- Order of elements
  - This is important. For example, the order of the paragraphs is significant

Figure 2.19: An example of a XML document containing data with low integrity requirements.

2.6.2 Producing XML documents

There are at least five different ways to produce an XML-document:

- Writing a special database program.
- Using SQL-extensions.
- Using special programming interfaces and techniques.

- Reading an XML document from a CLOB-attribute of a database table.
- Reading an XML document from a native XML-attribute of a database table.

The first alternative is to write a special program that generates the XML document according to the rules of XML specification. This solution does not need anything special from the database management system. The only requirements are to be able to connect to the database, have access to the data and naturally some programming skills.

The second alternative uses SQL-extensions for XML data management of the DBMS. Some database management systems have adopted the SQL/XML standard, which defined many easier ways to produce XML from a database (Eisenberg et al. 2003; Gennick 2003; ISO 2003). Three of the SQL/XML functions are presented in Table 2.3 and Figure 2.20 gives an example of the XMLElement-function.

Table 2.3: Examples of SQL/XML functions (Eisenberg et al. 2002).

Function	Description
XMLElement	A function that translates a relational value into a XML element
XMLForest	A function that creates a list of XML element from a list of relational values
XMLAgg	A function for grouping and composing XML content in Group by queries

```
Select XMLElement("Course", course)
From Coursegrades;
```

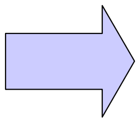
```
<Course>Database Design</Course>
<Course>Programming 1</Course>
<Course>Networks</Course>
…
```

Figure 2.20: An example of XMLElement function.

A third way to produce XML from a database is to offer special programming interfaces and techniques for XML data management. Usually these techniques exploit some XML-related standard techniques like XSLT style sheets and XPath-query capabilities. One example of this kind of technique is the XSQL-technique offered by Oracle DBMS (Munch 2000).

With the templates defined in *.xsql files, the result of a basic SQL query is presented as an XML document. The result can easily be transferred in different formats using XSLT stylesheets. An example of a simple XSQL file and the produced XML document is presented in Figure 2.21.

```
<?xml version="1.0" encoding='windows-1252'?>
<xsql:query connection="demo"
    rowset-element="Departments"
    row-element="Dept"
    xmlns:xsql="urn:oracle-xsql">
    SELECT * FROM DEPT
</xsql:query>
```



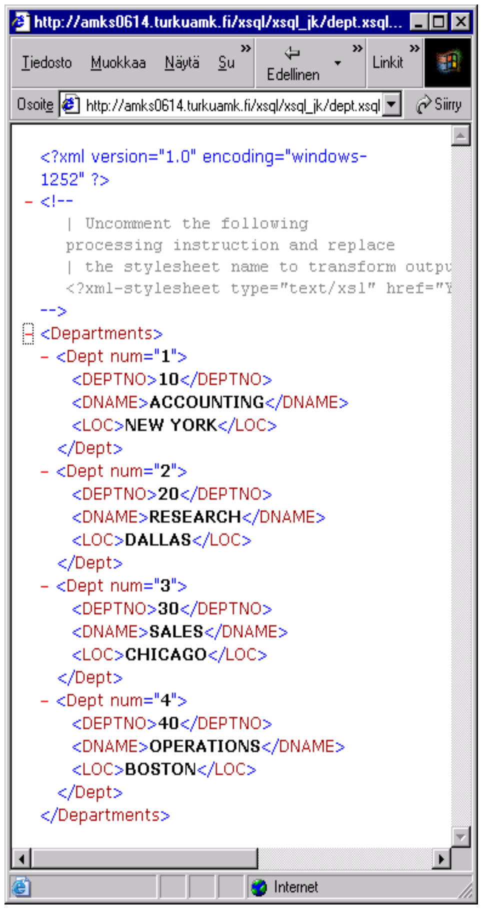


Figure 2.21: XSQL example.

The fourth alternative is to read an XML document from a CLOB-attribute of a database table. For example, received orders could have been archived as XML documents in a CLOB-attribute Order_xml in Order_archive-table. When the order is needed from the archive it can easily be displayed with a simple SQL query (see Figure 2.22).

Finally, the fifth alternative is to display an XML document from a native XML-attribute of a database table. The query used here could be exactly the same as in the previous case although the attribute containing the XML document is different. Therefore an example of this case is not presented.

ORDER_ARCHIVE

Order_id	Varchar2(10)
Order_xml	CLOB
Timestamp	Date

```
SELECT *
FROM order_archive
WHERE order_id='82728'
```

Order_id	Order_xml	Timestamp
82728	<?xml version="1.0" encoding="windows-1252"?> <Order ID="82728"> <From> <Company>Firm A</Company> <Address> <Street>Helsingintie 2</Street> <Zipcode>00001</Zipcode> <Area>Helsinki</Area> <Country>FINLAND</Country> </Address> <Contact> <Person>Scott Allen</Person> <Phone>+358 50 484838</Phone> </Contact> …	22-AUG-04

Figure 2.22: Displaying an XML document from a CLOB-attribute.

2.6.3 Storing XML documents in relational DBMS

Storing the content of an XML document in a relational database means that the content is transformed into rows and columns. The integrity of the data should be checked before storing the content of the XML document in a relational database. A basic requirement is the availability of an XML schema for the XML document, but additional checking will probably be needed. Anyway, the storage of XML documents in a relational DBMS typically means that the original hierarchic structure and order of elements in the XML document are destroyed (DeJesus 2000). For data-centric XML documents this is not a problem, since the basic idea with this type of XML documents was to move data between information systems. However, the order of the elements is vital for document-centric XML documents and it should not be destroyed (Schöning 2001). Therefore a number of writers have suggested ways of preserving the original structure storing XML documents using relational tables: set of tables (Chen et al. 2002), edge-strategy (Florescu et al. 1999) and nested sets model (Edwards 2003). An example of using Edge-strategy is presented in Figure 2.23. Using the SQL:1999 feature of recursive query it is possible to produce a listing that is quite similar to the XML document. An example of this is presented with Oracle DBMS in Figure 2.24. However, relational database management systems are not considered to be the best solutions for working with XML documents (Chen et al. 2002) although they offer reliability, scalability, tools and performance (Khan et al. 2001).

```
<?xml version="1.0"?>
<Dept>
   <Name>Marketing</Name>
   <Worker>
      <WId>1110232</WId>
      <Name>Jack Stone</Name>
      <Projects>
         <Project>A11</Project>
         <Project>A14</Project>
      </Projects>
   </Worker>
   <Worker>
      <WId>1110245</WId>
      <Name>Michael Fox</Name>
      <Projects>
         <Project>A11</Project>
      </Projects>
   </Worker>
</Dept>
```

SourceID	Tag	Ordinal	TargetID	Data
1	Dept	1	2	Null
2	Name	1	0	Marketing
2	Worker	2	3	Null
2	Worker	3	5	Null
3	Wid	1	0	1110232
3	Name	2	0	Jack Stone
3	Projects	3	4	Null
4	Project	1	0	A11
4	Project	2	0	A14
5	Wid	1	0	1110245
5	Name	2	0	Michael Fox
5	Projects	3	6	Null
6	Project	1	0	A11

Figure 2.23: Storing XML in Relational DBMS with Edge-strategy.

```
SELECT
LPAD(' ', 3 * LEVEL-3)||tag AS TAG,
DATA
FROM   XMLDOCS
START WITH SOURCEID=1
CONNECT BY PRIOR
TARGET_ID = SOURCEID
```

```
TAG              DATA
Dept
   Name          Marketing
   Worker
      Wid        110232
      Name       Jack Stone
      Projects
         Project A14
         Project A11
   Worker
      Wid        110245
      Name       Michael Fox
      Projects
         Project A11
```

Figure 2.24: An example of producing a hierarchical view from data stored with Edge-strategy.

2.6.4 Storing XML documents with third generation DBMSs

The third generation of database management systems offers new alternatives for storing XML documents. For example, an object-relational DBMS allows for the storing of an XML document in one piece, thus preserving its original structure. However, the database management system must be able to manage character large objects (CLOBs) in order to achieve this (Bertino et al. 2001). Actually CLOBs are seen as an ideal solution for storing XML documents that are unknown, irregular or dynamic (Banerjee 2002). For example, the previous document-centric XML document could be stored in a CLOB (see Figure 2.25). The drawback of using CLOBs is that the details of an XML-document are not easily accessed (Munch 2000; Champion 2001; Hohenstein 2003). Also, object types and collections have been suggested for storing XML documents (Ha et al. 2001; Varlamis et al. 2001). However, without native XML storage possibility, additional mapping and transforming, a remarkable amount of work is needed (DeJesus 2000; Bertino et al. 2001; Darrow et al. 2001), meaning actually a new impedance mismatch situation (Schöning 2001). Anyhow, object structures will require the XML-document to be modelled in order that it maps into the object-relational database management system (Ha et al. 2001). The definition of the XML document is typically presented with a DTD or an XML schema. However, a tool is still needed to do the actual operation, i.e. placing the XML-document into the object structures (Brown et al. 2001), unless this functionality is provided by the DBMS.

```
Create table XMLdocs(
Docname varchar2(200) primary key,
Xmldoc CLOB,
Timestamp date)
```

```
insert into XMLdocs
values('Document',
'<?xml version="1.0" encoding="windows-1252"?>
<Doc>
<Parag>
Document-centric XML-documents are common in
Business-to-Customer communication, where the
documents are designed for human consumption
like web pages <Reference>(Bourret 2002a)</Reference>.
<Definition>A document-centric XML-document is
irregular, heterogeneous and the order of
elements is significant </Definition> like
<Link>Figure 21</Link> shows
(<Reference>Bertino et al. 2001</Reference>;
<Reference>Renner 2001</Reference>;
<Reference>Rosenthal et al. 2001</Reference>;
<Reference>Schöning 2001</Reference>;
<Reference>Bourret 2002a</Reference>). Typically,
document-centric XML-documents are found in
magazine articles, books, press releases and
XHTML-based web sites.
</Parag>
</Doc>', '18-JAN-2004')
```

Figure 2.25: An example of CLOB-storage.

Although object-relational DBMSs offered new possibilities for managing XML documents, they are still unnatural and tedious solutions. They do not allow for XML to be handled efficiently (Benham 2003). The most natural way to manage all kinds of data is to have your own data type for it. This makes it possible to manage data in XML documents in the same way as numbers and characters have been managed in the relational DBMS. Also, the DBMS typically provides its own services and functions for the management of a certain data type. The XML-enriched third generation of DBMSs offer these native XML data management capabilities. For example, Oracle offers its own XMLType data type for managing XML documents and either CLOBs or XML schema based object structures are used. However, if the XML schema is defined, the DBMS will follow it to generate the necessary object structures for storing XML documents automatically.

2.6.5 Comparing XML-enabled and native XML DBMSs

The XML capable database management systems are usually classified as native XML and XML-enabled DBMSs. It is estimated that XML-enabled DBMSs are going to dominate the database solutions of application areas that manage data-centric XML-documents (Bourret 2002c), while native XML DBMSs will succeed in content management markets where typically document-centric XML-documents are managed (2001). Native XML DBMS differs from XML-enabled DBMS in three ways (Bourret 2002b):

1) It maintains the physical structure of the XML-document.
2) It can store the XML-document without any schema or document type description.
3) Its only interface to data is XML and related technologies.

This final point distinguishes for example the newest versions of Oracle and IBM from being a native XML DBMS since it is possible to manage XML data and traditional data both with SQL and XML tools. Some examples of native XML DBMSs are Tamino, dbXML, eXcelon and X-hive/DB (DeJesus 2000). Many writers have evaluated the suitability of XML-enabled and native XML DBMS for the management of XML-documents (Carr 2001; Darrow et al. 2001; Dyck 2001; Tillet 2001; Bourret 2002b; Chen et al. 2002; Fong et al. 2003; Schmauch et al. 2003). The evaluations are summarized in Table 2.4.

Table 2.4: Comparison of native and XML-enabled DBMSs.

	Native XML databases	XML-enabled (object-) relational databases
Strengths	• No XML conversation needed • Better performance with XML data • Better for B2B transactions and XML applications that need to be extended or modified • Better as an aggregate repository for XML data • Good for unstructured data	• Suited for transactional applications • One integrated platform for handling XML and other data • One management platform • SQL is still relevant • Good for structured data
Weaknesses	• Separate management of XML and other data • Consumes more disk quota • Additional costs	• Requires some XML conversion • Slower performance with XML data

2.7 From a customer database to Customer Relationship Management

At the heart of a customer relationship management (CRM) system is an integrated database (see Figure 2.26) (Fowler 2000).

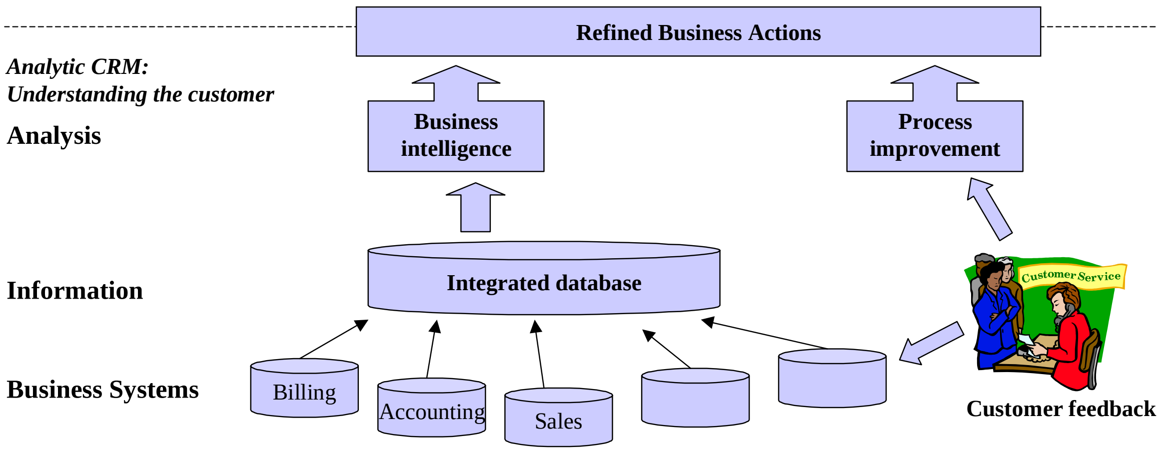


Figure 2.26: CRM as an integrated system (Dyché 2002).

A single customer database is valuable, but an integrated database together with a CRM system offers the tools and equipments for truly managing customer relationships (Clark et al. 2001). CRM has been defined as an integrated information system for managing the interactions between a customer and the organization (Anderson 2001; Harding 2001). However CRM is not only a technology, rather it is a strategy that is supported by the technology (Coffee 2002; Zaino 2003). A major challenge is to integrate data sources and information systems (Harding 2001).

Moving to CRM requires the successful integration of technology and confluence of techniques. Implementing CRM should be regarded as a business-oriented project rather than an IT-oriented project (Eberhardt 2001; King 2002). With the right approach CRM has the chance to improve business results significantly (Eberhardt 2001), but this requires a clear vision and strategy (King 2002; Lee et al. 2003; Tehrani 2003). At design phase all existing CRM-related processes within the organization should be identified and present information systems should be evaluated (Lee et al. 2003). Geographic information should be integrated into the CRM-system, since all main CRM-processes have the ability to do this (Morris et al. 2000). The requirements for integration and utilizing geographic information mean that the underlying database management systems have to provide more capabilities than a pure relational database. In practice, this means that exploiting a CRM system requires a database from the third generation.

2.7.1 Defining CRM

CRM has origins in the 1990s, when companies began to understand the value of a detailed customer database. Actually there are three preliminary phases to CRM that highlight the relationship between company and customer: direct mail marketing, segmented marketing and database marketing. CRM is however the first solution to offer means and tools for ‘managing’ customers. (Clark et al. 2001) We can define Customer Relationship Management as:

- A dynamic process of managing a customer-company relationship so that customers elect to continue mutually beneficial commercial exchanges and are dissuaded from participating in exchanges that are unprofitable to the company (Bergeron 2002).
- The infrastructure that enables the delineation of and increase in customer value, and the correct means by which valuable customers are motivated to remain loyal (Dyché 2002).

- A business strategy that is designed to optimise profitability, revenue and customer satisfaction (Fowler 2000).
- Having the technology to provide an integrated view of all customer interactions and changing the corporate culture to leverage this information to maximize the benefits to the customer and the company (Anderson 2001).
- An integrated information system that is developed from disparate data sources and is used to plan, schedule and control all client and prospect interactions within an organization (Harding 2001).

When talking about CRM, e-CRM is often mentioned and this expands the traditional CRM techniques by integrating technologies of new electronic channels, such as web, wireless and voice technology, and combines them with e-Business applications into the overall enterprise CRM strategy (Lee et al. 2003). Coffee (2002) has argued that CRM is now doing the same to enterprise systems as spreadsheet did to desktop computing in the 1980s.

CRM is often touted as a means of improving customer satisfaction, which in turn improves long-term customer loyalty. It is also part of a public relations campaign to improve the company’s image. (Bergeron 2002). The focus of CRM is on learning more about customers and using that knowledge to refine every interaction with them (2000). To manage customer expectations it is essential to identify all of the customer touch points, i.e. all the places and means the customer can interact with the company. Typically, these might include email, fax, post, media, personal contacts, phone and Internet. (Bergeron 2002) The CRM solution should encompass this entire network of touch points and gather information from each transaction. This information is then stored in a database, which is the heart of the entire CRM-system (Fowler 2000) and a mandatory element for a company to serve its customers well (Dyché 2002). This blended multi-channel approach to CRM enables a business to respond to the needs and desires of its customers, whilst also directing its less valued customers to lower cost service channels (Anderson 2001). To satisfy customer needs, companies have to maintain consistency across all interaction channels (Lee et al. 2003). All communications should also be structured in such a way that the preferences and information-seeking behaviours of individual customers are reflected in the way information is offered to them (Clark et al. 2001).

2.7.2 Implementing CRM

When starting a CRM project, a company should ask itself a) what business value is expected from the CRM and b) will these values lead to quantifiable improvements in customer retention and satisfaction (Dyché 2002).

CRM is difficult to execute successfully without first establishing a reliable customer retention strategy (CRS). After a solid CRS is established it becomes possible to think about broadening the client base and launching a CRM project. During the strategy formulation a profitability review of all products and services should be performed and a value code assigned to each customer. (2003) When dealing with CRM one should be aware of the role of technology, the importance of self-knowledge, the finite nature of customer-company relationships and the need for consistency in quality of service (Bergeron 2002). As mentioned earlier, CRM is not about technology, but is instead a strategy whose implementation is supported by technology (Coffee 2002; Zaino 2003). Consequently, the standard off-the-shelf products are not adequate solutions, rather most CRM systems are custom-built (Tehrani 2003). It is essential to customize CRM software according to current processes, rather than attempting to reengineer everything in order to fit some theoretical idea (James 2003). Users also prefer using a CRM system that is tailored to their specific needs (Tehrani 2003).

2.7.3 Critical success factors

Many factors affect the success of CRM. It also makes sense to list and consider them, since over 50 % of all CRM-implementations are considered to be failures at the end (James 2003). Usually, if a company is not able or not willing to change its business processes and culture to become more customer-oriented, the CRM-project will fail (Coffee 2002). However, when CRM is approached correctly, it has potential to improve business results significantly (Eberhardt 2001). It is possible to avoid CRM failure with careful design, which takes in account the organization's core competencies, technological infrastructure and customer needs (Tehrani 2003). Careful design also covers evaluating software and services from several different vendors and finally the selection of a vendor (Fowler 2000; Tian et al. 2002; Lee et al. 2003).

The CRM toolset should support company processes (King 2002) and the tools should be able to drill down to the biggest problems in current processes (James 2003). A company's real CRM needs should be analysed carefully and quantitative metrics should be set (Eberhardt 2001; Coffee 2002; King 2002). In practice this defines a clear CRM vision and strategy (Lee et al. 2003). The lack of CRM strategy, one that is not integrated with the business strategy or a lack of adherence to it all increase the risk of failure (Elmasri et al. 2000; King 2002). Customer expectations and hopes should also be collected and assessed during the design phase.

When implementing a CRM system the project must have full support from senior staff in the company (Lee et al. 2003), management (King 2002) and backing from the whole organization (Fowler 2000; Harding 2001). The role of employees is also central to the CRM project (Eberhardt 2001) and they should participate in the development process from start to finish (Dyché 2002). Furthermore, it is also important to set goals and be aware of the project's implications for the whole company (Tehrani 2003).

CRM should be integrated with all other activities (Coffee 2002) and integrating various data sources and information systems is a particularly big challenge (Harding 2001). Special attention should also be given to the integration of processes in the various touch points (Eberhardt 2001; King 2002). The whole CRM infrastructure should be designed and implemented carefully without hurry. (Eberhardt 2001; Tehrani 2003). Implementation should be a company-wide effort, but at the beginning the focus can be in some sub-areas (Fowler 2000; James 2003). A CRM project is not a one-off task, but rather a long lasting process (Fowler 2000; King 2002). In fact, it is an effort to move towards a more customer-centred way of doing business (Eberhardt 2001).

After implementation the company must be willing to change its processes and workings according to the metrics and customers' feedback (Eberhardt 2001; Coffee 2002). CRM systems are very complicated and problems often arise as a result of expectations that are too high (Fowler 2000; Tehrani 2003). However, with CRM an organization can concentrate on substantial customers (Coffee 2002).

2.8 Improving reporting with a Data Warehouse

The term data warehouse was first introduced as a subject-oriented, integrated, non-volatile, and time-variant collection of data that supports management decisions (Inmon 1992). A simpler definition says that it is a store of enterprise data that is designed to facilitate management decision-making (Kroenke 2004). A data warehouse differs from traditional databases in many ways, including in its structure and required functionalities (Elmasri et al. 2000). The aim is to integrate all corporate information into one repository where the information is easily accessed, queried, analysed and used as a basis for reports (Begg et al. 2002). A data warehouse provides decision support to organizations with the help of analytical databases and On Line Analytical Processing (OLAP) tools (Gorla 2003). It receives data from operational databases on a regular basis, which is added to the existing information. It contains both detailed aggregated data and also summarized data to speed up the queries. A data warehouse is typically organized in smaller units called data marts that support the specific analysis needs of a department or a business unit (Bonifati et al. 2001). Figure 2.27 describes a data warehouse environment and shows how its various components are connected with each other.

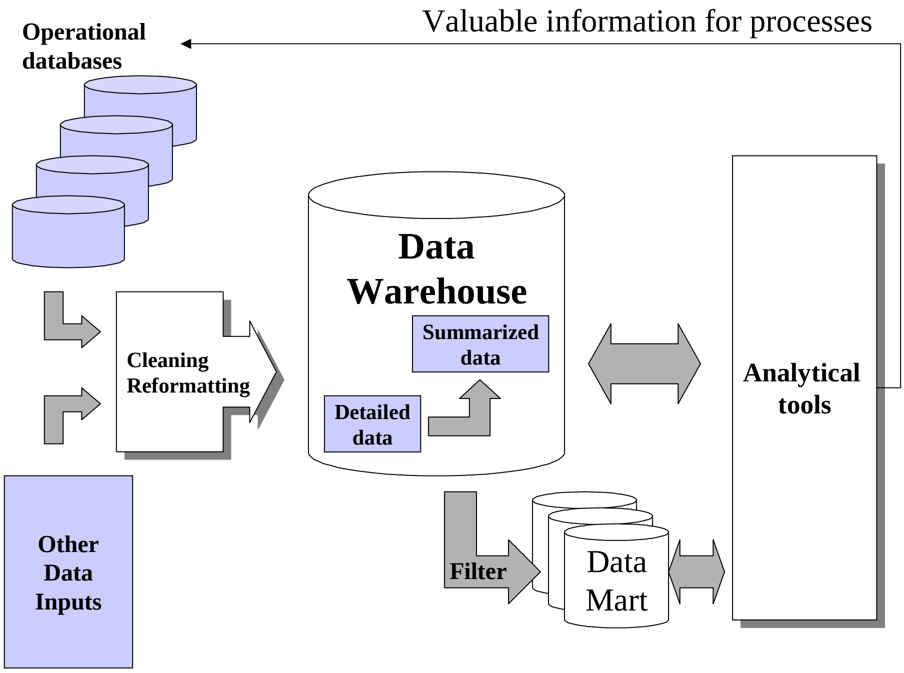


Figure 2.27: Data warehouse environment.

An enterprise data model is also a good starting point for building a data warehouse (Bonifati et al. 2001). Actually, building a data warehouse is an integration project, since the data in operational information systems is integrated into a single data warehouse database (Hasselbring 2000). The design of a data warehouse requires special modelling (Busborg et al. 1999) and several data model designs for a data warehouse have been suggested. Examples are StarER (Busborg et al. 1999) and Dimensional Fact Model (DFM) (Golfarelli et al. 1998).

Designing a data warehouse requires quite different techniques compared with designing an operational database (Golfarelli et al. 1998). Creating a data model for a data warehouse is seen as one of the most critical phases in the development process of a data warehouse (Bonifati et al. 2001). To begin with, the designer needs to understand the content of the operational information systems, the interconnections between them and the equivalent entities (Blackwood 2000). In practice this means studying the data models of the operational databases and developing an integrated schema to enhance data interoperability (Bonifati et al. 2001). The data modelling of a data warehouse is called 'dimensionality modelling' (Golfarelli et al. 1998; Begg et al. 2002). Dimensional models were developed to support analytical tasks (Loukas et al. 1999). Dimensionality modelling concentrates on facts and the properties of the facts and dimensions connected to them (Busborg et al. 1999). 'Facts' are numeric and quantitative data related to the business and 'dimensions' describe different dimensions of the business (Bonifati et al. 2001). Fact tables contain all business events to be analysed and dimension tables define how to analyse fact information (Loukas et al. 1999). The result of the dimensionality modelling is typically presented in a star model or in a snowflake model (Begg et al. 2002). 'Multi–Dimensional' (MD) Schema is a more generic term that is used to collectively refer to both schemas (Martyn 2004). When a star model is used the fact-tables are normalized, but dimension-tables are not. Normalization of the dimension-tables turns the star-model into a snowflake-model. (Bonifati et al. 2001.)

Ideally an information system like a data warehouse should be accurate, fast and user-friendly (Martyn 2004). Accuracy is especially important in data warehouses to ensure that decisions base on correct information. Actually, it is estimated that 30 to 50 percent of information in a typical database is missing or incorrect (Blackwood 2000). This emphasizes the need to pay attention to the quality of the source data in operational databases (Finnegan et al. 2000). A problem with multi-dimensional schemas is that when the business environment changes, the evolution of MD schemas is not as manageable as it is with normalized schemas (Martyn 2004). In addition, Gardner (1998) says that any architecture not based on the 3rd normalized form can cause the failure of a data warehouse project. On the other hand however, a dimensional model provides a better solution for a decision support application than a pure normalized relational model (Loukas et al. 1999). All of the above is actually related to efficiency since a large amount of data is processed during analysis. Typically this is a question about the number of joins that are required in the database level. Usually a star schema is the most efficient design for a data warehouse, since the de-normalized tables need fewer joins (Martyn 2004). However, recent developments in storage technology, access methods (like bitmap indexes) and query optimisation are indicating that performance with the third normalized form should be tested before moving to multi-dimensional schemas (Martyn 2004). From this 3rd normal form schema a natural step towards multi-dimensional schema is to use 'denormalization', which will also support both efficiency and flexibility issues (Finnegan et al. 2000). Still there is a possibility to define necessary SQL views on top of the 3rd normalized form schema without denormalization (Martyn 2004). Finally, the issues of physical and logical design should be separated: physical is about performance and logical is about understandability (Kimball 2001).

OLAP (Online Analytical Processing) tools access data warehouses for complex data analysis and decision support activities (Kambayashi et al. 2004). The architecture of the underlying database of the data warehouse categorizes the different analysis tools (Begg et al. 2002). Depending on the schema type terms, Relational OLAP (ROLAP), Multi-Dimensional OLAP (MOLAP) and Hybrid OLAP are used (Kroenke 2004). ROLAP is a preferable choice when a) the information needs change frequently, b) the information should be as current as possible and c) the users are sophisticated computer users (Gorla 2003). The main differences between ROLAP and MOLAP are in currency of data and in data storage processing capacity. MOLAP populates its own structure with the original data when it is loaded from the operational databases. ROLAP analyses the original data and the users can drill down to the unit data level. (Dodds et al. 2000) ROLAP typically utilize the SQL extensions like CUBE and ROLLUP, which are offered for example in Oracle DBMS (Begg et al. 2002).

2.9 New requirements for database design

Good database design is crucial to the successful implementation of databases (Ram et al. 1998; Browne et al. 1999). Traditional database design consists of three phases: conceptual, logical and physical (see Figure 2.28). The conceptual level model is critical in the overall design process (Barron et al. 1999; Noah 2000; Johnson et al. 2002). Both the conceptual and logical database design phases are independent of the final database management system. However, the logical schema is dependent on the selected data model that the DBMS will be using. Therefore, it is especially important for the logical schema to support the requirements of new application areas. Anyway, the conceptual schema is still the general description of the data in the specific application area, regardless of the next design phases.

To build better information systems it is necessary to improve the quality of the conceptual data models (Calero et al. 2001). It is therefore important to have a flexible model, but this might also cause problems (Wedemeijer 2001). Strategic data planning provides a wider perspective on organizations’ data and is a basis for identifying and implementing an integrated set of information systems that will meet the needs of the business (Goodhue et al.

1992). It is also an important aspect of data management, which aims to manage data as a corporate resource (Shanks 1997).

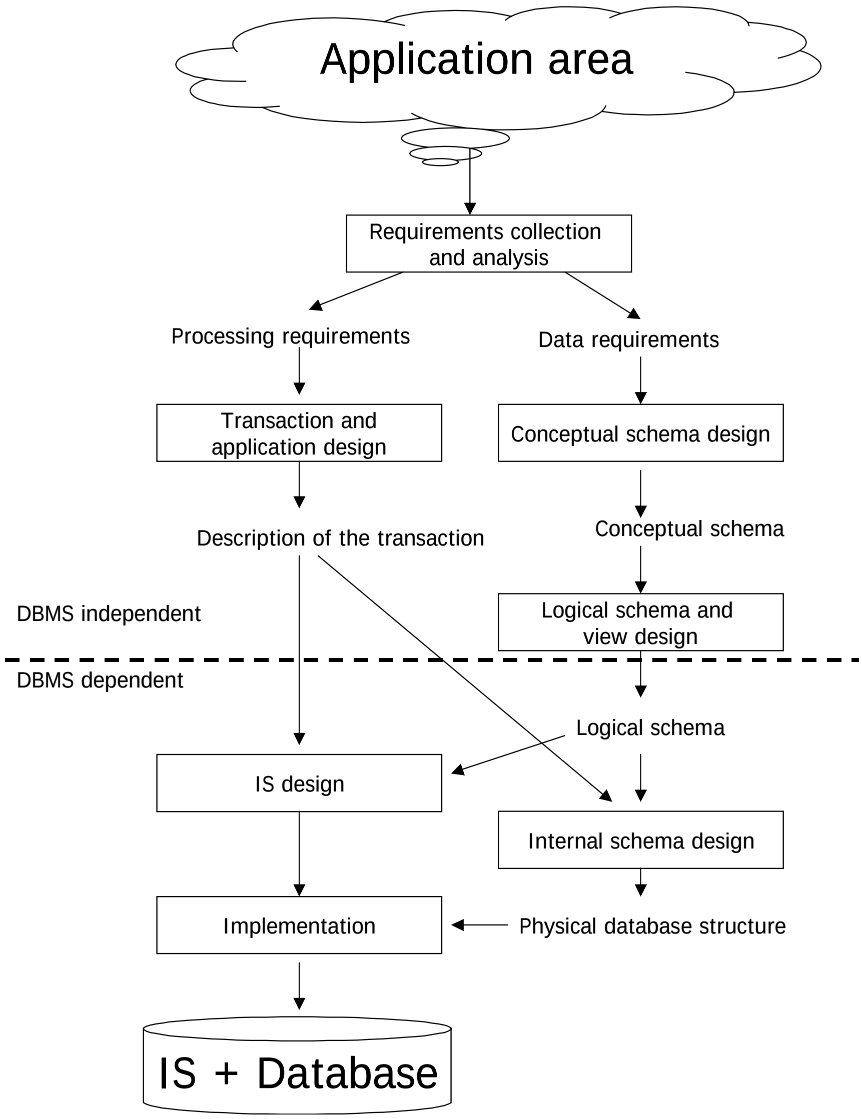


Figure 2.28: Phases of the database design.

Many studies have also emphasized that the integration of information systems is best done at the conceptual level (Batini et al. 1986; Barron et al. 1995; Patel et al. 1998) and one step towards doing this is constructing an enterprise level data model (Busichia et al. 1999). Integrating data is also one key challenge in managing e-business (Fremantle et al. 2002). In addition, surveys of management information systems issues consistently rate developing information architecture as being very important (Shanks 1997). Integration of information systems also supports the effectiveness of business processes (Hasselbring 2000) and the economic benefits of computing rely heavily on the integration of disparate systems (Sutherland et al. 2002).

However, the evolution of database management systems and the new data requirements have increased pressures on data modelling. Thus data modelling should be able to model the new structures (objects, collections, XML documents) but it should also support other functions of the business like integration, data warehousing and e-Business. These pressures are particularly relevant at the logical schema level, because it is designed to support the data model that is selected as the basis for the final DBMS. Typically, the final DBMS is based on a relational data model. Thus a traditional entity in a conceptual data model would lead to the creation of a relational table during the design process. However, the modelling began with the conceptual model, which was transformed into a logical model. Therefore these requirements have influenced the conceptual modelling as well. For example, how should an XML-document or a collection of objects be modelled? Or how could we model a collection of objects that should be embedded in an entity, or how should application-specific and user-defined types be drawn in the model? Therefore it seems obvious that the traditional modelling approach is no longer sufficient for modelling application areas in the third generation of database management systems. This is why alternative suggestions for data modelling and design process have emerged.

Representing complex objects and complex relationships in relational databases means that the objects must be broken down into a large number of rows (Cavero et al. 2004). As described earlier, object-oriented databases can solve this problem, but there must be a methodology for designing them. Figure 2.29 presents one suggested methodology for object-relational database design (Cavero et al. 2004).

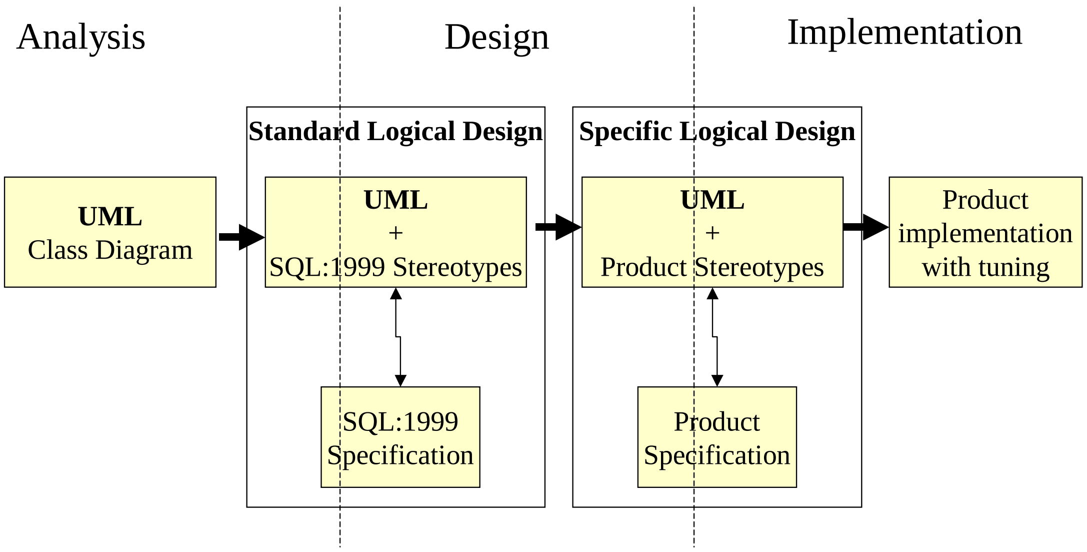


Figure 2.29: Object-relational database design methodology (Cavero et al. 2004).

The methodology is divided in three phases. At the analysis phase a UML class diagram is used to design the conceptual model. The design phase has two parts. Standard logical design is logical design independent of any product. Specific logical design is the design for a specific product. This specific logical design is needed because in different DBMS products SQL:1999 is implemented differently. The implementation phase is about physical design tasks like improving the response times. The whole methodology is based on UML and the extension possibilities of it. For the methodology, UML is extended with new stereotypes for designing object-relational constructs. (Cavero et al. 2004.)

Spatial data management is a typical area where the conceptual model needs to be extended. For example, the model should be able to express points, lines and polygons. Chawla et al. (2003) have proposed that a necessary miniature graphic is placed in an entity or in a class to represent a spatial object. There are numerous other extensions of how a conceptual model containing geographical objects should be presented. These suggestions include:

- GISER-model (Geographic Information System Entity Relational model) (Coyle et al. 1997).
- USM-model (Unified Semantic Model) (Ball et al. 1999).
- MADS-model (Modelling of Application Data with Spatio-temporal features) (Parent et al. 1999).

Even more models are analysed in (Friis-Christensen et al. 2001).

XML is presented as a solution for all types of integration projects (Linthicum 2001). Besides the benefits, XML also presents challenges to data modelling, since new ways of modelling and reconciling data models are needed (Lear 1999). As mentioned earlier, XML documents are typically described with a DTD or an XML schema. They allow us to express information in ways that better match the way the business is done. Therefore, they bring powerful capabilities to information modelling: heterogeneity, extensibility and flexibility (Brandin 2003). However, XML is not a language for conceptual modelling: it is possible to describe data, but not to model the relationships and interconnections of the whole information system. However, XML-capable DBMSs contain tools that wrap the existing data into XML formats by creating views based on XML schemas. Finally, to get full advantage of the object-relational DBMSs and the XML-enabled DBMSs, we need to model closer to the real world: the objects are in one piece and not artificially split into several objects.

In one idea, an XML schema is produced from the logical level data model indicating storage of XML documents (Bird et al. 2002). Pardede et al. (2004) have suggested using XML schema. This defines the whole design process starting from the conceptual level with a semantic model, turning into the logical model with a XML schema and finally mapping the XML schema into an object-relational structure.

As this section has shown, there is a real need to develop the database design process. The old traditional database design model is still valid, but the phases of it (and especially its modelling tools) should be redesigned. There are different proposals for this redesign, but a generally accepted model or kind of de facto standard does not yet exist that solves the requirements posed by recent innovations in databases.

3 DIFFUSION OF DATABASE INNOVATIONS

This chapter presents the key theory that was used to explain and analyse database innovations in the case organizations. First, it defines the concept of ‘diffusion of innovations’. Next, different factors affecting the adoption of innovations are presented and the lifetime of information systems are discussed. Finally, it introduces a developed framework of attributes that affect the adoption of database innovations.

3.1 Diffusion of innovations

Diffusion is a process by which an innovation is communicated through certain channels over time to the members of a social system. An innovation is an idea, practice or object that is perceived as new by an individual or other unit of adoption. (Rogers 1995) Defining ‘newness’ can be difficult and therefore Lyytinen and Rose (2003) write that the *perception* of newness counts, rather than whether the idea or artefact is new to the world. They also introduce the concept of the Information Technology (IT) innovation, defining it as the creation and new organizational application of digital computer and communication technologies. Swanson (1994) has presented a similar definition, but this focuses more on information systems (IS), while Lyytinen and Rose have a wider scope, starting from information technology and ending at information systems. Both Lyytinen and Rose (2003) and Swanson (1994) divide their concept further into subcategories. Table 3.1 details of Lyytinen and Rose’s subcategories and their relationship to Swanson’s categories.

The evolution of databases (database innovations) described earlier falls into the IT Base Innovations category and, more precisely, the *Base Technology Innovation* subcategory. The changes described earlier are all affecting the functionality and architectural principles of databases and thus they belong to the *Base Technology Innovation* subcategory. As we remember, one of the database innovations was combination of XML and databases. This also belongs to the *Base Technology Innovation* subcategory, but XML itself can be located in the Service Innovations subcategory *Technological Integration Innovation*.

Table 3.1: Subcategories of IT innovation (Lyytinen et al. 2003).

Category	**Description**	**Swanson's subcategories**
IT base innovations	• Base Technology Innovation o Changes in the base technology as defined by functionality, speed, reliability, architectural principle or other features o For example database management systems • Base Development Capability Innovation o Changes in IS development as defined by modelling and design principles or by the coordination of related processes • Base Service Capability Innovation o Changes in the service as defined by changes in general service features	n/a n/a n/a
System development innovations	• Administrative Process Innovation • Technological Process Innovation	Type I-a Type I-b
Service innovations	• Administrative Process Innovation • Technological Process Innovation • Technological Service Innovation • Technological Integration Innovation	Type II Type III-a Type III-b Type III-c

We can further adapt Lyytinen and Rose's (2003) concept of disruptive IT innovation for this research. A disruptive IT innovation is a necessary but not sufficient architectural innovation that originates in the IT base and radically and pervasively impacts systems development processes and services (Lyytinen et al. 2003). A disruptive innovation is also called a 'radical' or 'discontinuous' innovation (Rogers 2003). The evolution of databases can be understood as a disruptive IT innovation, while truly exploiting the new database technology means changes in systems development processes and enables and diversifies services. For example, designing an object-relational or XML-enabled database instead of a relational database changes the development process from traditional entities to new concepts like objects and XML documents. Also, services are influenced and data can be managed radically differently, e.g. by querying XML-documents or analysing geographical data with SQL.

Organizations considering innovations go through the 'innovation-decision process' (see Figure 3.1). Rogers (1995) defined this as a process through which an individual or other decision-making unit passes from gaining initial knowledge of an innovation, to forming an attitude toward the innovation, to making a decision to adopt or reject, to implementing the new idea, and finally to confirming this decision. As the Figure shows, the diffusion of innovations is a complicated process consisting of five stages. The knowledge stage consists of ascertaining the existence of the innovation gaining some understanding of how it functions. In the persuasion stage a favourable or unfavourable attitude toward the innovation is formed. Decision involves conducting activities that lead to the choice of whether to adopt or reject the innovation. At the implementation stage an innovation is put into use. Finally, the confirmation stage aims to seek reinforcement of an innovation-decision that has already been made, or to reverse a previous decision to adopt or reject the innovation if there have been conflicting messages about it. (Rogers 1995) Typically, the innovation is not copied or imitated exactly at the time of adoption. Therefore a concept of re-invention is defined. This is the degree to which an innovation is changed or modified in the process of its adoption and implementation. (Rogers 2003.)

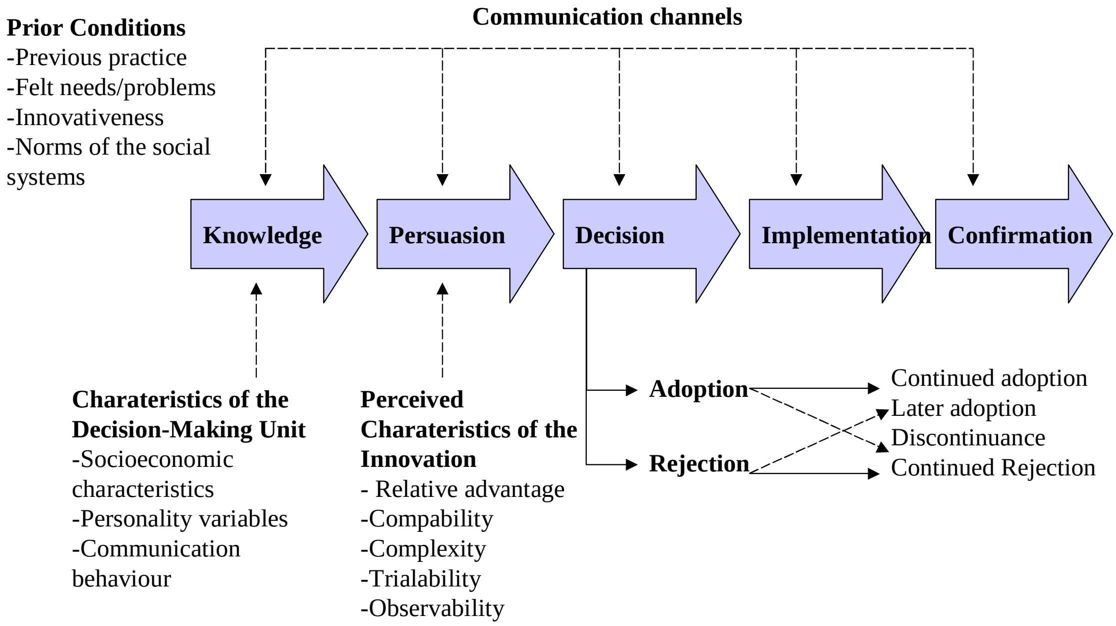


Figure 3.1: Innovation-decision process (Rogers 1995).

Thus, organizations and individuals have been faced with database innovations and according to Rogers (1995) there are five main attributes that influence their adoption of these innovations (see Table 3.2). The minus sign after ‘Complexity’ indicates that it has a negative influence on the adoption of the innovations.

Table 3.2: Attributes of innovations influencing adoption (Rogers 1995).

Attribute	Description
Relative advantage	The degree to which an innovation is perceived as being better than its precursor.
Compatibility	The degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters.
Complexity (-)	The degree to which an innovation is perceived as being difficult to use. This is negatively related to the adoption rate of the innovation.
Observability	The degree to which the results of an innovation are observable to others.
Trialability	The degree to which an innovation may be experimented with before adoption.

Other significant attributes influencing the acceptance of information technology are perceived usefulness and perceived ease of use (Davis 1989). Definition of perceived usefulness by Davis is quite similar to Rogers' definition of relative advantage: the degree to which a person believes that using a particular system would enhance his or her job performance. Rogers' definitions are based on perceptions of the innovation itself and not on perceptions of the persons actually using the innovation (Benbasat et al. 1991).

Most innovations have a S-shaped rate of adoption, but this curve is not consistent across all of them (Rogers 1995). For example, Figure 3.2 is a theoretical presentation of the diffusion of different database management systems. As the Figure shows there is so-called 'take-off stage' with every innovation (identified by the dotted area). We can also identify early and late adopters from the Figure.

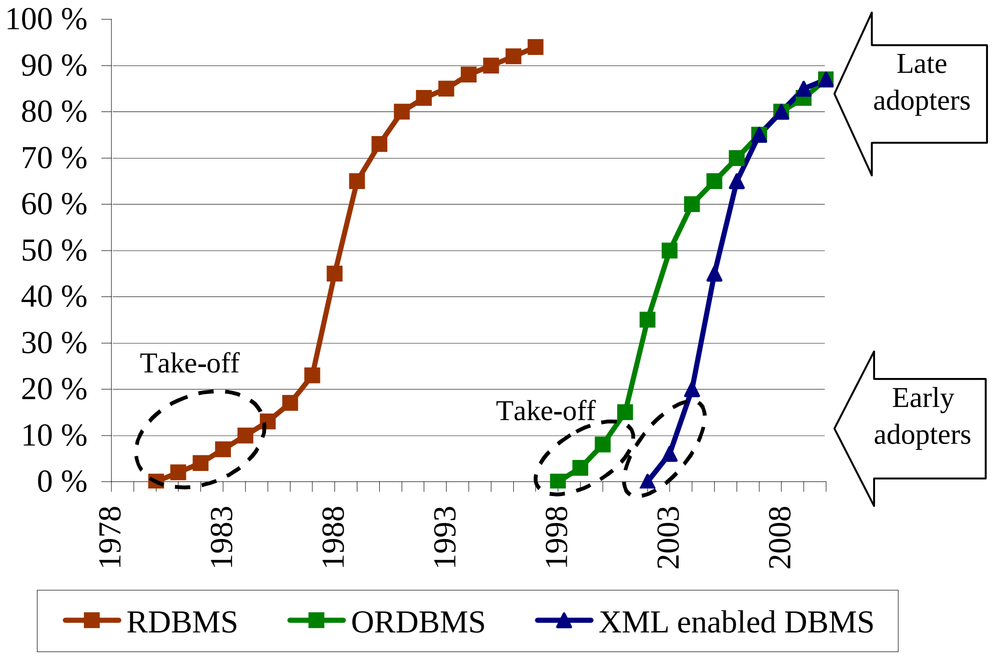


Figure 3.2: Diffusion of DBMSs in theory.

This innovation decision process works in an organization when individuals take innovation decisions independent of other members of a system. If they were made collectively or based on authority, the process would be described differently. The innovation process has now two sub-processes: Initiation and Implementation. (Rogers 2003) The innovation process was originally described in (Zaltman et al. 1973).

The initiation sub-process is divided into two stages. The agenda-setting stage occurs when a general organizational problem is identified that creates a perceived need for an innovation. The term ‘performance gap’ is used in describing this kind of problem and it is defined as the discrepancy between an organization’s expectations and its actual performance. In the matching stage the organization identifies a problem and opts for an innovation to solve it. Redefining/restructuring occurs when the innovation is re-invented for the needs of the organization and it prepares for the innovation. In the clarifying stage the meaning of the new idea gradually becomes clearer to the organization’s stakeholders. Finally, routinizing the innovation ensures that it becomes part of regular activities in the organization and it loses its separate identity. (Rogers 2003.)

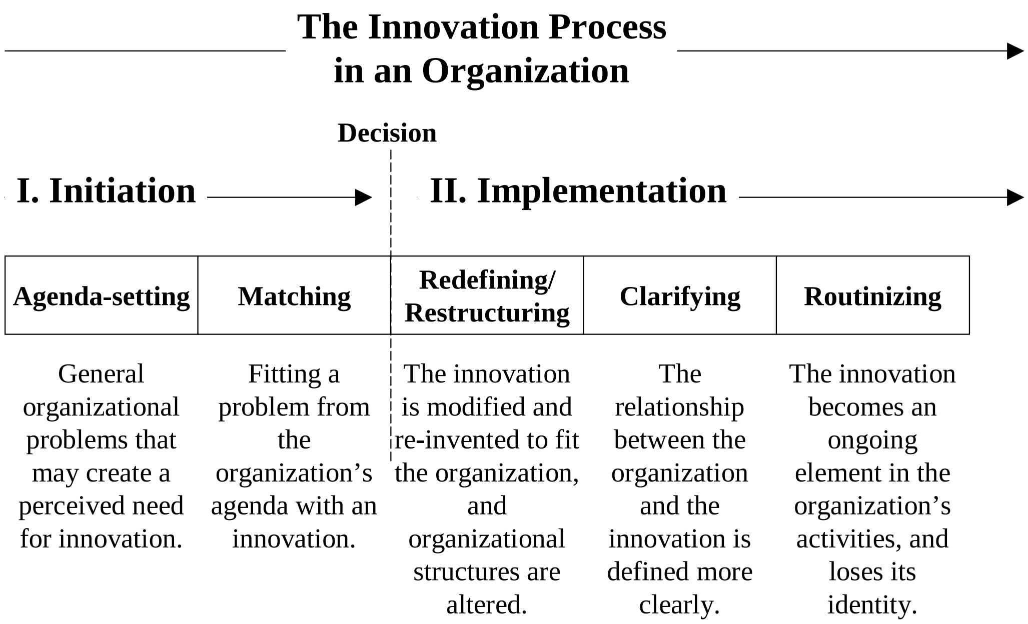


Figure 3.3: The innovation process in organizations (Rogers 2003).

Swanson (1994) presented an innovation diffusion circuit (see Figure 3.4). It describes a communication loop by which the innovation is supposed to spread throughout organizations. Figure 3.4 consists of two actors (the IS department and the host organization it serves) and two environments. Each actor operates within its own environment and in relationship to the other organization. The host business environment has a strategic impact on the host organization's way of doing business. The IS department works within the limits of the resource budget and provides services for the host organizations needs.

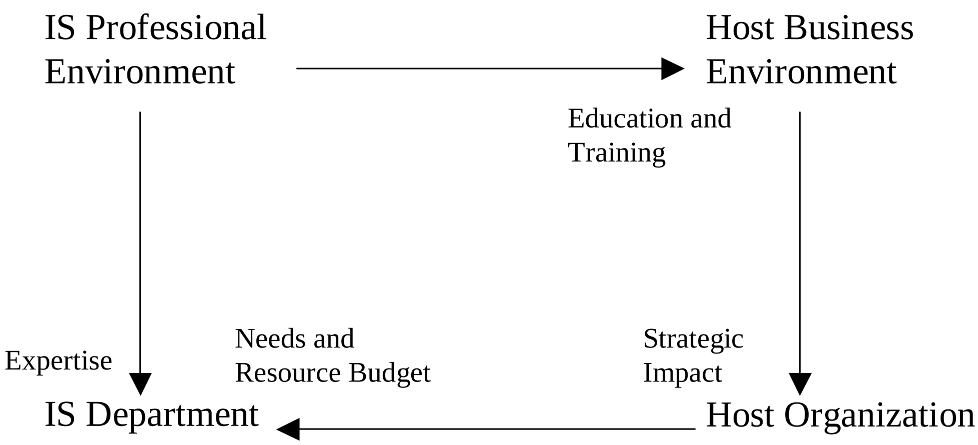


Figure 3.4: The IS Innovation Diffusion Circuit (Swanson 1994).

3.2 Factors affecting the adoption of innovations

Rogers (2003) presents variables that define the rate of adoption of innovations. There are variables that relate to the innovation itself and to the organization in the diffusion process. The innovation is described with the perceived attributes discussed above. For example, the organization is described, along with the type of innovation-decision and the nature of the social system.

Rogers (2003) also identifies a set of independent variables related to organizational innovativeness: 1) Individual (leader) characteristics, 2) Internal characteristics of the organizational structure and 3) External characteristics of the organization (see Figure 3.5). The variables having a positive impact to the organizational innovativeness are marked with a plus sign and respectively a minus sign indicates negative impact.

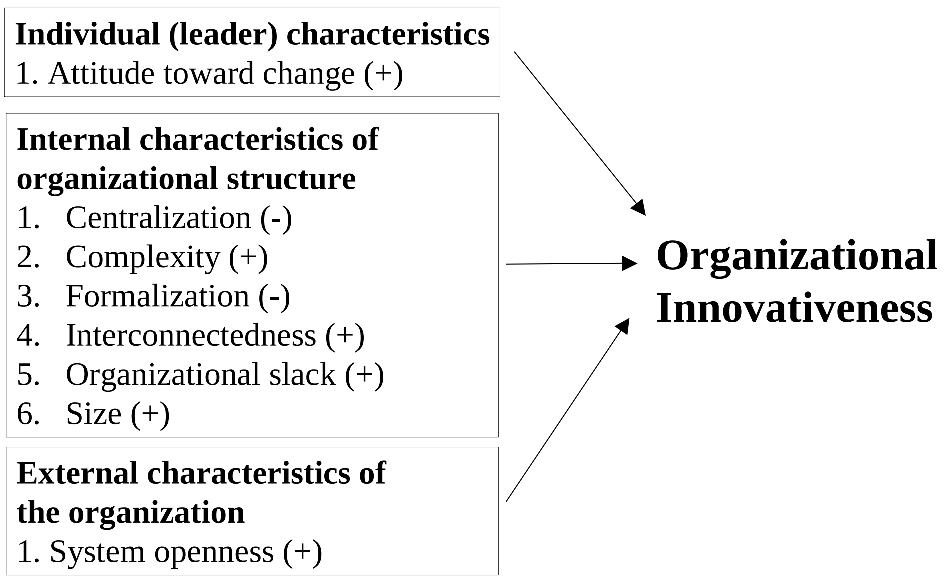


Figure 3.5: Independent variables related to organizational innovativeness (Rogers 2003).

Rogers (2003) defined these characteristics as follows:

- Centralization describes how far power and control is centralized in the organization.
- Complexity describes the level of knowledge and expertise that the organization's members possess.
- Formalization is the degree to which the organization emphasizes that its members should follow rules and procedures.

It can be described as the degree to which an organization is bureaucratic.

- Interconnectedness is the degree to which the units in a social system are linked by interpersonal networks.
- Organizational slack describes the availability of uncommitted resources to the organization. In practice this means that the organization has time, money and people for dealing with the innovations. Typically, bigger organizations have more slack resources.

Finally, Rogers (2003) states that the organizational variables might influence the innovation in one direction during the initiation phases of the innovation process and in the opposite direction during the implementation phases. For example, low centralization, high complexity and low formalization facilitate initiation of the innovation process, but they also make it difficult to implement the innovation.

As the above presentation shows, both internal and external characteristics affect the level of organizational innovativeness. Rogers concentrated on the internal environment (six of his independent variables were related to this, in contrast to only one that described the external environment). Zaltman et al. (1973) presented a more extensive list, originally by Duncan (1972), of components that affect to the innovation (see Table 3.3). All of the components are further divided into more concrete components. Most of Rogers' six internal independent variables can also be identified in this presentation.

Table 3.3: Factors and Components Comprising the Organization's Internal and External Environment (Duncan 1972).

Internal environment	External environment
Organizational personnel component A. Educational and technological background and skills B. Previous technological and managerial skills C. Individual member's involvement and commitment to attaining system's goals D. Interpersonal behaviour styles E. Availability of manpower for utilization within the system **Organizational functional and staff units component** A. Technological characteristics or organizational units B. Interdependence of organizational units in carrying out their objectives C. Intra-unit conflict among organizational, functional and staff units D. Inter-unit conflict among organizational, functional and staff units **Organizational level component** A. Organizational objectives and goals B. Integrative process integrating individuals and groups into contributing maximally to attaining organizational goals C. Nature of the organization's product service	**Customer component** A. Distributors of product or service B. Actual users of product or service **Supplier component** A. New materials suppliers B. Equipment suppliers C. Product parts suppliers D. Labour supply **Competitor component** A. Competitors for suppliers B. Competitors for customers **Socio-political component** A. Government regulatory control over the industry B. Public political attitude towards industry and its particular product C. Relationship with trade unions with jurisdiction in the organization **Technological component** A. Meeting the new technological requirements of own industry and related industries in production or product or service B. Improving and developing new products by implementing new technological advances in the industry

Walsham (1993) presents a framework for analysing organizational change associated with information systems. Also he emphasizes that changes occur at the organizational and process levels, in addition to the information system. He identifies four key components for the analysis framework: 1) Content, 2) Social Context, 3) Social Process and 4) Context/Process Linkage. Each of these components is further divided into associated conceptual elements:

- Content refers to changes at the organizational level (e.g. in processes) and at the information system level (e.g. software and related technologies).

- Social context concerns the social relations between those participants who are involved in the information system, the social infrastructure available or necessary for its support and the history of previous commitments made in connection with the information systems.
- Social process is about taking both cultural and political perspectives on how the information system resulted in organizational change.
- The Context/Process linkage is a framework that is quite similar to Rogers' view of factors affecting the rate of adoption of innovations.

CIOs (Chief Information Officer) and IT executives are required to deal with two broad constituents of the external environment: the external business environment and the external IT environment. Three dimensions can characterize the external IT environment (Ranganathan et al. 2000): IT Munificence, IT Dynamics and IT activity intensity (see Table 3.4).

Table 3.4: External IT environment Dimensions (Ranganathan et al. 2000).

Dimension	**Description**
IT munificence	The extent to which the external IT environment provides adequate resources to the organization.
IT dynamics	The extent to which the changes in the external IT environment are volatile and rapid.
IT activity intensity	The extent to which a firm's competitors, suppliers, customers, and other stakeholders use IT.

The business environment can be characterized according to two dimensions: product-based vs. service-based and high- vs. low-volatility context (Kankanhalli et al. 2003). In a stable business environment (low-volatility), it is possible to amend the information technology infrastructure and business model in response to relatively small rates of change. However, in an unstable business environment where the rate of change is much higher, it is more difficult to adapt the legacy information technology systems and business model fast enough to support the strategic vision. (Gibson et al. 1999) Therefore an unstable business environment is likely to speed up the diffusion of innovations.

So far organizational factors have been clearly emphasized as factors affecting the rate of innovations. When discussing IT/IS related issues of this nature, it seems important to understand how IT/IS is managed and what guides the management. Achterberg et al. (1993) identify four different paradigms for IT management (see Figure 3.6) based on strategic and operational activities. These paradigms are described through four concepts: control process, dominating internal structure, dominating political and power structures and cultural characteristics. They allow the dimensions of operational activities and strategic and tactical activities to be defined further, resulting a detailed model that means the organization can be positioned according to the paradigms presented.

In operational activities the dimensions are (from low to high): initial, experience, control, managed and flexible. In strategic and tactical activities the dimensions are (again, from low to high): start-up, growth, control, planning and strategic planning. (Achterberg et al. 1993) The paradigm an organization follows has of course influenced the diffusion of innovations process. For example, the management control paradigm is not characterized with a high level of innovation.

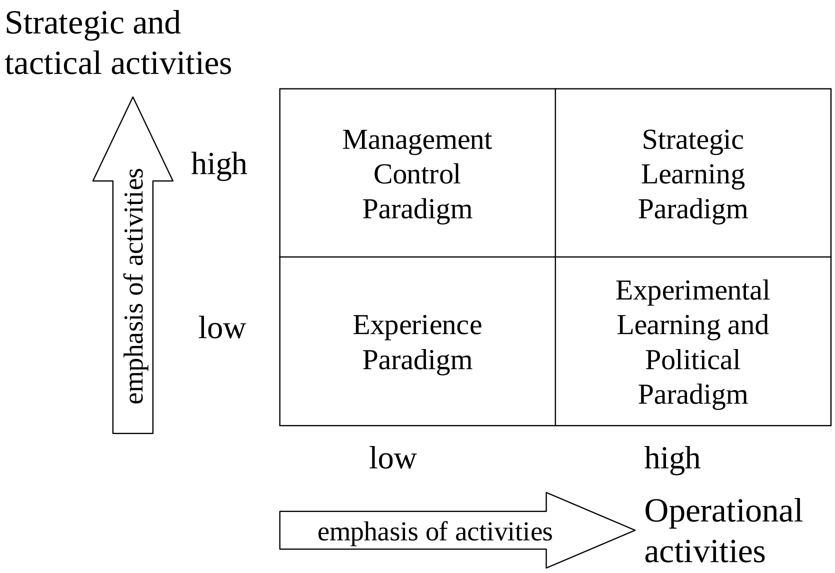


Figure 3.6: Four paradigms for managing IT (Achterberg et al. 1993).

Another point influencing the innovativeness of organizations' IT functions is the way in which the functions are organized. The IT function should be organized according to three underlying principles: 1) to foster co-evolution between the business and the IT function, 2) to nurture relationship networks for visioning, innovation, and sourcing and 3) to explicitly manage eight value-creating processes[10]. There are also three IT organizational models. The first model is the Partner Model, which aims to stimulate, catalyse and seed

[10] The processes are: Infrastructure management, Human capital management, Relationship management, Value innovation, Solutions delivery, Services provisioning, Strategic planning, Financial management.

thinking about strategic uses of IT. In this model IT is an active partner in setting the direction and timing of future IT capabilities. In the second model, the Platform Model, IT is primarily expected to provide infrastructure and tools to enable current and future business innovations. However, in contrast with the Partner Model, IT is not expected to be an active collaborator in initiating business innovations. In the third model, the Scalable Model, IT is viewed as a strategic differentiator and an important element of business innovation. (Agarwal et al. 2002.)

One of IT management's main tasks is to design and to implement an IS/IT strategy, which should be in line with the enterprise's business strategy (Gupta et al. 1997; McAuley et al. 1997; Jung et al. 2003). Indeed, better links between the IS and business strategies was ranked as the most challenging issue for IT managers in a Norwegian study (Christensen et al. 2000). In another study over 50 percent of respondents indicated that aligning business and IS/IT strategies was very critical for their organization (CSC et al. 2003). Enterprises must however firstly decide which technologies are relevant to them (Jung et al. 2003). Actually, this prioritising of technology investments was ranked as the most critical organizational issue in a 2003 study (CSC et al. 2003). The management must also decide whether the required technologies should be developed internally or produced by an external organisation. Finally, the enterprise has to decide which technologies should be used exclusively internally or transferred to other firms. (Jung et al. 2003) A framework is proposed that gives the IS/IT strategy four dimensions (see Figure 3.67).

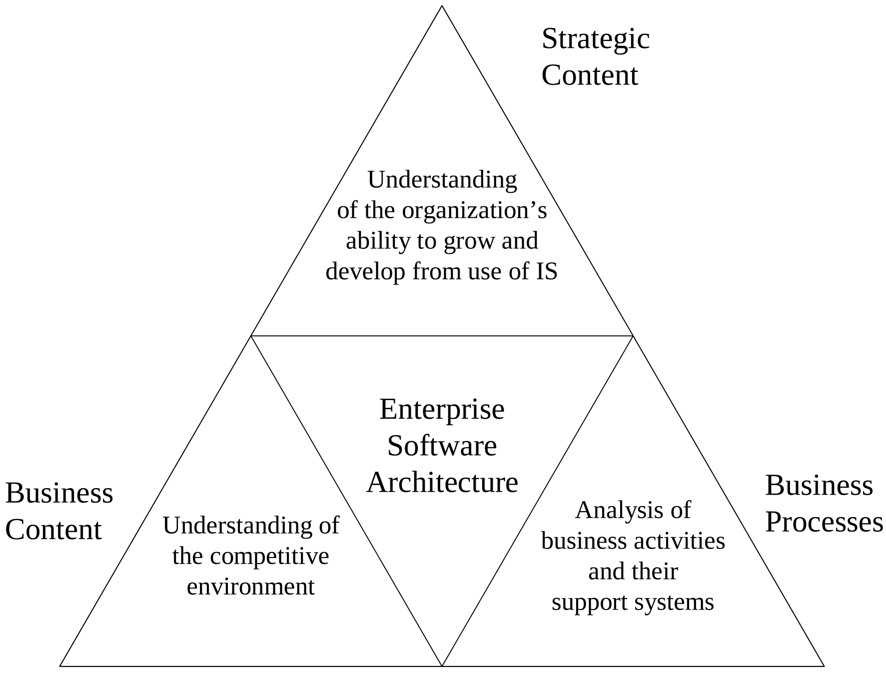


Figure 3.7: Four dimensions of IS/IT strategy (Cheong et al. 2001).

The framework requires a) understanding of the competitive environment, b) analysis of business activities and their support systems and c) understanding of the organization's ability to grow and develop from use of IS (Cheong et al. 2001). This framework emphasizes the importance of understanding the organization, the business and the enterprise software architecture. Research is also showing how aligning IS and business strategies can contribute to the success of the organization (CSC et al. 2003).

The IS/IT strategy can be divided in two distinct areas. The IS strategy is demand-oriented, focusing on the role of information and system requirements in meeting business objectives. The IT strategy, on the other hand, is supply-oriented and concerned with specifying the technology and how to deliver its applications. (Burn et al. 2000.)

The IS success model by DeLone and McLean (1992) provides some useful tools to help with analysing information systems. The original model states that the organizational impact of an information system depends on the quality of the system and quality of the information. These influence the use of the system and users' satisfaction with it. The organizational impacts of information systems probably lead to upgrade or replacement efforts and may also lead to the diffusion process for adapting new innovations. A revisited model was presented in 2002, but the basic structure was almost the same (DeLone et al. 2002).

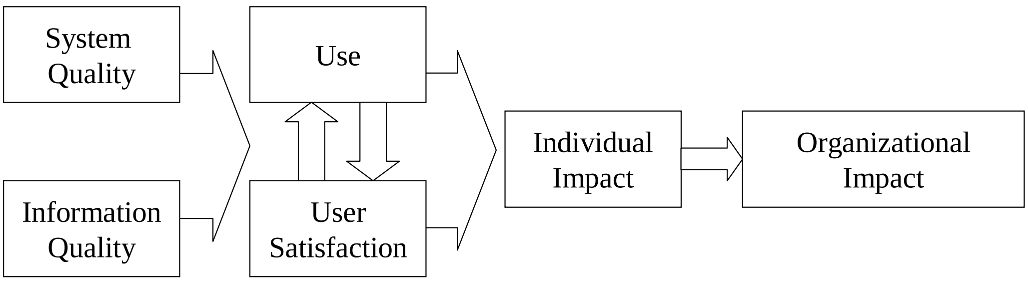


Figure 3.8: Model of Information System Success (DeLone et al. 1992).

3.3 Lifetime of Information Systems

The developments in database management systems influence also the maintenance of information systems (Akoka et al. 2003). Thus, if the evolving technological improvements are to be utilized, the information systems need to be maintained and upgraded or replaced. IS maintenance can be defined as correcting errors, improving performance or altering the system according to changed requirements after the information system is implemented (1993). IS maintenance is important since a) most of the costs during the lifetime of an information system are spent on it and b) if an information system cannot be altered quickly and reliably, the organization misses out on some business opportunities (Bennett et al. 2000). However, the combination of incremental change and their age, size and structural degradation means that legacy information systems are hard to maintain (Gibson et al. 1999). Key maintenance issues include a) the costs of IS maintenance and b) when an organization should start to think about replacing the system. According to one study many organizations are still making IS/IT investments to support ageing and inefficient processes, or adding to new processes that were designed without considering current IT capability (Feeny et al. 1998).

Information about the expected lifecycles of information systems is essential for an organization during the innovation process. Thus it is important to understand how the useful life of the system is determined and what consequences it has. As Swanson et al. (2000) write, the expected lifecycle of an information system also depends on how it has been maintained. They also acknowledge organizational context as an important factor relating to the expected lifecycles of information systems. However, they focus on (but are not limited to) the basic characteristics of information systems, like their age and size. The actual model they introduced is presented in Figure 3.9. Their findings confirm that system's age, size and portfolio complexity play significant roles in determining how much maintenance will probably be required and remaining life expectancy. In addition, the maintenance effort and life expectancy are positively related. They also write that the lives of information systems are extended or shortened through management decisions.

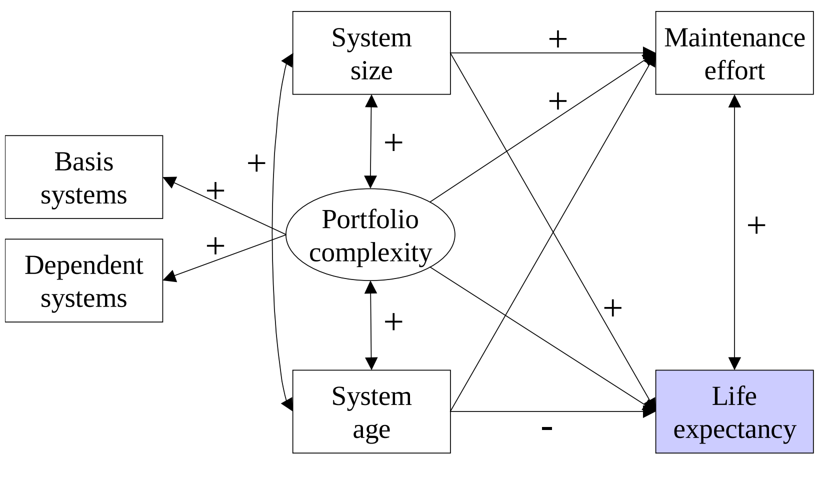


Figure 3.9: Life expectancy of an information system and the associated factors (Swanson et al. 2000).

An information system is maintainable whilst the benefits of its use justify the costs of maintenance and operation (Swanson 1999). Maintenance is seen as the most costly phase in a software life cycle (Idris et al. 2002). Usually there are fixed costs from the licences of the information system and the DBMS. The maintenance costs of an information system are said to follow a bathtub shape, i.e. at the beginning there is lot of maintenance, but over time this drops off and stabilizes, until it begins to rise again, eventually becoming very expensive (Lehman 1980; Heales 2000; Glass 2002). Basically, identifying where an information system is according to this graph will influence the diffusion process (Glass 2002). When information systems get older and have gone through many changes and updates, they become more fragmented and intertwined. As a result, maintenance becomes more difficult and time consuming, thus more expensive. (Heales 2000) Typically, these changes and updates also influence the database behind the information system.

Problems in maintaining software can also occur, depending on how much information is available from the supplier (Idris et al. 2002). Unfortunately the only reliable source of information about the system is often its source code (Gibson et al. 1999). Documentation can help to manage the development process as well as facilitate information system use and maintenance (Jazzar et al. 1995). Not only the lack (Moreira et al. 1998), but also the poor quality (Krogstie et al. 1994) of documentation has been identified as big problem in maintenance. The quality of documentation is low partly because it is usually generated in an ad-hoc manner (Huang et al. 2003). The lack of policies and standards also affects the maintenance process. If no policies and standards are present, the number of maintenance projects is likely to increase by 50% where the documentation is not in order. (Idris et al. 2002.)

3.4 Framework of the attributes that affect the adoption of database innovations

A developed framework of the attributes that affect database innovation adoption in the case organizations is presented in Figure 3.10. The framework has two main dimensions. The innovation is evaluated using the concepts of Rogers (2003) according to its perceived attributes. The organizational attributes affecting the diffusion process are divided into internal and external attributes, which are further categorized into business and technological environments. For the internal business environment the concepts of Rogers (2003) are selected, but also checked against the other theories presented in the previous sections. For example, instead of the first two properties of the organizational personnel component by Duncan (1972), Rogers' complexity concept (2003) is used. However, complexity is now evaluated from the perspective of database innovations. A study indicates that the level of technical skill amongst IS staff is the greatest contributor to the success of IS in an organization (CSC et al. 2003). Similarly the selected concept Organizational slack covers Duncan's (1972) concepts 'E. Availability of manpower for utilization within the system' and 'A. Technological characteristics or organizational units'.

For the internal technological environment the concepts are formed following. Theories influencing and explaining the strength of IT Management are selected from the following: IT management paradigm (Achterberg et al. 1993), IT organizational model (Agarwal et al. 2002), IS/IT strategy (Cheong et al. 2001; Jung et al. 2003; Gupta et al. 1997). The IS Portfolio consists of the following parts: system and information quality (DeLone et al. 1992), and system size, age and portfolio complexity (Swanson et al. 2000).

The concept of system openness is defined in Rogers (2003) and the concept of business volatility comes from Kankanhalli et al. (2003). Ranganathan et al. (2000) provide the external technological environment attribute.

The innovation part of this framework was used as a tool to analyse database innovations. The organization part of the framework was used to assess case organizations' orientation towards the adoption of database innovations. A detailed description of the use of this framework is presented in Section 4.5.

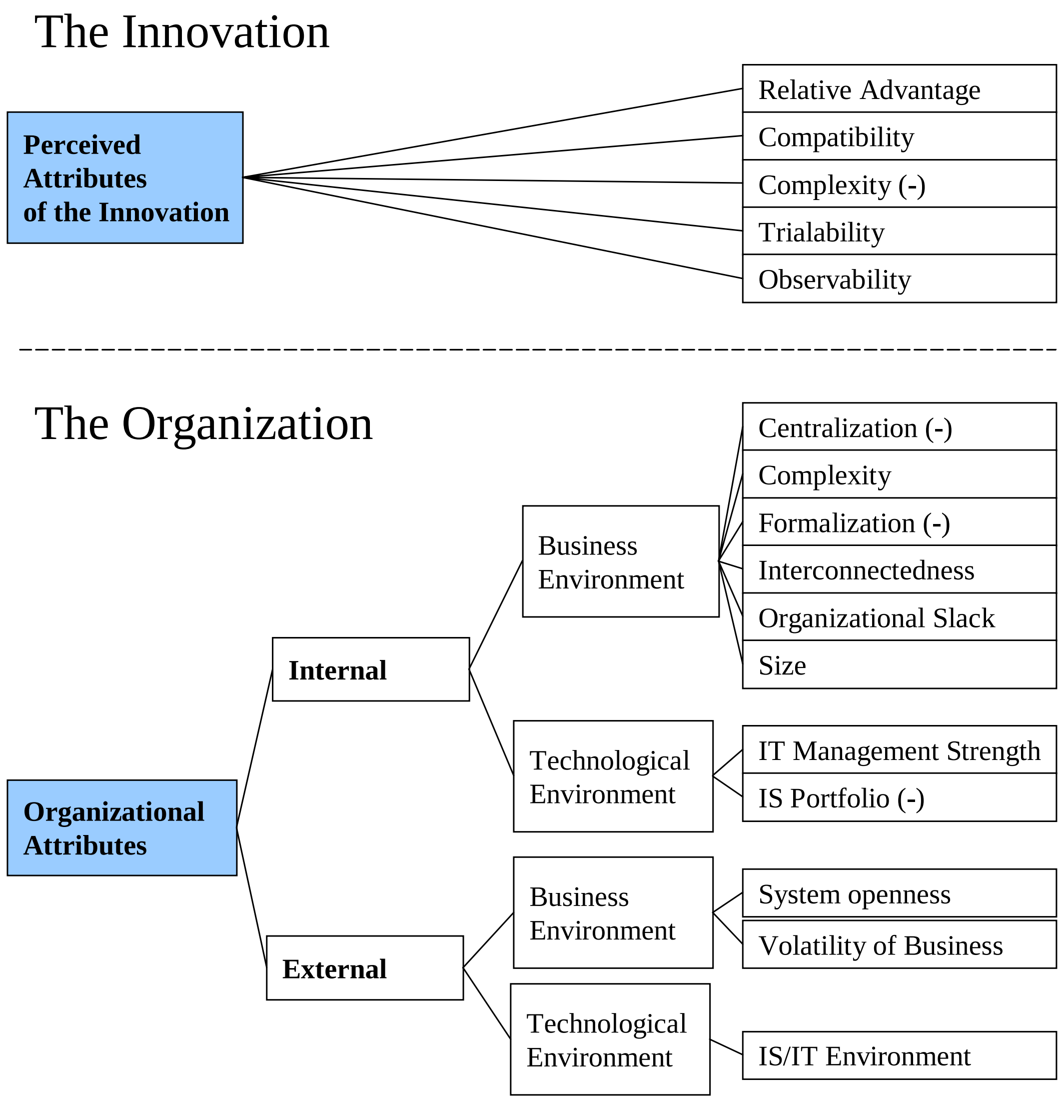


Figure 3.10: Developed framework of the attributes that affect the adoption of database innovations.

4 THE RETRO-PROJECT

This chapter is about the empirical part of the research. It describes the methodological choices, design of the research, data collection process and data analysis.

4.1 The RETRO-project and the goals of the research

The name of the research project is RETRO, which is an abbreviation of Relational Database Trends in Organizations. It began by looking at database evolution and their diffusion in organizations. Studies on innovation use two distinct perspectives for analysis: adoption and diffusion. In this research the focus is on diffusion, when we attempt to understand why and how an innovation spreads and what characteristics of the innovation lead to widespread acceptance (Premkumar et al. 1995). In other words, the purpose of this research is, on the one hand, to evaluate and describe the current state of information systems and databases in different organizations. On the other hand, it aims to identify and describe forthcoming database trends in organizations.

The goals of the research are thus:

1. To describe database exploitation and other activities related to databases in different organizations
   - Types of DBMS.
   - Administration and maintenance.
   - Interfaces between databases and information systems.
   - Level of technology adoption.
2. To identify typical shortcomings and problems in the exploitation of databases in different organizations.
3. To describe the present development plans that the organizations have in the field of databases.
4. To describe the requirements and effects of the diffusion of database innovations in different organizations.
5. To identify, if possible, general trends in the diffusion of database innovations.
6. To describe the pros and cons of database innovations.

7. To identify critical organizational factors that support the adoption of database innovations.

This research also tries to interpret how database innovations are handled in the innovation-decision process. In order to do this, the current business and technological environment must be understood and the present level of knowledge with the database innovations in the organizations must also be recognized.

4.2 Research methodology

Methodologically, this research is a multiple case study research. This approach is a viable research methodology in IS-research for several reasons. First, it allows both academia and practitioners to keep up with the rapid changes occurring in the IT world as well as in the organizations (Dudé et al. 2003). Second, the researcher can study information systems in a natural setting, learn about the real-life situation, and generate theories from practice (Benbasat et al. 1987; Cavaye 1996). Third, in-depth investigations help to develop new ideas and new lines of reasoning, and pinpoint the opportunities, challenges and issues facing IT specialists and managers (Dudé et al. 2003). Fourth, case methods allow the researcher to understand the nature and complexity of the processes taking place (Benbasat et al. 1987). Finally, capturing the context of people-related and organizational phenomena is always important, making case studies an appropriate research strategy in IS (Cavaye 1996). Case study research is also well suited for understanding the interactions between IT-related innovations and organizational contexts (Broadbent et al. 1998).

There is no generally accepted definition for case study research (Cavaye 1996), but for example Yin (1994) defines it as *an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident.* Benbasat et al. (1987) use this definition as a basis for their own description. According to them, *a case study examines a phenomenon in its natural setting, employing multiple methods of data collection to gather information from one or a few entities.* Additional characteristics defined for case study research are (Benbasat et al. 1987; Yin 1994):

1. The complexity of the unit is studied intensively.
2. No experimental controls or manipulation are involved.
3. The investigator may not specify the set of independent and dependent variables in advance.

4. The results derived depend heavily on the investigative power of the researcher.
5. Changes in site selection and data collection methods could take place as the investigator develops new hypotheses.
6. Case research is useful in the study of “how” and “why” questions, because these deal with operational links to be traced over time rather than with frequency or incidence.

Case study research can be used both in positivism and interpretivism research traditions. In positivism facts and values are distinct and scientific knowledge consists only of the former, but in interpretivism facts and values are intertwined and both are involved in scientific knowledge (Walsham 1995). In the interpretivist tradition the aim is to understand the phenomena from the point of the view of the participants who are directly involved in it. In the positivist tradition the aim is to understand and explain the social setting by identifying individual components of a phenomenon and to explain it in terms of constructs and relationships between constructs. (Cavaye 1996) Research suggests that most of the case studies are positivist. Positivist case research is further divided into theoretical and descriptive research. For example, in descriptive case research, investigators attempt no theoretical interpretation of the phenomena; rather, they present what they believe to be straightforward, objective, factual accounts of events to illustrate some issue of interest. (Dudé et al. 2003.)

This research follows the interpretivist tradition of the case research, since the aim is to understand the exploitation of information systems and databases through the views of the interviewees. The information systems and databases are not directly examined. In interpretivism there is no objective reality that can be discovered by researchers and replicated by others (Walsham 1993; Broadbent et al. 1998). An interpretive researcher attempts to gain a deep understanding of the phenomena being investigated and their subjectivity is acknowledged as part of this process (Broadbent et al. 1998). Interpretive researchers attempt to understand the way others construe, conceptualise and understand events, concepts and categories, in part because these are assumed to influence individuals’ behaviour (Duchon et al. 1988). Interpretivism is thus an epistemological position, concerned with approaches to the understanding of reality and asserting that all knowledge is necessarily a social construction and thus objective (Walsham 1993). At the lowest level an understanding of the reality is achieved, as it appears in the natural settings of the interviewees. At a second level the researcher interprets what is happening in the field. The highest level involves conceptualising and abstracting the interpretations of the second level. This carries the research partly into the positivist sphere. However, as Walsham (1993) argues, the relationship between the theory and the research is different in the interpretive tradition: "*There are no correct and incorrect theories but there are interesting and less interesting ways to view the world. A reader might well ask 'interesting to whom'? An author can only respond in the first instance that the theories, which he or she presents, are interesting to themselves and may be interesting to others.*"

4.3 Research design

This section describes the research questions, case selection and the actual cases that were studied. Finally, the research stages are presented.

4.3.1 Research questions

The database innovations of recent years inspired me to study their diffusion in organizations. The main research question is based on the research setting outlined in Figure 4.1.

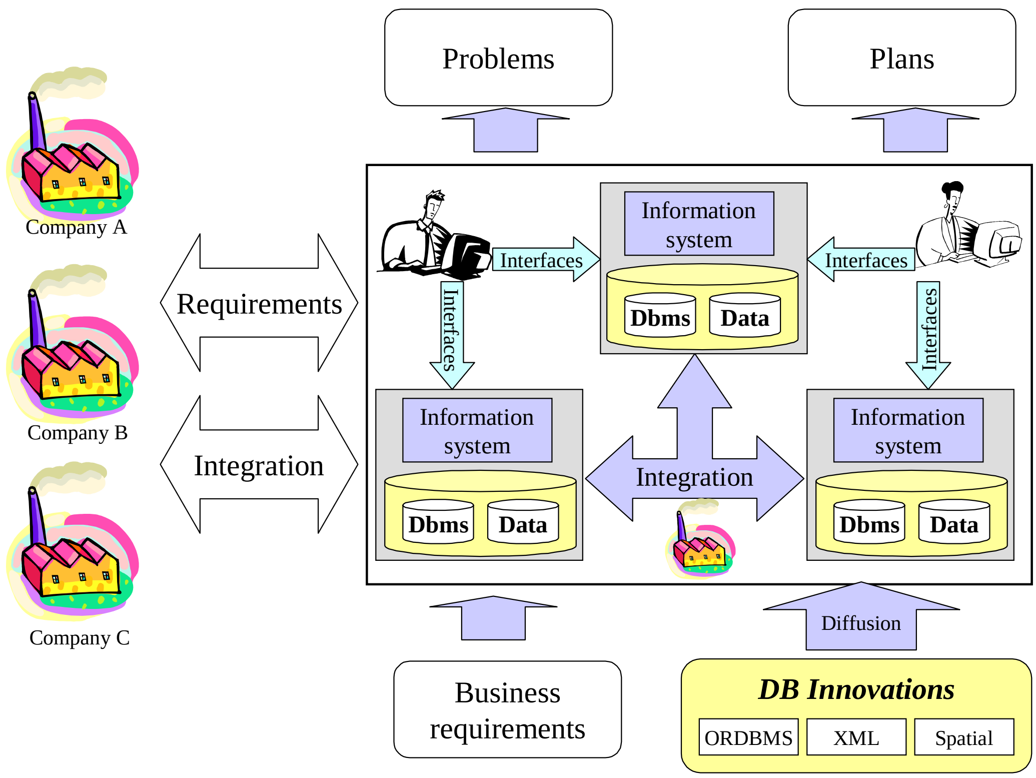


Figure 4.1: Research setting.

Since the diffusion is happening inside organizations, the organization itself plays a major role in adopting database innovations. Also relevant are the business requirements that influence the way in which information technology is exploited. These requirements may for example require integration of information systems and exchange of the data used in the business processes. There are many essential information systems and databases in a typical organization. These manage large amount of different data and are probably integrated in some way. The IS environment is not static and new ideas and plans may exist partly because of the problems and shortcomings inherent in current information systems and databases. Also, other organizations and companies may influence the information systems and databases by expressing their hopes and requirements. An organization has to work with and fulfil pressures and requirements from both internal and external stakeholders.

The main research question in this research setting is *How do database innovations diffuse in organizations?* This question is further divided in empirical and theoretical dimensions. In the empirical dimension two focused sub-questions are defined:

E1) How have database innovations diffused in the cases?

E2) How will database innovations diffuse in the cases in future?

The first subquestion concerns the present exploitation of databases by the case organizations. This thesis will describe what kinds of database solutions are currently exploited, how the data is modelled, how maintenance and administration of databases is organized and problems with the present database architecture. It also describes the level of database integration. The second subquestion focuses on the future of database innovations in the case organizations. This study describes the plans and trends that the case organizations have for information systems and background databases. In addition, it discusses how organizational factors contribute to the diffusion of database innovations in each case.

In the theoretical dimension three focused subquestions are defined:

T1) How do database innovations influence data management?

T2) How can diffusion of innovations theory predict the adoption of database innovations?

T3) What are the critical organizational success factors for the adoption of database innovations?

The first subquestion addresses the improvements, possibilities and drawbacks of the database innovations when they are adopted. It also covers the changes that are needed in database architecture and how organizations can start exploiting the possibilities of database innovations. The response of this thesis to the second subquestion analyses the use of diffusion of innovations theory in predicting the adoption of database innovations. The third subquestion describes critical success factors that contribute to the adoption of the database innovations.

Most of the research questions are ‘how-questions’ that are typical of case study research. ‘What-questions’ can be presented as part of these how-questions, following Yin’s (1994) advice that case studies that address what-questions are most appropriate when the purpose of the study is to explore a new phenomenon. (Yin 1994) As explained earlier, the new phenomenon for this research is the diffusion of database innovations in the case organizations.

4.3.2 Multiple case design

Multiple case designs are desirable when the intent of the research is description, theory building or theory testing (Benbasat et al. 1987). Cases must be selected so as to maximize what can be learned in the period of time available for the study (Dudé et al. 2003). The number of required cases also depends on the focus of the research question (Broadbent et al. 1998). Each case must be carefully selected so that it either predicts similar results (literal replication) or forecasts contrasting results but for predictable reasons (theoretical replication) (Yin 1994; Cavaye 1996). In this research literal replication strategy was used. The cases were selected as equally representative, with no predetermined ideas. If we had chosen theoretical replication the cases would have been selected differently.

Research on organizational-level phenomena requires for the site selection to be based on the characteristics of organizations, i.e. industry, company size, public or private ownership (Benbasat et al. 1987). Studying multiple cases enables analysis of data in various situations (Cavaye 1996). As a research strategy, case studies are good for revealing the depth and complexity of situated processes or phenomena, but often lack generalization. In turn, to increase generalization, multiple case studies must be employed, and their analyses conducted at different levels. (Jazzar et al. 1995.)

Many companies and organizations were contacted during the case selection process. The contacts were made by phone and an introductory letter was emailed. This letter described the nature and the context of the research project and its objectives. Finally, a total of six cases were selected for the project and studied in the following order: SOK Corporation (SOK), Salon Seudun Puhelin Ltd (SSP), Statistics Finland (STAT), State Provincial Office of Western Finland (WEST), TS-Group Ltd (TS) and Optiroc Ltd (OPTI). Table 4.1 gives some detailed information about the case organizations and additional information is also available in Appendix 1.

The cases were selected according to certain criteria and the selected six cases satisfied all of these requirements. First and most important of these was a willingness to co-operate. Secondly, the organizations should be at least medium-sized. The EU Commission has defined that a medium-sized organization has 50 to 250 employees and an annual turnover between 10 and 50 million euros (EU Commission 2003). All except one case are medium-size when looking the annual turnover, but only one is still in this category according to its number of employees. The rationale for aiming at medium-size or larger organizations was the desire to study more complex environments than probably would have been the case otherwise. Third, the organizations should not be direct competitors with each other. This criterion was based on the idea that the organizations could hopefully learn something from each other during the research. It also enabled more open discussions at the beginning and in the final seminars. Furthermore, we could write the reports without worrying about issues of competition. Fourth, the organizations should represent different lines of businesses. This provided a broader perspective in relation to database innovations throughout the Finnish economy. Fifth, the organizations should not have IT as the major business. This criterion focused the research more on business than technology issues. It also meant that interviewees provided a more current and realistic view of the diffusion of database innovations rather than information about those in the development phase that would enter the market later. Finally, public sector organisations should be among the cases. This criterion also broadened the view of the diffusion of the database innovations.

Every case should also serve the research in some special way (Yin 1994). Research questions should also be interesting and important to the participating organizations in order to gain their support and ensure access to their people and resources (Broadbent et al. 1998). Therefore, in addition to the research setting and the research questions, a specific and more detailed research area was defined with every case, as presented in Table 4.1. These additional research areas focused on organizations' main interest areas around database innovations and provided a rich description and understanding of the nature of the specific research area. Actually every organization received three reports:

- The case-specific report.
- Cross-case report.
- This PhD-thesis.

The case-specific report concentrated on the specific topic in question, described the current state in the organization in relation to the specific research field and suggested proposed targets for development. The cross-case report, which was based on the different cases, described general trends in database usage and offered proposals for better data management. Finally, this PhD-thesis provides deeper analysis of the diffusion of database innovations.

Table 4.1: The case organizations in RETRO-project.

Organization/ Abbreviation used in this research	**Line of business**	**Private/ Public**	**Turnover in 2002 *(€ millions)***	**Number of employees in 2002**	**Specific research field**
SOK corporation/SOK	Co-operative society (Food and groceries, Consumer goods, Hotels and Restaurants, Hardware and Agriculture, Car dealerships, Fuel sales)	Private	2998	4645	Data modelling for information systems integration
Salon Seudun Puhelin Ltd/SSP	Tele-communications	Private	28	121	Management of information systems
Statistics Finland/STAT	National statistics	Public	52	1 074	Storing and managing spatial data in databases
State Provincial Office of Western Finland/WEST	Regional administrative authority	Public		350	Managing inquiries with databases
TS-Group Ltd/TS	Printing services and Communi-cations	Private	69,7	2 052 (Consolidated corporation)	Managing customer information with databases
Optiroc OY/OPTI	Building materials	Private	149	388	Maintenance and customisation of IS

The basic units of analysis were the interviewee’s knowledge, meanings and interpretations, as expressed in their answers to questions about essential databases or information systems. In each case multiple databases and information systems were studied. This specification of the unit of analysis is critical if we want to understand how the case research relates to a broader body of knowledge (Dudé et al. 2003).

4.3.3 Research stages

The stages of the research are presented in Figure 4.2, which has been adapted from Yin (1994). The research had three main stages: 1) Define & design, 2) Prepare, collect & analyse and 3) Analyse & conclude. Every stage was also divided into smaller steps. One relevant point is that every case stands alone and after each was complete a final meeting of 1 to 2 hours was held to present and discuss the results of the study.

The design and define stage started in May 2002 and it lasted six months. During that period an understanding of the development in the field of databases was acquired and the case study was designed. The second stage, i.e. the empirical part of the research, started in October 2002 and it continued until September 2003. In this stage six single-case studies were conducted, largely following the interpretive approach. In third stage the cases were cross-analysed and the cross-case report was published at the beginning of February 2004. Since then this doctoral thesis has been the focus of the researcher.

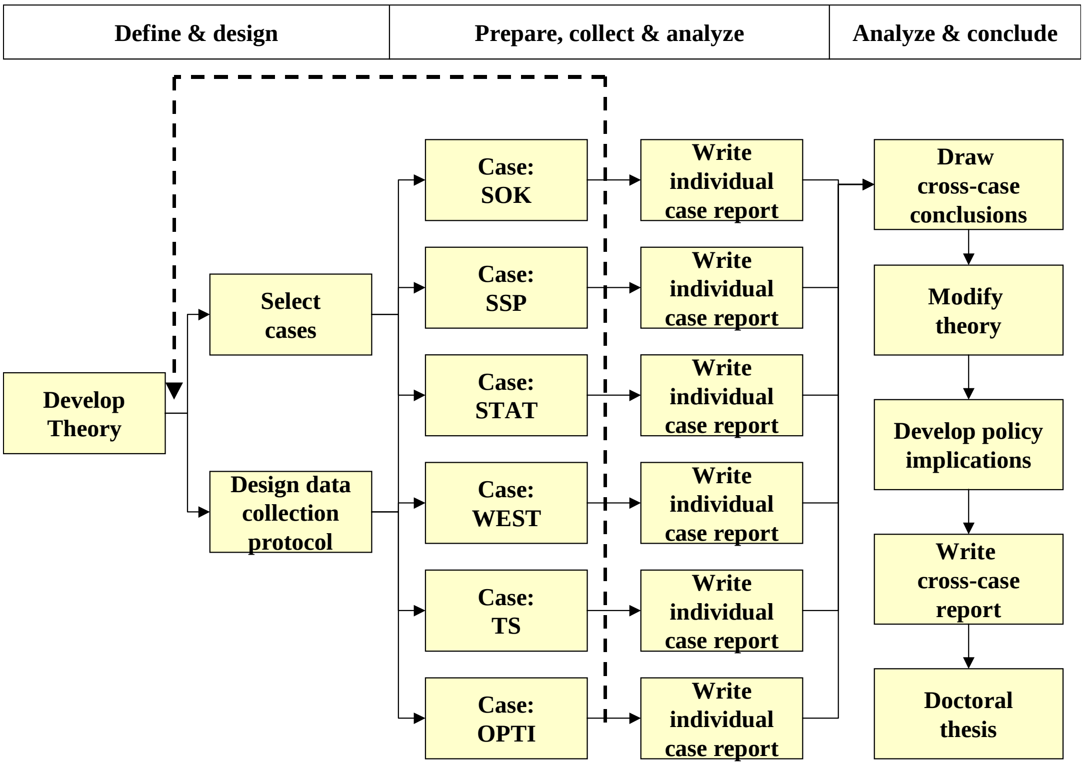


Figure 4.2: Phases of the research.

4.4 Data collection

This section describes issues relating to data collection. It starts with discussing and describing the multiple data collection methods used in this research. It then outlines the interviews and the themes of the interviews, before describing the actual data collection process.

4.4.1 Multiple data collection methods

In a typical case study research multiple data collection methods are used (Benbasat et al. 1987). This involves combining different techniques, such as interviews, observation, and document and text analysis (Broadbent et al. 1998). All of these techniques were employed in this research.

Semi-structured interviews were used as the main data collection method. This is the case for most interpretive case studies, since the method provides the best access to the interpretations of participants regarding information systems and databases in the organization (Walsham 1995). Actually, interviews are recognized as essential sources of information for all kind of case study research (Yin 1994). The strengths of interview are that a) it is targeted and focuses directly on the case study topic and b) it is insightful, providing perceived causal inferences (Yin 1994).

During this research a total of 54 interviews were held and the number of interviews per organization ranged from 6 to 14 (see Table 4.2). The length of a single interview was approximately 100 minutes and the total time spent conducting interviews was over 91 hours. The researcher prepared well for every interview since interview time should be used only to obtain information that cannot be obtained in any other way (Broadbent et al. 1998). The interviewees were set up by the organization's contact person, who was usually its IT manager. The interviewees represented various positions, but they were typically persons in administrative positions toward the information systems. Examples of the posts held by the interviewees include CEO, IT manager, IS Manager, Project manager and Main User. All conducted interviews are listed in the Appendix 1.

Table 4.2: The interviews.

Source	SOK	SSP	STAT	WEST	TS	OPTI	Sum
Interviews	12	10	7	10	8	7	54
Interviewees	14	10	6	10	8	7	55
Total length of interviews	26 h	17 h	11 h	12 h	12 h	13 h	91 h

Data was also gathered from existing documentation and archival records like annual reports. Documentation provides stable, unobtrusive, exact and broad coverage of the cases (Yin 1994). Most of the documentation (memos, publications, presentations) and archival records were received during the interviews. Table 4.3 present the many documents used as data sources in the different cases.

Direct observations were also used in some instances, such as when better understanding of the meaning and functionality of the information system was needed. For example, in Salon Seudun Puhelin Ltd and the State Provincial Office of Western Finland two information systems were described in more detail. In Statistics Finland the researcher participated in a GIS seminar to get more information about the state of spatial data management in the area.

During the data collection period the researcher acted as an outside observer. Acting this way means that the interviewees will be relatively frank in expressing their views (Walsham 1995). This happened frequently during the interviews, since the interviewees were often very critical of the information systems. However, as an outside observer the researcher might also be prevented from accessing certain data (Walsham 1995) and this also proved to be the case a couple of times when some information was not provided, e.g. costs of licences.

Table 4.3: Documents as sources of evidence.

Case	Documents
SOK	• Annual report • Current EDI solutions • EAI research report • IS documents • Data models of present databases
SSP	• Annual report • Current IS environment • IS documents
STAT	• Annual report • Memos of spatial data management issues • Powerpoint presentations of GIS services in the organization • IS documents • Data models of present databases • Report of one information system project • Documents of current database management systems
WEST	• Annual reports • Official reports of basic services in West-Finland • Final report of developing data collection • Excel-sheets used in data collection • Description of the data managed in WEST • Draft of spatial data exploitation in the future • List of information systems • A number of other reports
TS	• Annual report • Memos of the DBA group • Description of the database infrastructure • Draft of the CRM infrastructure • Report of the customer segments
OPTI	• Annual report • Current IS environment • Process models • List of the databases in use • User manual of one information system

4.4.2 Themes of the interviews

The interviews were divided in two parts (see Figure 4.3).

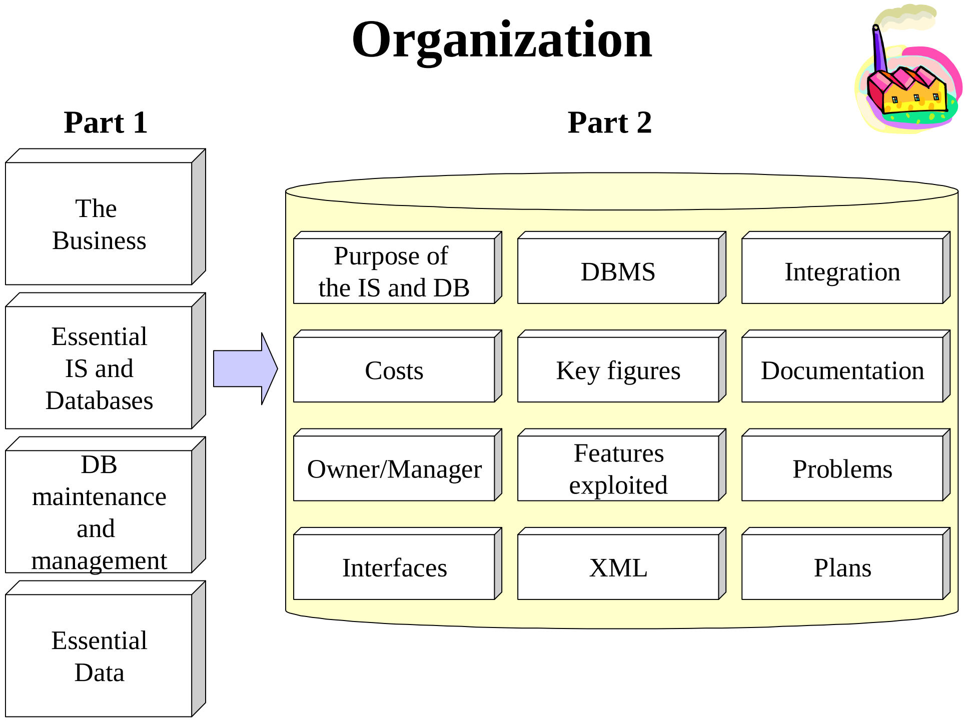


Figure 4.3: The themes of the interviews.

The first part of the interview concentrated on the general environment of the information systems and the second part concentrated in detail on the specific information systems and databases. The purpose of the first part was to get acquainted with the line of business the interviewee represented, the important data that was managed and produced and the important information systems and databases in use.

The second part looked in a more detailed and technical way at the essential information systems and databases identified in the first part of the interview. If more than one information system and database were identified this part was conducted several times with each relevant interviewee. The second part of the interview discussed the following themes of the information system and database in question:

- The purpose of the IS and the database.
- License, maintenance and administration costs.
- Owner/Manager of the IS and the database.
- Different interfaces of the IS and the database.
- The database management system (DBMS).

- Key statistics associated with of the information system and the data related to it, i.e. their size (in bytes, in number of the rows, in number of the tables), number of users, number of concurrent users, number of transactions, type of transactions.
- Features exploited in the database, i.e. object types, LOBs, triggers, stored procedures, views.
- XML and information systems, i.e. whether XML documents are produced from the database, or stored in the database, whether the information system can receive XML documents and the level of XML knowledge.
- Integration with other information systems and databases.
- The state of the documentation.
- Problems and critical issues, i.e. typical problem situations, usage breaks, shortcomings of the current database.
- Future plans in the field of databases, i.e. plans with object features, XML data management and other issues.

The translated interview script is presented in Appendix 2.

4.4.3 Data Collection Process

Each case began with at least one preliminary meeting with the contact person or persons in the organization. These meetings defined the specific research field, identified the interviewees and finalised necessary preparations for the investigation. Usual protocols suggest that the contact person informs the interviewees of the project and the fact that a researcher will contact them to agree the interview schedule. On most occasions, documentation describing or relating to the specific research field was received during these meetings. Working in this way, the organization could ensure that adequate preparation for the study has been carried out, which was one concern of Broadbent et al. (1998).

After the contact person had informed the interviewees, the researcher either phoned the interviewee directly to agree the interview time or sent an email suggesting some possible dates, to which the interviewee responded. Some days before the actual meeting the script was sent to the interviewees. The purpose of this was both to remind the interviewee about the forthcoming meeting, to orientate the interviewee for the occasion and also give them time to check some details relating to its themes.

All interviews began by presenting an overview of the whole research and explanation of how the particular case fitted into the overall project. The slide presented in Figure 4.4 played a central role in this presentation.

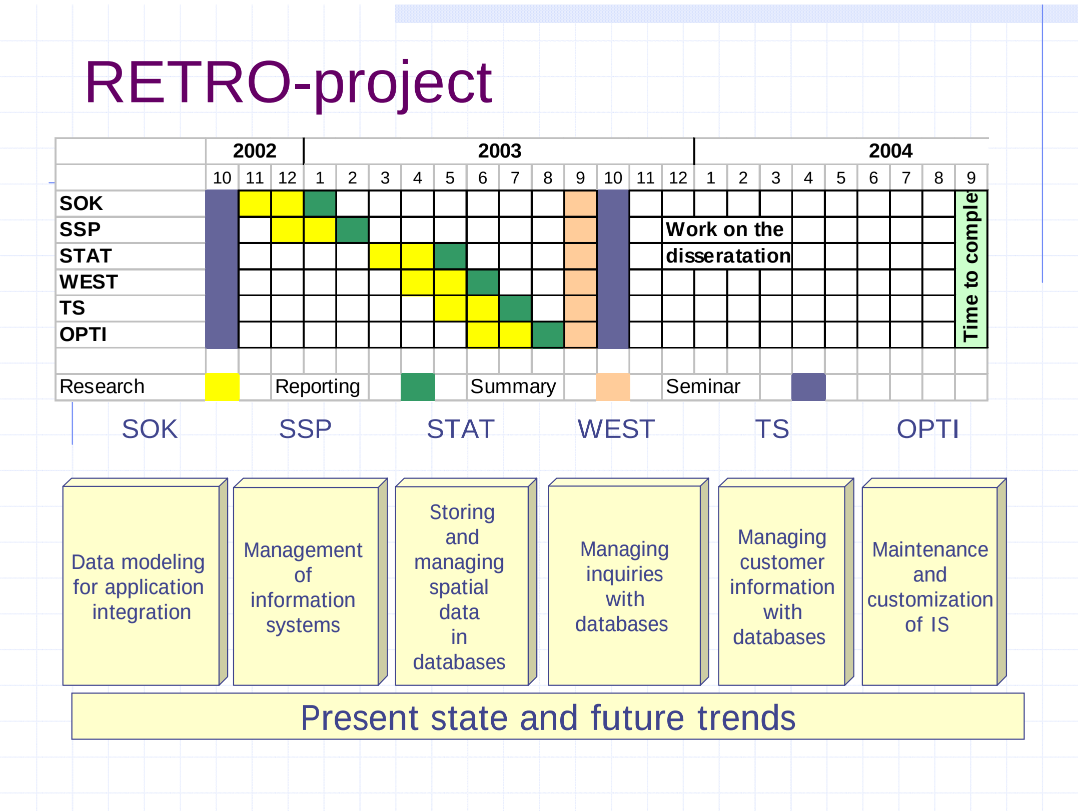


Figure 4.4: Introduction slide used in interviews.

All interviews were recorded and notes were taken that were transcribed afterwards, before being returned to the interviewee for corrections and possible additions. The main advantage of tape-recording is that it provides a full description of what was said, but a disadvantage is that it requires lot of time to transcribe the recordings (Walsham 1995). Additional materials were received during the interviews, and these were put into a database, along with the transcriptions, once the discussion was completed. This database application was developed to improve and to ease the analysis of the information, since many different search conditions could be used. It also met the following requirements:

- The researcher should be very accurate in record keeping. Vital data may be lost if entrusted to memory and not organized as soon after collection as possible (Benbasat et al. 1987).
- The case study data needs to be organized in a way that will ensure ready access to the case data at any point during or after the study (Yin 1994).

The interviewees made some corrections and additions to the interview transcriptions, but altogether their changes were quite minor. Typically, only some small details were changed.

According to Yin (1994) three principles should be followed during the data collection:

- Use multiple sources of evidence.
- Create a case study database.
- Maintain a chain of evidence.

In this research multiple sources of evidence were exploited to achieve so-called 'triangulation'. The use of multiple data collection methods was also justified since triangulation 1) can alert potential analytical errors and omissions and 2) can lead to new insights and modes of analysis (Duchon et al. 1988). Yin (1994) used the term 'case study database' to describe the collection and archiving of all case relevant data. This research also followed this principle by keeping all of the data together and easily accessible. The third principle allows the reader of this report to follow the derivation of any evidence, ranging from the initial research questions to case study conclusions.

4.5 Data Analysis

The analysis of case study evidence is especially difficult because the strategies and techniques are not well defined (Yin 1994). However, it is important to describe the methods used in data analysis to guarantee that the chain of evidence is not lost. This will also improve the reliability of the data. (Benbasat et al. 1987) A project becomes more credible when the researcher describes in detail how the research results were achieved (Broadbent et al. 1998). The goal of the data analysis in this kind of interpretive case research is to produce an understanding of the context of information systems and the interactions between these systems and their concepts (Broadbent et al. 1998). The researcher develops categories and meanings from the data through an iterative process that starts by developing an initial understanding of those being studied (Duchon et al. 1988). In an interpretive case study four possible types of generalizations are identified (Walsham 1995): development of concepts, generation of theory, drawing of specific implications and contribution of rich insight. In this research the main type of generalization was to create rich insights into the cases.

In this multiple case study project the data was analysed in three phases. The first phase involved writing the single case reports, the second was concerned with writing the first part of the cross-case report and the third consisted of writing the second part of the cross-case report. Table 4.4 gives additional details about these phases.

Table 4.4: Data analysis in this research.

	Single case analysis	**Cross-case analysis I**	**Cross-case analysis II**
Focus	• The specific research area	• The empirical research questions	• The theoretical research questions
Aims	• Description of the case	• Description of the database trends	• Explaining the diffusion of the database innovations
Analysis	• Interpreting the interviews with additional materials and observations	• Building categories and classifications and describing the field of databases • Quantitative analysis	• Describing the cases against the developed framework • Describing the database innovations against the developed framework • Qualitative analysis

Cross-case analysis I involved handling each interview theme individually and building categories and classifications. The classifications were not based on any specific absolute metrics, instead they were defined from the thoughts and feelings of the interviewees associated with the topic in question. The categories and classifications are presented in Table 4.5.

Table 4.5: The categories and classifications used in data analysis.

Interview theme	**Categories/Classifications**
What kinds of database solutions are currently exploited?	DBMS Generations • 2nd generation: Relational database • 3rd generation: Object-oriented database • Enriched 3rd generation: XML database Type of information systems • On-Line Transaction Processing System • Decision support system, Reporting • Enterprise Resource Planning system • Information channel system Importance to the business • Business Critical IS • Supporting IS Interfaces to the database • PC-terminal • Client/Server • Browser • Mobile
What is the state of data modelling considering the essential data?	Data modelling needs in the cases State of the enterprise data model • Not defined • Data is gathered, but not modelled • First level data model designed during the research • Data model designed during IS implementation
How is the maintenance and administration of databases organized?	IS maintenance • Outsourced • Mostly outsourced DB maintenance • Outsourced • Mostly outsourced • In-house Main user –ideology • Yes • Partly State of documentation • Unsatisfactory, outdated • Documentation started • Almost satisfactory • Satisfactory Documentation template or standard in use • Yes • No

Table 4.5: The categories and classifications used in data analysis (continued).

Interview theme	**Categories/Classifications**
How are databases integrated in the organizations?	Implemented integration solutions • Manually • Email attachments • FTP file transfers • System-to-System • EDI • Database link • XML message
How is XML used and/or will be used in combination with databases in the case organizations?	Levels of XML use with databases • Level 0: XML not used at all • Level A: XML produced from the data in the database • Level B: XML data stored in relational tables • Level C: XML data stored in CLOBs • Level D: XML data stored in native format
What kind of black spots and shortcomings current database and information systems have?	Problem areas of IS and databases • Logical structure of the database • Data integrity • DBMS • Response times and hardware • Integration • Reporting • Personnel
What plans do organizations have to consider developing databases and information systems?	Development plans • Data modelling • DBMS upgrade/replacement • IS upgrade/replacement • Integration, Web Services • Browser based interfaces • Hardware upgrade/replacement • Strategy

During the cross-case analysis II each case organization was characterised according to the organizational part of the developed framework presented in section 3.4. This phase also involved analysing the database innovations, but against the innovation part of the framework. The researcher evaluated all attributes of the developed framework on a five-level scale: Very Low – Low – Moderate – High – Very High. These evaluations were converted to numbers 0 to 4 corresponding to the previous scale. However, three attributes have negative influence adopting database innovations. In these cases the scale was turned upside down in the conversion to numbers. For example, if an organization's IS portfolio were evaluated highly it would be converted to 1. Finally, all numbers were added up to calculate an orientation value towards the adoption of database innovations and this was expressed in percentage terms. This made the maximum orientation value 44 (100 percent) and the lowest orientation value zero (0 percent). The orientation value is not a timeless indicator, rather it describes the situation at certain point of time. The higher the value, the better prerequisites the organization has to adapt to current database innovations. Thus the evaluation tries to point out areas in which improvements are needed to enhance innovativeness.

The aim of this analysis of database innovations and framework is to describe how database innovations can help and improve business processes. This means producing innovation-evaluation information that according to Rogers (1995) answers questions like *what are an innovation's consequences?* and *what will its advantages and disadvantages be in different cases?*

5 DATABASE ARCHITECTURES AND TRENDS IN THE CASE ORGANIZATIONS

This chapter presents the results of the empirical part of this research, organized around the interview themes.

5.1 Description of the database environment

The research analysed a total of 44 different information systems. Most (40%) of these were online transaction-processing (OLTP) systems (Figure 5.1), such as the order-entry systems in the SOK Corporation, Optiroc Ltd, Salon Seudun Puhelin Ltd and TS-Group Ltd. The second biggest category was decision support and reporting systems (30%). There were no data warehouse systems, but customized reporting systems were used by SOK, Salon Seudun Puhelin and Optiroc. All accounting systems were also classified in this category. A quarter of the total were enterprise resource planning systems like the production management system at Optiroc. In addition, two were interactive information channels for customers. Most of the studied information systems (75%) were produced by external IS/IT-companies. The remaining quarter were developed by the organization itself.

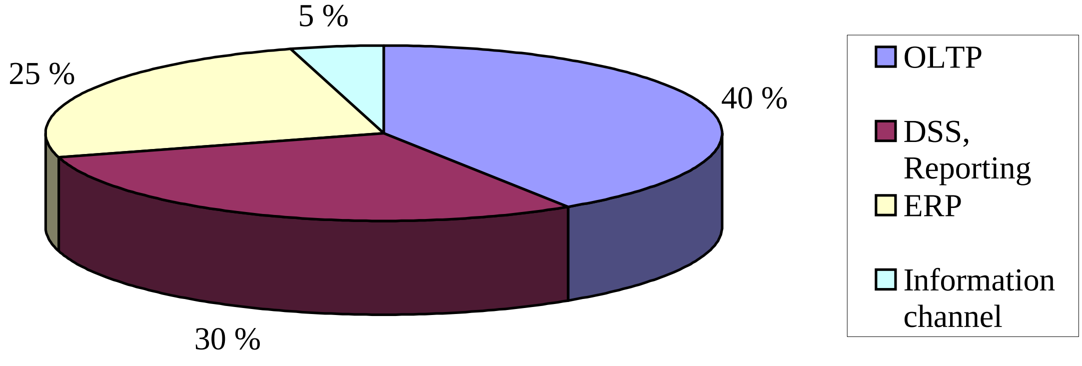


Figure 5.1: Distribution of types of the information systems.

Over half (52%) of the information systems and the databases studied during this research were used in business-critical processes. The study identified business critical information systems in all cases except the State Provincial Office of Western Finland. Figure 5.2 shows how business critical information systems were categorised. As the Figure shows, most (65%) of them were online transaction processing systems. Analysis of variance shows also that the differences between IS type categories are statistically very significant (p=0.0016). In the OLTP-category 83% of the information systems were business critical, but this was the case with only 23% in the DSS, Reporting–category and only 45% in the ERP-category.

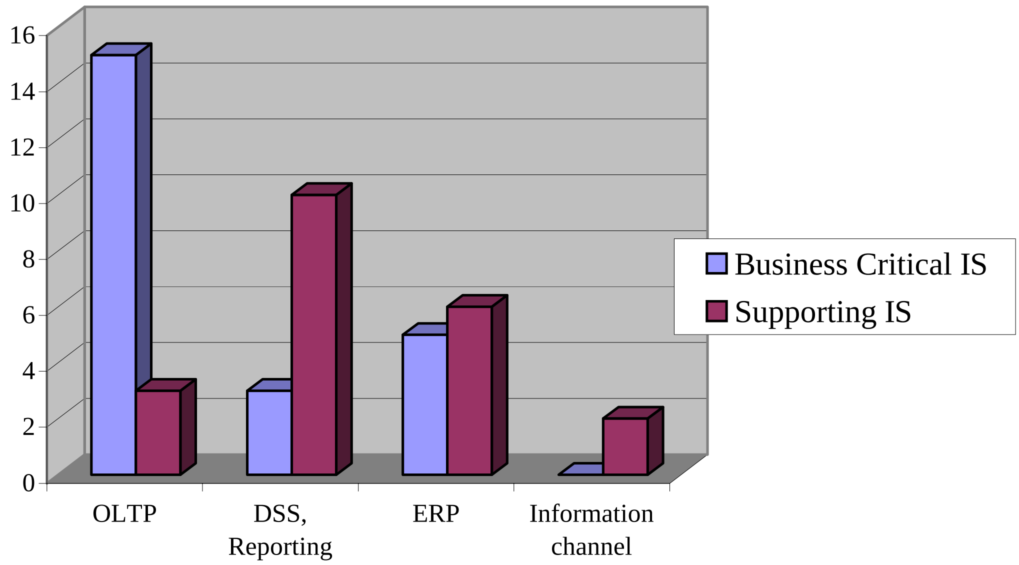


Figure 5.2: The distribution of business critical information systems.

On average, the databases in the OLTP-category were a little over 9 Gigabytes in size (in fact they ranged from 0.5 GB to 50 GB, with a median of 4.5). The number of concurrent users in the OLTP-category ranged from 4 to 200, with a mean of 84 (median 50). The biggest database was SOK's reporting database, which was 2 Terabytes.

The average implementation year of the information systems was 1997, making their average age almost 7 years. The newest information systems studied during the research were implemented in 2003 and the oldest as far back as 1985. Within different cases their average age ranged from 4.6 to 10.7 years. The average year of implementation in the various IS-type categories is presented in Table 5.1. As the table shows, online transaction processing systems were the first to be implemented, on average almost 9 years ago. In contrast, the average date for implementing their supporting information systems was 1999. This difference between these two categories is also statistically significant (p=0.0048), indicating that business critical information systems are not easily changed just to follow new technological flows.

Table 5.1: Average year of implementation for different categories of information system.

Category	Average year of implementation
OLTP	1995
DSS, Reporting	1998
ERP	1998
Information channel	2001

The majority (68%) of information systems were implemented in 1999 or earlier (see Figure 5.3). Their ages already show that the database technology behind the information systems is mainly relational DBMSs. Actually, relational DBMS were used in all of the information systems studied and no information systems exploited even third generation database technology, not to mention any further enrichments of them. Although some DBMSs represent the third generation, their information systems utilize only the features of the second generation – in other words, they do not utilize user-defined types, collections or LOB-types.

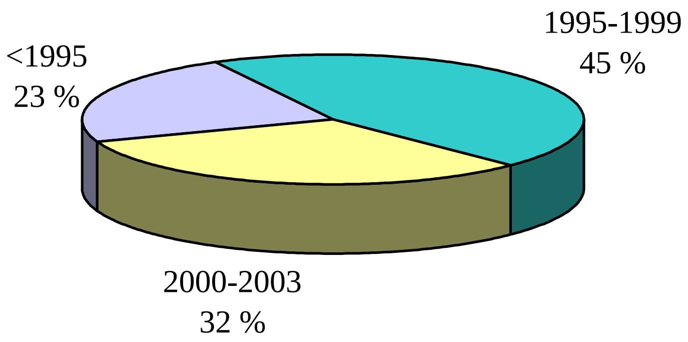


Figure 5.3: Date of information system implementation.

Typically, the organizations use more than one DBMS. The normal reason for this is that each information system requires a particular DBMS. Only in Optiroc was a decision made that all the information systems should utilize Oracle DBMS. Most of the information systems utilize one of the three major DBMSs (Oracle is involved with 29.5% of them, SQL Server 29.5% and DB2 27.3 %, although the distribution is more balanced here than it was in a recent market share analysis, which gave Oracle 39.4 %, IBM DB2 33.6% and Microsoft SQL Server 11.1% (IDC 2003)). The use of many different DBMSs in the case organizations can increase the pressure and challenges on the administration and maintenance of the databases.

Three quarters of the information systems had only one kind of interface to the database, although the remaining 25% had multiple interfaces. The focus of these interfaces between the information systems and the databases was slightly different in the various organizations (Figure 5.4). Altogether, 59% of the information systems had at least a client/server interface. In practice this means that in these information systems, client software installations are required to the client computers. Only at SOK were there possibilities to connect to the information system through a mobile device. In addition, SMS-messages are used at SOK to send additional information to customers. At Optiroc all studied information systems had a browser-based interface.

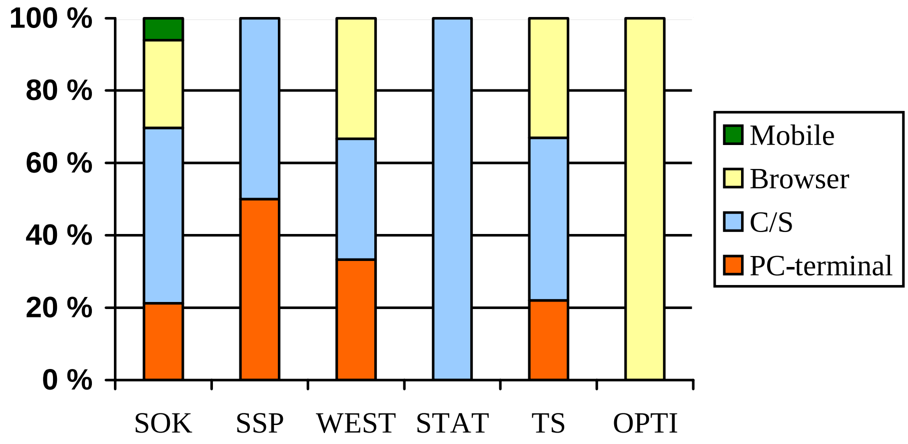


Figure 5.4: Percentages of different interfaces per organization.

5.2 State of data modelling

The interviews confirmed the importance of the enterprise data model when designing and implementing integration between information systems and databases. The interviewees mentioned that the enterprise data model helps to understand the information systems and their possible and existing connections. In three cases (SOK, State Provincial Office of Western Finland and Statistics Finland) the investigation expected a first level data model to result from the research (see Table 5.2). Almost all cases regarded data modelling as a very important and a critical part of IS development. Only in Salon Seudun Puhelin was the role of data modelling not emphasized. The results show that in all other cases one major rationale for data modelling is the need to integrate information systems and databases. This is not a surprise when one considers the theory that a data model is a central starting point for integrating information systems.

TS-Group Ltd gathered a large amount of background data during its CRM project, but surprisingly did not design a data model for customer management, even though the interviewees agreed that it is essential to define what is actually a customer. In Salon Seudun Puhelin Ltd the focus was on modelling processes, rather than data modelling. This is based on the idea that process modelling can improve processes and decrease problems.

The information systems studied in this research manage mainly traditional types of data, i.e. numbers, characters, strings and dates. However, the case organizations had very challenging data modelling environments. For example, modelling a product at SOK proved to be very complex, because the corporation handles such a broad variety of products. The same model would need to describe anything from sweets to tractors and from milk products to home electronics. In Statistics Finland and the State Provincial Office of Western Finland data modelling has got to the stage where complex objects need to be included in the data model. They both have plans to use geographical data in their future information systems and these points, lines and polygons are now appearing in their data models. XML on the other hand has not yet influenced data modelling in any of the case organizations.

Table 5.2: State of data modelling in case organizations.

Case	Rationale for data modelling	State of data modelling
SOK	Integration of information systems	First level data model designed during the research
SSP	No data modelling needs recognized; Modelling of processes seen more important	Not defined
WEST	New information system and integration of information systems	First level data model designed during the research
STAT	Management of spatial data and integration of information systems	First level data model designed during the research
TS	Management of customer data and integration of information systems	Data is gathered, but not modelled
OPTI	New production planning system and integration of information systems	Data model designed during IS implementation

5.3 Maintenance and administration of the databases and information systems

None of the organizations does business in IS development and therefore it is no surprise that information system maintenance is mostly or totally outsourced (see Table 5.3). The only exception is Statistics Finland because it needs special information systems for statistical analysis, which are developed in-house. Neither SOK, Salon Seudun Puhelin nor TS-Group rely totally on outsourcing, as they all run information systems that were developed in-house[11].

Only the State Provincial Office of Western Finland and Optiroc have outsourced database administration, although this is also the case for SOK's information systems that were developed elsewhere. Three other organizations employ their own personnel to administer the databases. In practice, database administration includes DBMS upgrades, backup, optimisation, user management, quota management and problem-solving. All organizations sought to coordinate, identify and discuss their user's comments and requirements with the information system supplier. SOK formed large user groups to draw out this information.

[11] Actually 25 % of all studied information systems were developed within the case organizations.

Depending on the selected maintenance organization, the practical processes might also be influenced. Total outsourcing was seen as a risk, since it means that knowledge of the operational information systems and databases is held outside the organization. It might also slow down development processes, because the developers are not bound to the organization itself. In addition, a recent study argues that without in-house expertise, a company cannot understand the viability of addressing new demands, or the potential for meeting existing demands on a new technology platform (Feeny et al. 1998). Problems were also identified in strong main user –ideology, which means that an information system has a responsible person that acts as middleman between the users and the IS vendor. In these cases one information system might be emphasized over the others and the IS wholeness could be disturbed.

Table 5.3: Information systems maintenance issues in the case organizations.

Case	IS maintenance	DB administration	Main user–ideology
SOK	Mostly outsourced	Mostly outsourced	Yes
SSP	Mostly outsourced	In-house	Partly
STAT	Mainly themselves	In-house	Yes
WEST	Outsourced	Outsourced	Partly
TS	Mostly outsourced	In-house	Yes
OPTI	Outsourced	Outsourced	Yes

As mentioned earlier, documentation plays an important part in the maintenance process. The quality of information system and database documentation studied in this research proved to be quite high (see Figure 5.5). It was ‘satisfactory’ or ‘almost satisfactory’ in 68% of the cases. In a further 9% of the information systems and databases a project has begun to produce or update the documentation to a satisfactory level. The remaining 23% were ‘unsatisfactory’ or ‘outdated’. There was no correlation (-0.05) between the age of the information systems and the state of their documentation. The analysis of variance also showed no difference in the documentation between business critical and supporting information systems (p= 0.856). Similarly, there was no difference in the quality of documentation in the different IS categories (p=0.135). We analysed whether the interface and the state of the documentation had any relationship, but again there proved to be almost no correlation (-0.19). The average quality of documentation in the various interface groups followed the correlation coefficient. For example, the PC-terminal group had the highest average (3.2) and figures for the others were slightly smaller (for client/server it was 2.85, browser 2.71 and mobile only 2). We might think that the newer interfaces are more self-explanatory and therefore less documentation is needed. However, the analysis of variance did not prove any significant statistical difference between the groups (p=0.489).

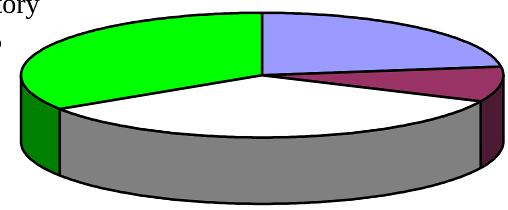


Figure 5.5: State of documentation.

Although 68% of the information systems documentation was regarded as ‘satisfactory’ or ‘almost satisfactory’, this still leaves a significant amount (23%) that was unsatisfactory or outdated. Most of the information systems (75%) were developed externally by an IT company. The average interviewee opinion of this information system documentation was 2.56, which is less than ‘almost satisfactory’. The rest of the information systems (25%) were developed in-house and the average figure for this documentation was 3.45. The averages of these groups were analysed with t-test and it showed, at 5% significance level, that there are significant differences (p=0.019) in the state of the documentation depending where is was developed, thus indicating that in-house information systems documentation is significantly better than that which comes from a third party.

Statistics Finland and TS-Group were the only organizations to define documentation templates or standards. For Statistics Finland this has been successful, as the quality of information systems documentation was good. Salon Seudun Puhelin also achieved a satisfactory result. In SOK and TS-Group part of the information systems were documented satisfactorily, but some were unsatisfactory or outdated. SOK has recognized this and initiated several documentation projects to improve the situation. One reason for this at TS-Group is that the documentation template has not been updated, meaning that use of it is decreasing. Optiroc's documentation is partly satisfactory, but it does not detail the essential rules and logics of the application programs. In the State Provincial Office of Western Finland the documentation is almost satisfactory.

CASE tools have been employed for documentation purposes in a few information systems. For example, Optiroc has used Oracle Designer for system development and documentation. WEST and SOK have documented (or tried to document) their database structures with Erwin CASE tool.

5.4 Present integration solutions

Currently, the information systems and databases in the case organizations exploit many different integration solutions. Among these solutions seven different categories were identified (see Table 5.4).

Table 5.4: Description of the identified integration categories in this research.

Category	**Description**
Manually	The data is read from one information system and then inserted into another system manually.
Email attachment	An individual from one information system produces a data file and sends it on. The file is then processed and written into another information system by some program.
FTP file transfer	A data file is produced and transferred automatically from one information system to another. A program processes the file and writes the data into the second system.
System-to-system	The data is transferred automatically from one information system to another by an additional.
EDI	The data is transferred in large batches from one information system to another. Additional programs/systems are needed to convert the data into the correct format.
Database link	An information system accesses the database of the other information system directly with SQL.
XML message	An information system requests some data with an XML message and the other information system produces an XML message containing the necessary data in real time.

The most primitive integration solution category identified in this research is manual integration, which means that the data is first queried from one information system and then updated manually in the other information system. A little more sophisticated is the email attachment category, where a file is generated from the first information system and emailed to someone working with a second system. The received file is then processed and the data employed by the other information system through an application. Also here a certain level of manual work is required. The third category is FTP file transfer. Here a file is automatically processed and transferred to the other information system where it is automatically processed in batches. System-to-system integration means that data moves automatically between information systems, but not directly from database to database, instead it may travel via a temporary file. The EDI-category is for solutions that exploit EDI-messages in communication between information systems, typically transferring large batches of data. Database link means that one information system is able to read and write directly to the database of another information system. Database links were the main integration method in Optiroc, since their DBMS environment was standardized to Oracle, but in other organizations a number of different database management systems were used, which meant that the usage of direct database links was not so natural. XML-message is the last integration category. It means that data is transferred in real time in XML format from one information system to another. To sum up the integration categories are presented in the Figure 5.86.

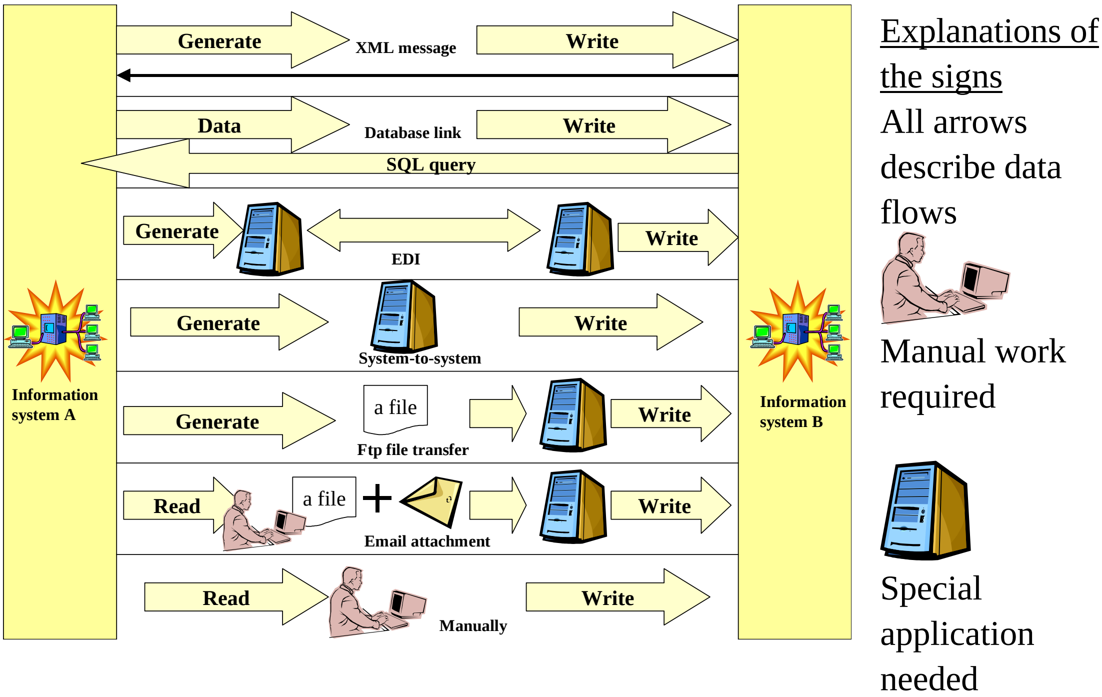


Figure 5.6: Summary of the integration solutions.

Table 5.5: Integration solutions utilized in the case organizations.

Case	Manually	Email attachments	FTP file transfers	System-to-System	EDI	Database link	XML-message
SOK	X	X	X		X		X
SSP			X	X		X	
STAT		X	X	X			
WEST	X	X					
TS	X	X	X	X	X	X	X
OPTI			X		X	X	

Table 5.5 shows how different organizations have integrated their information systems. It shows that the most typical integration solution is currently FTP file transfer, which is used in five case organizations. According

to the interviews, only SOK and TS have utilized XML-messages, but also in these cases the use of XML-messages is an exception and typically old-fashioned integration solutions are exploited. WEST has not yet really started the integration of its information systems. The integration solutions exploited in the case organizations can also be grouped according to the degree of internal integration (Enterprise Application Integration; EAI) and with other organizations (B2B). Table 5.6 shows how the case organizations have exploited EAI and B2B integration solutions.

Table 5.6: EAI vs. B2B solutions in the case organizations.

	Integration within the organization (EAI)							**Integration with other organizations (B2B)**						
Case	**Manually**	**Email attachments**	**FTP file transfers**	**System-to-System**	**EDI**	**Database link**	**XML-message**	**Manually**	**Email attachments**	**FTP file transfers**	**System-to-System**	**EDI**	**Database link**	**XML-message**
SOK	X		X				X	X	X	X		X		
SSP			X			X				X	X			
STAT				X					X	X				
WEST	X	X							X					
TS		X		X		X		X		X		X		X
OPTI			X			X				X		X	X	

The table shows that EDI, FTP file transfers and email attachments are typically used in B2B integration efforts. On the other hand, system-to-system and database link integration solutions are typically used within the organization.

At SOK, FTP file transfer is the fundamental integration solution, both internally and with partner organizations. Examples of FTP file transfer integration within the organization include the file transfers of a) invoicing data to the central financial management IS and b) product and price data to the local cash systems in grocery shops all over Finland. An example of B2B integration is the file transfer from a car manufacturer in France that sends data about exported cars via FTP. Purchase invoices or messages containing data about suppliers' products are received through EDI. Email attachments are used for example to deliver sales statistics in Excel-files to suppliers.

In Salon Seudun Puhelin two main information systems are integrated with each other through a database link. From these operational systems the data is transferred via FTP to the reporting system. FTP file transfers are also used to deliver telephone call data from call centres to Salon Seudun Puhelin. This company also uses system-to-system integration: Electronic Worldwide Switch Digital (EWSD) centres deliver telephone call data three times a week to its server, where a special program processes this information.

For STAT, the most important integration solution is with the Finnish Population Register Centre. It transfers the spatial data of people and buildings via FTP to this body. Within the organization the information systems are also integrated using system-to-system solutions. For example, basic data to produce statistics about population centres is transferred from other systems to the production system using this method.

At WEST, basic service evaluation data is received as email attachments from the Centre for Research and Development of Welfare and Health. Some data is also queried from one information system and manually inserted into the other information system. For example, operating permits are managed this way.

TS has used EDI for the last 10 years and it is their basic integration solution. For example, paper mills send their freight invoices in EDIFACT-format. EDI is also used in communication with 30 major newspaper publishers, to deliver advertising information, for example. However, order confirmations are based on XML-messaging. This method is also used to send customer advertisements to the online advertising administrator. Customer advertisements for the local newspaper are collected via a web-based application, which sends on the data as email attachments to a person working in the newspaper production system. A similar solution exists with newspaper orders. A customer can make an order through a web-based application, which is delivered to the sales department as an email attachment. FTP file transfers are also used, for example to provide data about newspaper delivery to other newspaper publishers. In addition, database links and system-to-system integration solutions are exploited within TS-Group.

Within OPTI, database links are the basic integration solution as already mentioned. Also FTP file transfers are utilized, for example when factories send production statistics to their main office. OPTI and its partners have developed database links across organizational boundaries, which are used as B2B integration solutions that can read and write privileged data from each other's databases. Currently, these database links manage the data that is necessary for organizing product transports. EDI messages are used for example in delivering invoices to a wholesale firm.

5.5 Role of XML

Most of the studied information systems do not deal with XML documents at all (see Figure 5.7). The databases represent the 2nd generation of databases, which have limited XML capabilities. With these XML can only be produced from the data in relational tables and XML documents can be split into pieces and stored in relational tables. Three of the cases had information systems that can manage XML data, but most of their systems do not deal with XML data at all (see 'Now (mode)' bar in Figure 5.8). At SOK, Optiroc and TS-Group, XML-documents are produced from relational data. SOK and Optiroc also have information systems that store XML-data in a relational database. Often, information systems that have been renewed/upgraded within the last couple of years also exploit XML data in some way. However, all case organizations have plans to exploit XML in the future and typically these plans related to the integration of information systems. Actually, the cases where XML is stored in a database are examples of data transfer from one information system to another. Figure 5.7 presents how XML and databases are combined now and how the situation will change in the next one to three years, according to the interviewees.

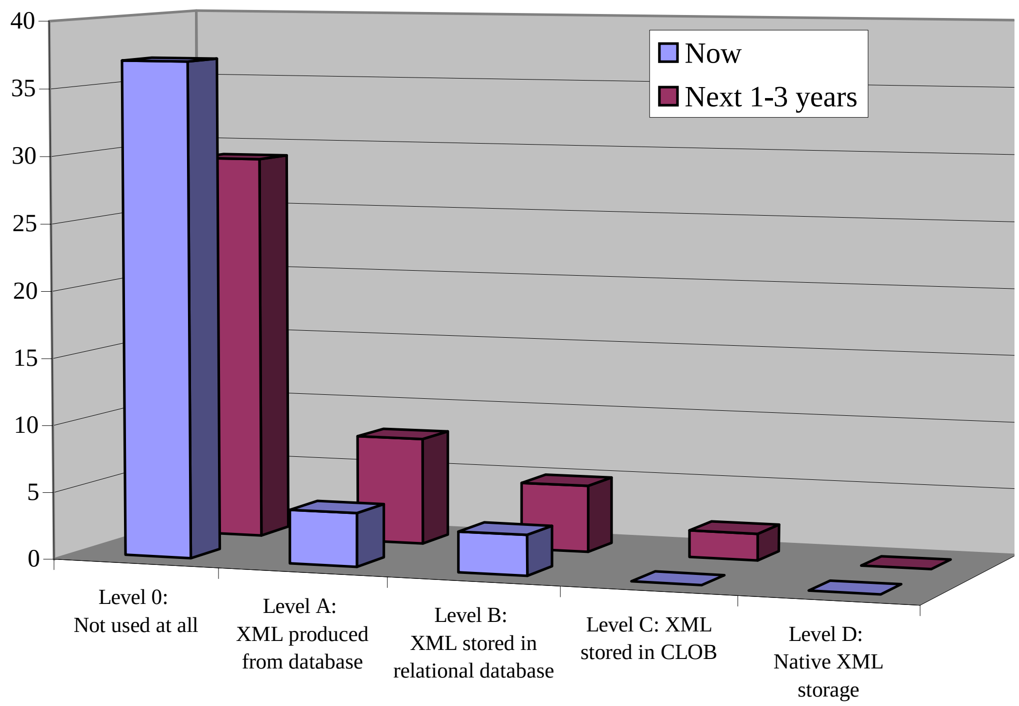


Figure 5.7: Information systems managing XML now and plans for the future.

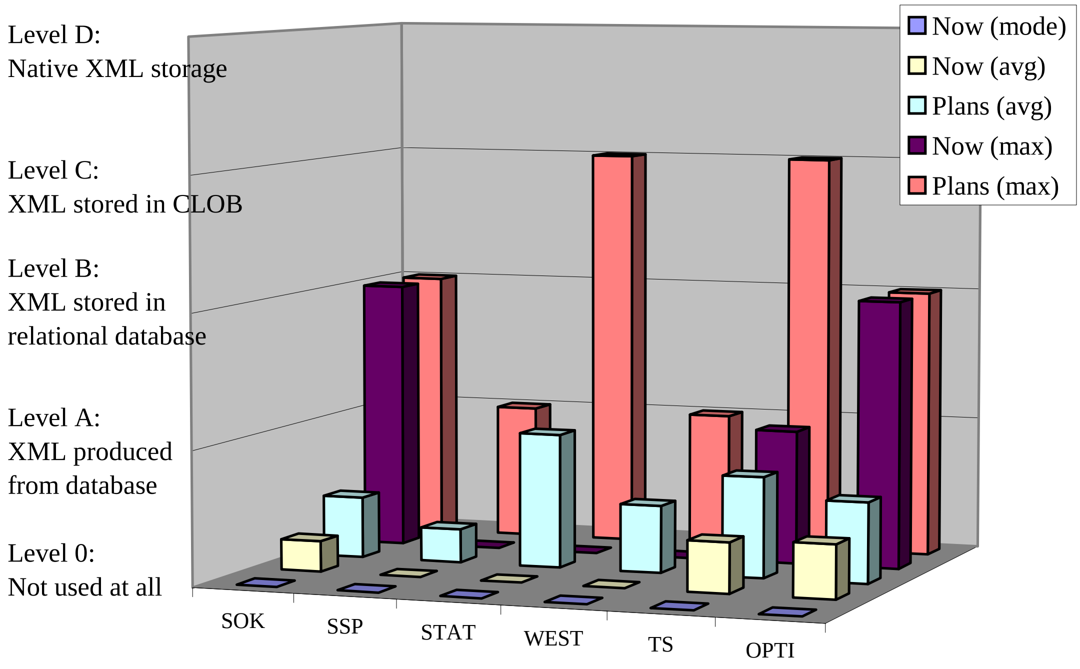


Figure 5.8: Role of XML in the case organizations.

Figure 5.8 shows the current mode, average and highest levels of XML usage in each of the case organizations. It also shows the average and highest levels according to the plans outlined by interviewees. As the non-existing ‘Now (mode)’ bars show, XML is not yet a typical part of the information systems and the databases. However as the ‘Now (avg)’ bars show, three cases (SOK, TS, OPTI) have XML-enabled information systems and database solutions. The ‘Now (max)’ bars show the highest level of XML usage of all studied information systems in each case. These bars show that both SOK Corporation and Optiroc Ltd have at least one information system that stores XML in a relational database. The bars also show that in TS-Group an information system or systems exist that produces XML from the database. Comparing the ‘Now’ and ‘Plans’ bars shows the trend that XML is becoming increasingly important to the information systems and databases. In all cases the ‘Plans’ bars are higher than the corresponding ‘Now’ bars. For example, in TS-Group the average and the maximum levels of XML usage are increasing. A more complete description of the role of XML in the case organizations can be found in (Kontio 2004) and in (Kontio 2003).

The interviews show also that using XML has been discussed within the organizations and technological solutions will be adopted later. It has also been recognized that the level of XML knowledge is not sufficient in the case organizations and therefore some of them (including Statistics Finland and TS-Group) have already arranged for their staff to undergo some XML training. However, it should be recognised that the lack of XML knowledge is not such a critical problem when we remember that development of the information system is mostly outsourced.

The organizations also have plans to develop their integration solutions and in four cases (SOK, Statistics Finland, TS-Group and Optiroc) XML-messages are mentioned as the future goal. They have realized that XML messages can eliminate the problems they have with FTP file transfers. However, moving to XML messages requires an information system that is able to both produce XML and store it in databases.

5.6 Role of spatial data

Spatial data management was the specific research field in Statistics Finland, but spatial data also proved to play a role in other organizations. Table 5.7 shows examples of spatial data exploitations in the case organizations. Only at SOK did the interviewees mention no examples of utilizing spatial data.

Table 5.7: Examples of exploiting spatial data in the case organizations.

Case	**Example**
SSP	Structure of data communication network
STAT	Production of statistics
WEST	Presentation of evaluation results
TS	Distribution of newspapers
OPTI	Delivery arrangements

Although many of the present database management systems are able to manage spatial data through their extensions, none of the organizations exploit these functionalities at the moment. Instead, spatial data was typically stored in files and managed with special software. However, there were signs that these possibilities will be exploited in the future. For example, in Statistics Finland all spatial data will be moved from sequential files to a spatial database.

The interviewees also mentioned that spatial data could be utilized in a much more versatile way than has been the case previously. At TS-Group, the interviewees mentioned that addresses and coordinates are becoming increasingly important. For example, the future CRM system could utilize spatial data quite effectively. In the State Provincial Office of Western Finland, spatial data could be combined with other evaluation information in a database, thus enabling more versatile analysis and reporting possibilities. Staff at Statistics Finland had a similar rationale for investigating this, whilst also wanting to combine many data sources with spatial data to offer new possibilities for analysis. They were especially interested in maintaining the integrity of geographic data within a spatial database. In addition, they wanted to use SQL functions and procedures for the analysis and therefore avoid the necessity to write special programs for this.

5.7 Problems in the information systems and databases

The interviewees identified many different problems with the present information systems and databases. These were grouped into seven categories, as listed in Table 5.8. One should keep in mind that these problems are not universal, the organization has to deal with these problem areas in some of the information systems and databases.

Table 5.8: Identified problem categories in the information systems and databases.

Case	**Logical structure of database**	**Data integrity**	**DBMS**	**Response times and hardware**	**Integration**	**Reporting**	**Personnel**
SOK	X	X	X	X	X	X	X
SSP		X	X	X		X	X
STAT	X	X			X		
WEST	X				X	X	
TS	X	X	X	X	X	X	
OPTI				X		X	X

5.7.1 Logical structure of the database

The category ‘logical structure of the database’ covers problems inside the database structure. Typically this means that:

- The database design has not succeeded in capturing all of the necessary elements; or
- The logical structure should be (re)designed; or
- The requirements have changed over time, but changes have not been implemented in the database.

In one of the information systems at SOK, the environment is so new that the logical structure and the concepts have not yet settled down. In another, the system has been upgraded in a piecemeal fashion over the years, which has resulted in a myriad of tables that blurs the logical structure. At Statistics Finland, the environment has changed so much that there is a need for a new logical data model. In the State Provincial Office of Western Finland, plenty of data is gathered but no real data modelling has been done and therefore this information cannot be utilized in an efficient manner. As described earlier, TS-Group is implementing a CRM system, which requires its customer information to be modelled logically.

5.7.2 Data integrity

The second category, ‘data integrity’, refers to the validity and consistency of stored data (Begg et al. 2002). These problems usually come from the duplication of relevant data meaning that the same information is managed independently in different systems, without any integration. For example, Salon Seudun Puhelin stores its customer information in two different information systems, even though it could use common tables for customer data. In one of SOK‘s information systems customer data can be inserted in the database more than once, generating additional work and tricky situations. Finally, Statistics Finland stores the same data in many places, which can make it inaccurate during distribution.

5.7.3 DBMS

The third category, ‘DBMS’, includes all the problems relating to the use of the DBMS. These include its stability/reliability, lack of adequate tools, limited capability to manage the data and other similar problems. In one of SOK’s information systems the database corrupts easily and requires new indexing, which results in unnecessary downtime. Salon Seudun Puhelin moves original data from operational systems to Microsoft Access for use in reports. However, the volume of this data is too great for the capacity of MS Access and temporary solutions have been developed to deal with this issue.

At TS-Group a DBMS does not include any tool for ad hoc queries and indexing choices are limited.

5.7.4 Response times and hardware

The fourth category, 'response times and hardware', represents mainly problems relating to hardware performance and age. These include poor response times (for example at Optiroc), although in most cases these are only temporary rather that everyday problems. Optimising the database usually solves them. At Salon Seudun Puhelin the hardware is becoming so old that reliance on it constitutes a risk. Similarly, TS-Group is considering whether to give up the mainframe computer. In those organizations where old mainframe computers were used, a number of interviewees mentioned that usually the hardware is blamed for the difficult and boring character-based user interfaces.

5.7.5 Integration

The fifth category, 'integration', concerns difficulties organizations may have in transferring data from one information system to another, using one of the integration solutions described earlier. Integration can be problematic for various reasons:

- Data transfer is uncertain and difficult.
- It is not clear whether the data is accurate.
- Integration is currently expensive and time-consuming.
- Integration requires too much tedious manual work.

A basic problem in the case organizations is the lack of integration of the information systems and databases. There are still many information systems that are not integrated anywhere, or if they are only non-automated solutions exist. Poorly integrated solutions also produce inefficient business processes. For example, SOK delivers a large amount of paper reports to locations all around Finland, since it has no electronic archive. As mentioned earlier, some data is manually processed, which also slows down processes. Furthermore, the organizations often find it difficult to make the information available when and where it is needed. Since it is located in many diverse systems that are not integrated properly, it is almost impossible to provide the necessary information to those who need it. The plethora of information systems also raises the issue of interoperability and how the integration should be implemented technically. For example, should there be a tailored or a common interface for every information system?

The organizations are not completely satisfied with their current integration solutions. For example, the most common solution, FTP file transfer, is criticized because of the offline nature of the transfer. In addition, interviewees are unhappy with the uncertainty and difficulties of file transfers between information systems. For example, TS has experienced trouble with accuracy of data when data is moved from one information system to another. At WEST the difficulty is getting up to more elegant and automatic solutions from the very simple and tedious integration mechanisms like email attachment files. In STAT the problem is that the data is stored in various systems and it is currently very expensive and time-consuming to bring it all together. After all, the integration difficulties in the case organizations related generally to the challenges of implementing solutions and the fact that the level of integration should be raised. One of the interviewees summed up the dilemma: "The more data is coming in to the information system, the more problems there will be if the data is processed manually".

5.7.6 Reporting

The sixth category, 'reporting', represents problems like insufficient reporting tools and difficulties in producing reports. Table 5.8 shows that reporting is the most common problem in the case organizations: only in Statistics Finland is the IS-based reporting not considered to be a problem. The case organizations had discussed the idea of a data warehouse, but so far none had been adopted, although one organization was at a pilot stage. The rationale for these discussions of data warehouse solutions is that at the moment the reporting and the analysing possibilities are not serving the organizations very well. In fact, the interviewees identified many problems in reporting.

At SOK the interviewees complained that data is distributed in numerous information systems and thus it is difficult to get a comprehensive view of it. Another problem is in financial reporting. A financial report taken from different information systems gives different results, even though they should be equal. This may be because the data is not harmonized and processed at the same time and in the same way. In SOK's restaurant business product sales figures are essential. The company should be able to analyse what, where and how many products have been bought. At the moment it is impossible to analyse in detail different customers and their behaviour across the whole corporation. As the interviewees mentioned, a common database containing all of the co-operative society's products might help reporting, but defining a common classification of every product will be a demanding task. SOK managers have discussed centralizing their data to try and improve reports.

In Salon Seudun Puhelin Ltd the interviewees mentioned that users entering data somehow use the main information system inconsistently. Therefore the data is not consistent and this influences the reports. Salon Seudun Puhelin has developed it own reporting application with MS Access, but it is not capable of managing files over 1 GB, which reduces the possibilities of using the system. According to the interviewees this prevents the follow-up of daily sales. Another problem concerning the reporting system is that users are incapable of defining their own reports when their needs change. It is also difficult to analyse customer data since bringing all of this information together is a very burdensome task. Therefore, Salon Seudun Puhelin has also considered a data warehouse solution, a) to get rid of the size limits, b) to provide a system to the users where they can define new reports easily and c) to have more versatile analysis possibilities.

The State Provincial Office of Western Finland is a joint regional administrative authority of seven ministries. One of their annual responsibilities is to evaluate the basic services in their region. In practice this means that large amount of data is gathered and analysed. The first problem is that no special data management tool for managing this data has been used. Another issue is that the results should be easily distributed through various reports, but at the moment this is not the case. Also, the lack of adequate tools in data management makes it difficult to do any time-series analysis, which was anticipated in many interviews.

TS-Group has begun a data warehouse pilot, which enables versatile reporting and, as the interviewees mentioned, this opportunity should not be lost. However, some problems still exist with the reports. For example, the distribution and format of these reports has not yet been decided. In the department of financial management the reporting systems do not support the latest operating systems and therefore they can only be used on some computers. It should also be easy to modify these reports further in Excel, which sets some new requirements for the future development. Altogether, in TS-Group the reporting is fragmented and uncoordinated. Therefore it has begun to develop a new data warehouse.

At Optiroc reports are generated directly from the operational databases, even though they are not designed for reporting purposes. This means that they run slowly. In principle users should also be able to create reports by themselves, but in reality only a few of them have the necessary skills to do this. Maybe this is one reason why the interviewees were very critical of the reports. They also mentioned that the implementation of a data warehouse system was strongly supported and it was seen as a solution for these problems. A data warehouse was also justified since it also means that customers could be served better and customised reports could be produced.

At the moment customer reporting is analysed to help develop reports and to define necessary tools.

5.7.7 Personnel

The last category is ‘personnel’, which details the problems organizations can experience due to staffing changes. It also covers the difficulties associated with a limited amount of capable personnel. For example, at Optiroc the IS development is totally outsourced and only a couple of consultants are capable of maintaining the information systems. The company considers this to be a serious risk. SOK has had to replace some personnel and this resulted in some practical problems. In Salon Seudun Puhelin the main operating system is sometimes used incorrectly, since staff require additional training in using the system and entering data.

5.8 Plans for the future

When the interviewees were asked about future needs, plans and developments in their information systems and databases, two issues were considered especially important. They were integration and browser-based interfaces. Thus one of the main efforts in the organizations is to continue developing the integration of their information systems and databases in the future. Most future plans for the development of information systems and databases are quite easily derived from the problems described in previous section and they have been placed in seven categories (see Table 5.9).

Table 5.9: The development plans in the information systems and in the databases.

Case	Data modelling	DBMS upgrade/ replacement	IS upgrade/ replacement	Integration, Web services	Browser based interfaces	Hardware upgrade/ replacement	Strategy
SOK	X	X	X	X	X	X	X
SSP		X	X	X	X	X	X
STAT	X	X	X	X	X		
WEST	X	X	X	X	X		
TS		X	X	X	X	X	X
OPTI				X	X		X

SOK, Statistics Finland and the State Provincial Office of Western Finland will continue the development of data modelling. SOK will focus on building an enterprise data model, but the actual motivation is the future idea of an integrated information system environment. Statistics Finland is upgrading its file-based geographic information systems to database driven-systems. The State Provincial Office of Western Finland is also moving to database-driven information system and in both cases a reasonable starting point is to model the environment.

In all cases except Optiroc the upgrade or replacement of an information system will lead to an upgraded or new database management system. Actually this is a natural development given the average age of the information systems. Optiroc's information systems are newer and consequently there is no need to upgrade or replace them.

All organizations plan to continue their integration projects, but the focus is moving towards XML-based solutions such as web services. The aim is to use message-based integration and to get rid of unnecessary middle-phases and delays. For example, TS plans to move from EDI PAP to the XML-based PAPINET[12] solution. STAT is considering the integration alternatives and one central issue is an agreed format of the data. XML-based solutions are regarded as a strong candidate in this organization and database links are also on the agenda. In practice the movement to utilize XML means that the language must be produced from the database but it must also be processed and stored in the database. This might cause a problem when we weigh up the

[12] Look more information from www.papinet.org.

fact that the information systems and the databases are an average of seven years old. These old systems do not offer any support for managing XML, on the contrary special programs are needed. SOK is concentrating on integrating different business areas into a common centralized database solution where all customers and products could be managed in the future. For Salon Seudun Puhelin the customer is the focus of integration, with discussions centring on Business-to-Customer integration and the possibility of providing customers with access to their information systems through browsers. In the future, WEST is wanting to implement solutions for integrating many data sources in a sophisticated way. This should also serve customers better by offering them more accurate information. Customer relationship management is one example that is driving the organizations to integrate their information systems and databases. At TS the integration issues are currently linked to the implementation of a CRM (Customer Relationship Management) system and integration of various operational systems will be the focus in the near future. SOK, Salon Seudun Puhelin and Optiroc are discussing similar topics, along with the need to shape marketing. One important aspect in marketing today is the management of customer relationships. In Optiroc future integration plans are in B2B integration, which would provide online ordering possibilities to their dealers and allow them to receive sales estimates in the other direction.

There is also a clear trend towards browser-based user interfaces. Currently 37% of information systems have browser-based user interfaces and this figure is increasing.

SOK, Salon Seudun Puhelin and TS-Group have plans to upgrade or change their server hardware. They all rely on very old mainframe computers, purchased at around the same time as their information systems.

SOK, Salon Seudun Puhelin, TS-Group and Optiroc need to come up with a strategy. TS-Group has chosen to develop a CRM-strategy, while the others are aiming to update their IT-strategies.

5.9 Orientation values of the database innovations in each case organization

In this section each case organization is evaluated using the developed framework presented in section 3.4. The cases are presented in same order as they were carried out during the research.

5.9.1 SOK Corporation

Table 5.10 presents the evaluation of the organizational attributes of the SOK Corporation. SOK does not devolve very much power and control and therefore centralization is evaluated 'high'. The business groups of SOK Corporation have power to act within their own spheres, but again the degree of centralization within these divisions is high. Formalization is typically related to the centralization and it is considered to be moderate. The knowledge of database innovations is also moderate. Although some business groups do have a high level of understanding, across the whole organization it is only moderate. The nature of SOK requires different business groups to work together and create a common view for the future. An example of this joint working is this research, which aimed to produce an enterprise-level data model. Thereby the degree of interconnectedness in SOK is very high. SOK also has the necessary resources to adapt the database innovations, although it does not really have enough time to do this. The size of SOK Corporation is evaluated as very high; indeed it was the 14th biggest enterprise in Finland in 2003 (Talentum 2004).

The internal technological attributes show that IT management strength is high and IS portfolio complexity is moderate. The IT management of SOK follows the partner model and the strategic learning paradigm (see section 3.2 for explanations of these). The company is currently working on its new IS/IT strategy. As we can remember from section 3.3, the IS portfolio is made up of various attributes. The portfolio complexity of SOK is very high. There are a large variety of information systems that must be taken into consideration when new innovations are analysed. The present systems are seven years old on average and their volumes (the number of users, number of transactions, or size of the database) are quite high. However, the information that can be produced from these systems is sometimes lacking in quality. Also the quality of some of the systems themselves was criticized during the interviews.

SOK is operating in business areas that are quite stable; therefore volatility is 'low'. The system openness of SOK is high, as is its external technological environment. The external IT environment provides adequate resources to the organization. For example, most of the information system management and database administration is outsourced. On the other hand, changes in the external IT environment are not rapid and volatile. IT activity intensity in the external technological environment is high, since competitors, suppliers, customers and other stakeholders use IT intensively.

The evaluation of SOK's organizational attributes resulted in an orientation value of 64 percent for the adoption of database innovations.

Table 5.10: Organizational attributes in SOK Corporation.

I/E	Environment	Attribute	Evaluation	Points
Internal	Business	Centralization (-)	High	1
		Complexity	Moderate	2
		Formalization (-)	Moderate	2
		Interconnectedness	Very High	4
		Organizational slack	High	3
		Size	Very High	4
	Technological	IT management strength	High	3
		IS portfolio (-)	Moderate	2
External	Business	System openness	High	3
		Volatility of business	Low	1
	Technological	IS/IT environment	High	3
Sum of organizational attributes				28
Orientation value for the adoption of the database innovations				64 %

5.9.2 Salon Seudun Puhelin Ltd

Table 5.11 presents the evaluation of the organizational attributes of Salon Seudun Puhelin. In this company power and control is centralized only a little. Actually, different departments are quite autonomous and during the research it sometimes looked like they were operating too independently. Similarly, the degree of formalization is also considered to be low. The IT department is very small and resources for staff to study or inform themselves about database innovations are limited. This is not a question of money, but rather one of finding time. However, employees do have a basic knowledge of database-related topics. Therefore the complexity is low and the organizational slack very low. Salon Seudun Puhelin is the smallest of the cases, but this ensures a high level of interconnectedness, as people know each other in the organization.

The internal technological attributes show that both its IT management strength and IS portfolio are moderate. IT management is following the platform model and the experience paradigm. The research identified internal pressures to move towards the partner model and make the paradigm more strategic. Actually, they are currently working on an updated and more complete IS/IT strategy.

The IS portfolio complexity is evaluated as being low. There are only few critical information systems, thus indicating low portfolio. The average age of the information systems is six years and considered moderate. The size of the information systems is low. The quality of the information systems is high, but unfortunately the usage of these systems is not consistent and this has lowered the quality of the information.

Salon Seudun Puhelin operates in a business area where volatility is high. New products and services are developed constantly and the basic communications technology and infrastructure is evolving. The system openness of Salon Seudun Puhelin Ltd is moderate. The organization is part of the FINNET group, which facilitates links to members of other organizations. Also outsourcing IS management indicates system openness.

The external technological environment is very high according to the measurement framework. It provides adequate resources to the organization, although they are not true alternatives for the most critical information systems. On the other hand, changes in the external IT environment are not rapid and volatile. IT activity intensity in the external technological environment is very high: the whole business is based on technology.

The evaluation of the organizational attributes of Salon Seudun Puhelin Ltd resulted in an orientation value of 57 percent for the adoption of database innovations.

Table 5.11: Organizational attributes of Salon Seudun Puhelin Ltd.

I/E	**Environment**	**Attribute**	**Evaluation**	**Points**
Internal	Business	Centralization (-)	Low	3
		Complexity	Low	1
		Formalization (-)	Low	3
		Interconnectedness	High	3
		Organizational slack	Very Low	0
		Size	Low	1
	Technological	IT management strength	Moderate	2
		IS portfolio (-)	Low	3
External	Business	System openness	Moderate	2
		Volatility of business	High	3
	Technological	IS/IT environment	Very High	4
Sum of organizational attributes Orientation value for the adoption of the database innovations				25 57 %

5.9.3 Statistics Finland

Table 5.12 presents the evaluation of the organizational attributes of Statistics Finland. In this organization power and control is moderately centralized. Public organizations are typically bureaucratic and formal. Statistics Finland is not an exception and thus the formalization is deemed high. It develops many information systems in-house and has knowledge of database innovations. Staff have also worked on improving their expertise in this field. Thus the complexity of Statistics Finland is very high. At the moment they are working on a large project to reorganize the production of statistics and data management issues play a role in this. Therefore they have necessary resources available and the organizational slack is very high. However, the degree of interconnectedness is moderate. The different departments work mainly independently from each other, but also inter-departmental work groups are recognized. The size of Statistics Finland is high.

The internal technological attributes show that IT management strength is moderate. IT management also follows the platform model and the experience paradigm; they work pragmatically and are result-oriented. The IS portfolio is moderate. The portfolio complexity is high, because there are numerous different statistics and classifications that need to be supported by corresponding information systems. The average age of the studied information systems in Statistics Finland is ten years and therefore high. The size of the information systems is rated low. The quality of the information systems is moderate, because their data management capabilities are not as versatile as expected. Similarly the quality of the information is low, because the information is partly erroneous and inaccurate, since all features of the original data are not accessible yet.

Statistics Finland operates in a business area that is very stable. The same statistics are produced yearly and only few minor changes are expected per year. The system openness is moderate, with nearest partners in other Nordic countries. The organisation also operates at a European level to design commensurable statistics.

The rating for this organization’s external technological environment is low. It may provide standard tools for statistical analysis, but all special analysis problems are solved with information systems developed in-house. The external IT environment is quite stable from Statistics Finland’s point of view.

The evaluation of the organizational attributes of Statistics Finland resulted to an orientation value of 52 percent for the adoption of the database innovations (see Table 5.12).

Table 5.12: Organizational attributes of Statistics Finland.

I/E	Environment	Attribute	Evaluation	Points
Internal	Business	Centralization (-)	Moderate	2
		Complexity	Very High	4
		Formalization (-)	High	1
		Interconnectedness	Moderate	2
		Organizational slack	Very High	4
		Size	High	3
	Technological	IT management strength	Moderate	2
		IS portfolio	Moderate	2
External	Business	System openness	Moderate	2
		Volatility of business	Very Low	0
	Technological	IS/IT environment	Low	1
Sum of organizational attributes				23
Orientation value for the adoption of the database innovations				52 %

5.9.4 The State Provincial Office of Western Finland

Table 5.13 presents the evaluation of the State Provincial Office of Western Finland's organizational attributes. In this organization the power and control is moderately centralized, since departments each have their own focus. Again, as a public organization, the degree of formalization is high. Interconnectedness is considered moderate: the different departments work mainly independently from each other, but inter-departmental work groups are also recognized. The State Provincial Office of Western Finland does not have knowledge of database innovations. All of the maintenance tasks are also outsourced. Therefore the evaluation for complexity is very low. Neither are there any resources available for analysing and working with the database innovations. The organizational slack is also very low. Size of the organization is moderate.

The internal technological attributes show that IT management strength is moderate. IT management follows the platform model and the experience paradigm; it is working pragmatically and is result-oriented. The IS portfolio is low, since they have only a few essential information systems. The average age of the studied information systems is nine years and thus evaluated high. The size of the information systems is low. The quality of the information systems is also low, because there is a lack of proper information systems. The quality of the information is high, but the collection of the raw data requires plenty of manual work.

The State Provincial Office of Western Finland operates in a business area that is very stable. Their main responsibilities are defined in the law. The system openness is considered to be very high. Members of the organization are constantly in contact with individuals of the external business environment when they gather information from municipalities and other actors.

The external technological environment is rated ‘low’. There are only five provinces in Finland and their data processing needs are not typical. There are no information system solutions tailored to these needs available for implementation in the external IT environment. Rather, all operational information systems must be customized. IT dynamism of the external IT environment is complex but stable. Major national IT projects that would also ease operations in the State Provincial Office of Western Finland are in the pipeline, but it will take time before their benefits come on stream. The density of IT activity in the external technological environment is high. The municipalities are well computerized, although these solutions do not help the State Provincial Office of Western Finland.

The evaluation of the organizational attributes of State Provincial Office of Western Finland resulted in an orientation value of 39 percent for the adoption of database innovations.

Table 5.13: Organizational attributes of the State Provincial Office of Western Finland.

I/E	Environment	Attribute	Evaluation	Points
Internal	Business	Centralization (-)	Moderate	2
		Complexity	Very Low	0
		Formalization (-)	High	1
		Interconnectedness	Moderate	2
		Organizational slack	Very Low	0
		Size	Moderate	2
	Technological	IT management strength	Moderate	2
		IS portfolio (-)	Low	3
External	Business	System openness	Very High	4
		Volatility of business	Very Low	0
	Technological	IS/IT environment	Low	1
Sum of organizational attributes Orientation value for the adoption of the database innovations				17 39 %

5.9.5 TS-Group Ltd

Table 5.14 presents the evaluation of TS-Group Ltd’s organizational attributes. TS-Group Ltd is a family-owned business and power and control is very highly centralized in the family. The organization is however not highly bureaucratic and therefore the degree of formalization is only moderate. It developed two of the main information systems in-house and has a very high knowledge of database innovations. Staff have also worked on improving their expertise in this field. The complexity of TS-Group Ltd is very high. They also have quite good resources in their IT/IS department, thus indicating high organizational slack. This is shown by the CRM project currently being implemented. This implementation has required and will require a lot of integration and customisation throughout the organization. The employees have worked for this common goal and this has required a number of discussions and meetings. Therefore the interconnectedness of TS-Group Ltd can be said to be very high. The size of the TS-Group Ltd is high; in 2003 it was the 136th biggest company in Finland (Talentum 2004).

The internal technological attributes show that IT management strength is high. IT management follows the partner model and the strategic learning paradigm. The development and maintenance of information systems are emphasized, but so are strategic issues. For example, the organization is looking to write a CRM-strategy. The IS portfolio is high and it is also highly complex. The information systems are integrated and the level of integration has risen due to the CRM-implementation. The size of the information systems is high. The age of the information systems is moderate, with an average of seven years. The quality of the information systems was not covered during the interviews; respondents mentioned only some minor issues. From this we can ascertain that they were not problematic and systems' quality is therefore high. Also the quality of the information is high, but as the CRM project shows there are still possibilities to improve it further.

TS-Group Ltd operates in a business area that is quite unstable and highly competitive. The CRM project is an example of how it is trying to gain competitive advantage. The system openness is considered to be high. There is plenty of co-operation with other publishers and the TS employees are constantly in contact with individuals in the external business environment.

On the scale outlined in section 3.4, the external technological environment is ranked as being high. It can provide solutions for TS-Group Ltd as it can do it for other organizations in the same line of business. Dynamism in the external IT environment is high. There are for example initiatives to develop data exchange formats (see www.papinet.org). The whole communications business exploits technology, therefore we can conclude that the IT activity density of the external technological environment is high.

The evaluation of the organizational attributes of TS-Group Ltd resulted in an orientation value of 66 percent for the adoption of database innovations.

Table 5.14: Organizational attributes of TS-Group Ltd.

I/E	Environment	Attribute	Evaluation	Points
Internal	Business	Centralization (-)	Very High	0
		Complexity	Very High	4
		Formalization (-)	Moderate	2
		Interconnectedness	Very High	4
		Organizational slack	High	3
		Size	High	3
	Technological	IT management strength	High	3
		IS portfolio	High	1
External	Business	System openness	High	3
		Volatility of business	High	3
	Technological	IS/IT environment	High	3
Sum of organizational attributes Orientation value for the adoption of the database innovations				29 66 %

5.9.6 Optiroc Ltd

Table 5.15 presents the evaluation of the organizational attributes of Optiroc Ltd. The centralization of power and control is moderate. The organization is not bureaucratic and the degree of formalization is low. Optiroc has outsourced IS development and does not have knowledge of database innovations. Expertise is bought in from the external IS/IT environment. Therefore the complexity can be considered very low. The organization does not have human resources available, but must have monetary resources available. Therefore the organizational slack is moderate. Its interconnectedness is also moderate, as is its size, being the 227th biggest company in Finland in 2003 (Talentum 2004).

The internal technological attributes show that IT management strength is moderate. IT management follows the scalable model and the experimental learning and political paradigm (see section 3.2). The IS/IT department uses sourcing to achieve flexibility in resources, but it is also emphasizing the maintenance and development of its information systems. On the other hand, little emphasis is put on the IS-strategy, something that is recognized by the interviewees. They are however finally starting to formulate an IS strategy.

The IS portfolio is considered to be ‘high’. The information systems are highly customized, making the portfolio complexity very high. The size of the information systems is high. The age of the information systems is low, as they are on average just four and half years old. The quality of the information systems is high. Although criticisms were aired in the interviews, these comments were directed at the skills required to use the systems rather than the information systems themselves. The quality of the information is high, but there is a need to develop reporting to produce even better and more accurate information.

Optiroc Ltd operates in a business area that has low volatility. The system openness is considered moderate, as is the external technological environment. The external IT environment can provide off-the-shelf solutions for the organization, but it has decided to use customised information systems. This lowers the evaluation of the external IT environment, because only a few people are capable of maintaining the information systems. Dynamism in the external IT environment is rated moderate, but its activity density is high as there are numerous different information systems used by suppliers and customers. It is a real challenge to develop integration solutions suitable for all.

The evaluation of the organizational attributes of Optiroc Ltd resulted in an orientation value of 43 percent for the adoption of database innovations.

Table 5.15: Organizational attributes of Optiroc Ltd.

I/E	**Environment**	**Attribute**	**Evaluation**	**Points**
Internal	Business	Centralization (-)	Moderate	2
		Complexity	Very Low	0
		Formalization (-)	Low	3
		Interconnectedness	Moderate	2
		Organizational slack	Moderate	2
		Size	Moderate	2
	Technological	IT management strength	Moderate	2
		IS portfolio (-)	High	1
External	Business	System openness	Moderate	2
		Volatility of business	Low	1
	Technological	IS/IT environment	Moderate	2
Sum of organizational attributes Orientation value for the adoption of the database innovations				19 43 %

6 CONCLUSIONS

The conclusions are presented in two parts. The first section concentrates on answering the empirical research questions and the second looks at the theoretical part of the study.

6.1 Diffusion of database innovations in the cases

This section describes the empirical conclusions.

6.1.1 RDBMS in front, but XML is entering into databases

This study has shown that organizations place a lot of trust in using relational database management systems as the data storage solution of their information systems. The database innovations described in earlier sections of this thesis have not been exploited in any of the studied organizational databases. Thus this research does not confirm the ideas of Brown et al. (1999) when they estimated that ORDBMSs would be 50% bigger than RDBMSs by 2005.

The information systems are also quite old. However, both these and the database management systems function at least at a satisfactory level. Nevertheless, they also lack properties that will become increasingly important in the future. For example, they do not support the storage of complex objects, not to mention natural management of XML format data. Part of the new requirements can be managed with the existing systems, but this requires much more work than with newer DBMSs. Many organizations have recognized that old information systems and DBMSs are a barrier to strategic innovation (Gibson et al. 1999). Newer DBMSs also offer better development tools, thus lowering the costs of development tasks. The interviews show that in a couple of years the information systems and databases studied in this research will be in a situation where they may need to be replaced. Altogether, we may say that the diffusion of database innovations (object-relational, XML-enabled and spatial databases) has not truly started in the organizations. The reasons for this are discussed in section 6.2.2.

Figure 6.1 describes the diffusion of DBMS technology. The curves *RDBMS, ORDBMS* and *XML enabled DBMS* describe the theoretical diffusion of each DBMS technology. These curves are based on the diffusion of innovations theory and discussion in the research community. The Present-curve describes how the current relational DBMS-based information systems have been implemented in the case organizations. The data for this curve comes from the databases that were studied as part of this project.

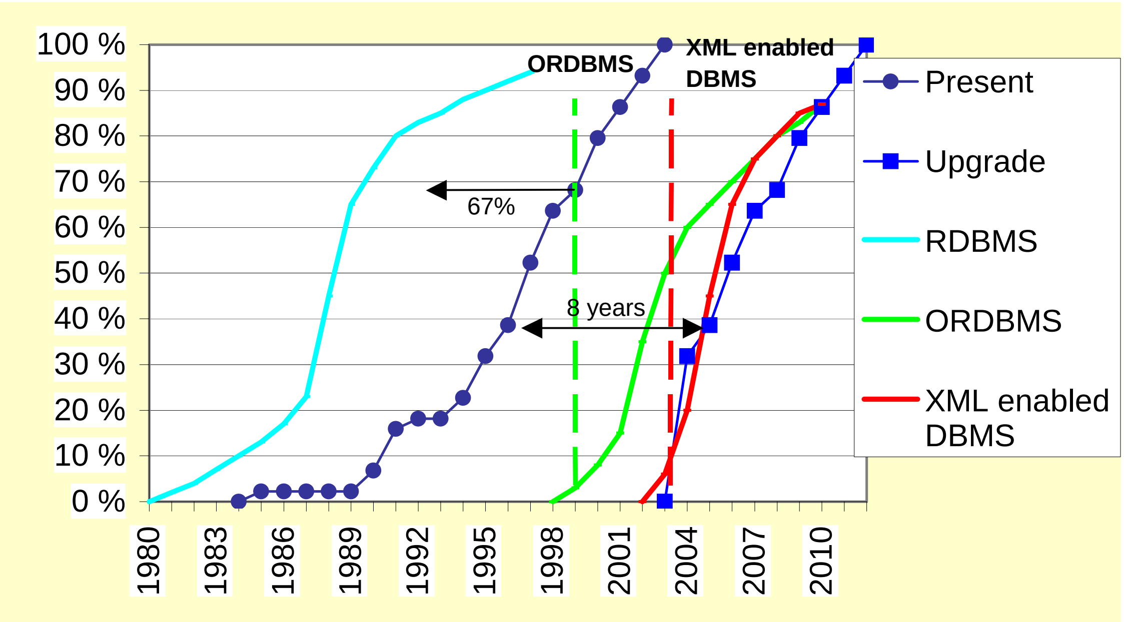


Figure 6.1: Diffusion of DBMS technology.

The Figure reminds us that 67% of the information systems were implemented before the introduction of the object-relational databases in 1999. It also shows that all of the information systems and databases were implemented before relational-based XML-enabled database management systems with native XML storage possibilities were introduced. The Upgrade-curve plots the diffusion of current information systems eight years after their actual year of implementation. As we can see, 100% diffusion only occurs eight years after the actual year of implementation. Eight and half years have been recognized as the average age when information systems are upgraded or replaced (Krogstie et al. 1994). The interviews suggested a similar upgrade/replacement cycle. A recent survey indicates an accelerating rate of upgrade/replacement, since 41 percent[13] of respondents defined upgrading or replacing legacy systems as a very critical IT issue in 2002 (CSC 2002b). However, many of the oldest information systems and databases have still not been upgraded/replaced and they had already been in place for more than eight

[13] Indicating a 10 percent growth since 2000.

years by 2003, i.e. more than eight years after the actual year of implementation. This is the reason why the Upgrade-curve is steeper than the Present-curve at the beginning.

In another study the average age of information systems was seven years and average life expectancy was a little under five years (Swanson et al. 2000). Thus combining the results of this dissertation and the results from other researches we can make a conclusion that information systems are seven years old on average and are typically upgraded/replaced when they are eight and half, but can be used until twelve years after the implementation date (see Figure 6.2).

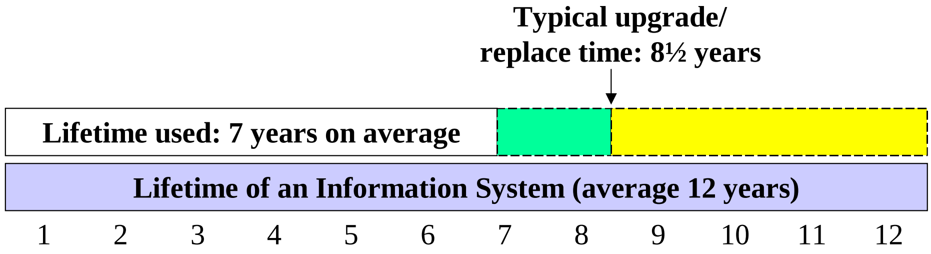


Figure 6.2: Lifetime of information systems.

The upgrade curve in Figure 6.1 follows the shape and timing of the theoretical curve of XML-enabled DBMS almost exactly. In practice this means that all of the information systems and databases are updated or replaced after the database innovations presented in this research are introduced. It also means that the theory of the diffusion of innovations is supported by practice in the case organizations. Thus, it seems that a number of new information systems and databases will be capable of XML data management. The object-relational DBMS were not widely accepted when they were introduced, but these new upgrades will probably mean that they can incorporate these features. We can say that organizations are entering into the confirmation stage of the innovation-decision process. Actually, new XML enabled database management systems that have developed from a traditional relational DBMS have gone through the evolutionary phase of object-relational DBMS and contain these features automatically. Finally, we must remember that the benefits of XML will only be delivered when there are several XML-capable information systems. The chart is suggesting that more than half of all information systems will have XML capabilities by 2006.

Figure 6.1 shows that 33% of information systems and databases were implemented after object-relational DBMSs were introduced. However, ORDBMSs were not adapted for these systems. According to Rogers (1995) there are two kinds of information associated with a technological innovation: software information and innovation-evaluation information. Software information concentrates on technology and reduces uncertainty of the technology itself, while innovation-evaluation information reduces uncertainty surrounding the expected consequences of the innovation. Considering the extent of diffusion of object-relational databases, the software information was probably clear but the innovation-evaluation information was not. The idea of object-relational databases was understood, but since real implementations were rare, little innovation-evaluation information was available.

This research estimates that database innovations will diffuse in the information systems and databases of the organizations in the near future. The view is that new information systems and database solutions will be based on the enriched third generation database management systems, which manage the whole coverage of data from numbers to objects to XML documents (see Figure 6.3).

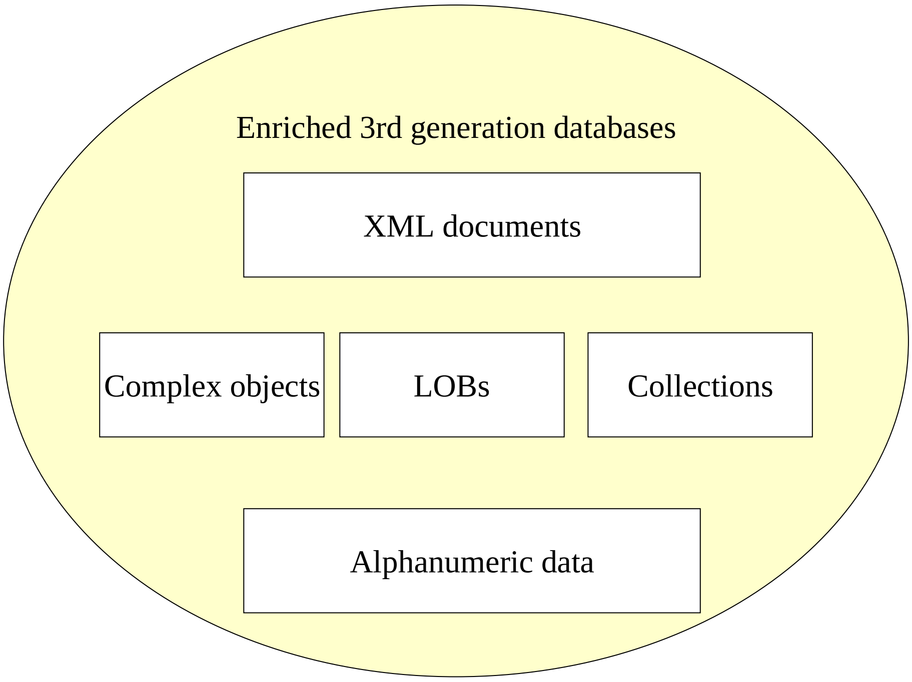


Figure 6.3: Data managed with enriched 3rd generation databases.

Object structures are exploited in managing XML documents, but object structures are becoming more important also because the rising exploitation of spatial data: spatial data structures offered by new databases use typically object structures. So far spatial data has been managed in different ways without object structures. Actually, the organizations are not familiar with software information relating to spatial extensions in database management systems. This research proved however that spatial data management has the potential to play a major role in the future information systems of the case organizations. Roddick et al (2000) reached a similar conclusion. Examples of spatial data management already exist in the case organizations and it is beginning to feature in databases due to the upgrades and replacements of information systems.

Diffusion is easier with XML-enabled database management systems than object-relational DBMSs, because the benefits (innovation-evaluation information) have already been recognized in integration efforts. On the other hand, the software information of XML-enabled DBMS is now trickier to extract because there has not been much time to test them. This is true especially with information systems that are developed within the organizations (25%) and not by an IS/IT company. In the latter circumstance, the information systems are more or less dependent on the IS/IT companies and their decisions on XML usage. However, the case organizations have close communication channels with these companies and can affect the solutions made there at some level. Thus the estimation is that relational database management systems will change to enriched third generation database management systems within the next few years and XML, spatial data and objects will be managed fluently in the database management systems. The information systems are finally turning into universal applications, as Brown et al. (1999) predicted. In practice this means that organizations upgrade or replace their relational databases to XML-enabled DBMSs that are capable of managing not only traditional relational data and XML documents, but also other information, such as spatial data. This means also that the organizations need to improve their knowledge of these topics, as the analysis in section 5.9 confirmed (see values for the *Complexity* attribute). For example, staff should be trained in XML.

The enriched third generation DBMSs can also improve performance. Organizations operating in the Internet typically produce information for the web from operational data. The DBMS could be utilized to produce the content as XML documents in the database. When users access this data through the Internet they receive the nicely formatted XML documents. This will improve the website's performance, especially if a large number of users are accessing the data simultaneously. If the XML documents have to be created from the operational data every time a user accesses some information, this will require additional, non-operational resources. The DBMS acts here as a XML document cache. TS-Group could benefit from this kind of structure. The content of their newspaper is turned into XML documents and readers accessing their web version of the newspaper receive all of the relevant information from the XML document cache. Another example is Amazon, which provides a lot of information about their products through their web site. This might be generated directly from the operational data or it might have been generated earlier into XML documents in a database. If the latter option is adopted, the database becomes a kind of XML document cache, which provides users with easy and quick access to the data.

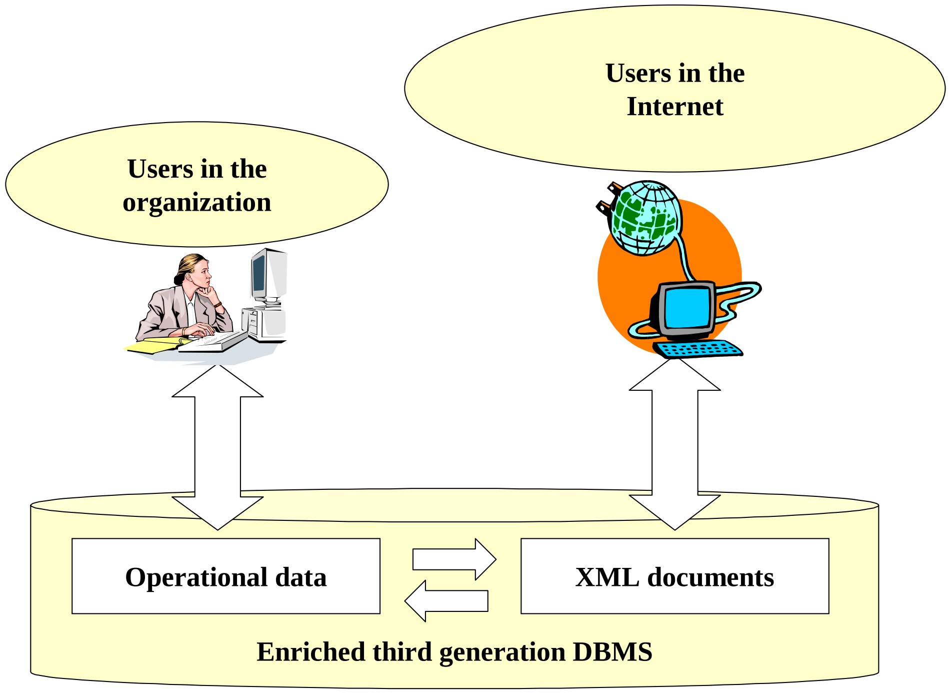


Figure 6.4: XML document cache.

If we place the organizations in Rogers' (1995) Innovation-Decision process, we end up with the situation shown in Figure 6.5. It demonstrates that organizations are convinced of the capabilities of the relational DBMS, but have stopped in the decision stage with the object-relational DBMS and thus far failed to use it. With a spatial DBMS, an appliance of the ORDBMS, the organizations are mainly at the *Knowledge* stage (except Statistics Finland, which is entering into the *Decision* phase). Organizations have already gained some knowledge of the XML database management systems and attitudes are becoming favourable, but additional understanding is still needed.

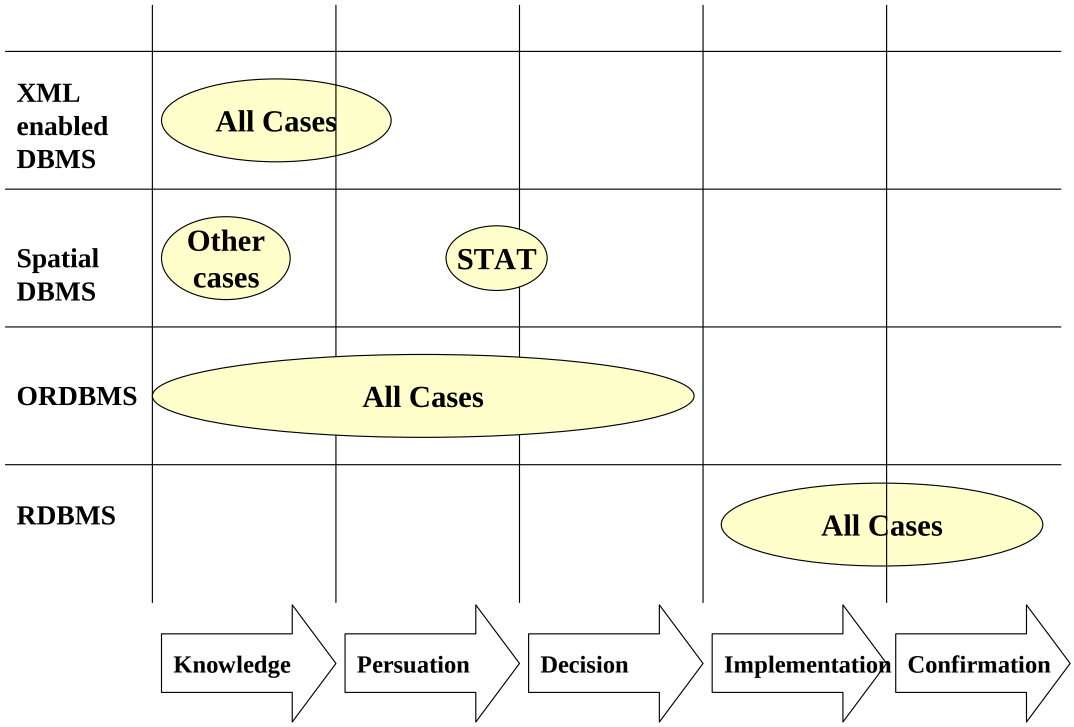


Figure 6.5: The organization in the Innovation-Decision process of database innovations.

6.1.2 Integration is emphasized

This research shows that organizations are emphasizing the importance of integrating their information systems and databases. Similar results have been achieved in other studies as well. For example, in (CSC et al. 2003) respondents indicated a definite need for greater integration to improve enterprise performance. In this research lack of integration is identified as one of the main problems with current information systems and databases, but it is also identified as one main development area in the forthcoming years. Integration issues influence and emphasize the need for upgrading and replacing information systems and databases. Indeed, integration requirements are a common reason for renewing information systems (Krogstie et al. 1994). Also the financial benefits of computing depend on the success of creating single information systems (Sutherland et al. 2002).

Technically, the integration can be performed in many ways, as is shown by the current integration solutions in the case organizations. In this study the organizations emphasized the role of XML in future integration efforts, although other alternatives also exist. In practice this accentuated role for XML means that the language has to be produced and managed in the information systems and databases. However, for the case organizations these plans to exploit XML will mean major changes to their existing systems. Current information systems and databases do not support XML natively. The possible change to XML-enabled database management systems will also entail a number of costs: hardware, software and personnel. This may include spending money on:

- Updated/upgraded information systems and databases require powerful new servers.
- Costs may vary, depending on the degree to which the present software can be upgraded.
- Personnel must be trained to use XML-enabled solutions. A couple of cases have already started to do this.

These integration issues confirm the ideas presented in the previous section, namely that XML is entering into the databases and XML-capable databases will diffuse in the organizations in the coming years. However, we must remember that a single information system managing XML fluently will not create an integrated environment of information systems. The benefits of XML will only come on stream when there are several XML-capable information systems in the organization (Lear 1999). A very similar idea is presented in diffusion of innovations theory by using the concept of 'critical mass'. Critical mass is defined as *the point when the rate of the adoption of the innovation is at a level where further rate of adoption becomes self-sustaining*. (Rogers 2003).

Integration of information systems and databases is also a prerequisite to many areas of computing (see Figure 6.6). Designing an enterprise-level data model is a logical starting point for integration. This model acts as a guideline for integrating information systems and databases and it defines the basics of a possible integrated database. Once an integrated IS and database environment is created, new possibilities become available. For example, customer relationship management requires that all customer-related data is accessible from one place when it is needed. Without an integrated solution the gathering of customer-related data would require the use of multiple information systems. Combining the data would take a large amount of time and work. The requirements for data warehousing are similar, but the focus is probably not on the customer. Again, an integrated environment supports the aims of data warehousing, in which the operational data is integrated into a data warehouse structure. If integration of operational data sources has already been achieved, it is much easier to design the data warehouse solution. An integrated environment also serves the requirements of electronic business, where information is needed from many sources. The share of online transactions in total business-to-business trade is expected to grow rapidly as organizations integrate their information systems (United Nations 2003). E-Business and business process management are increasing the integration efforts within organizations (Fremantle et al. 2002). Integration also improves the possibilities to use special features associated with some data effectively throughout the organization. For example, spatial data and spatial analysis can be combined in many different areas of the business, which again can provide new value for the organization.

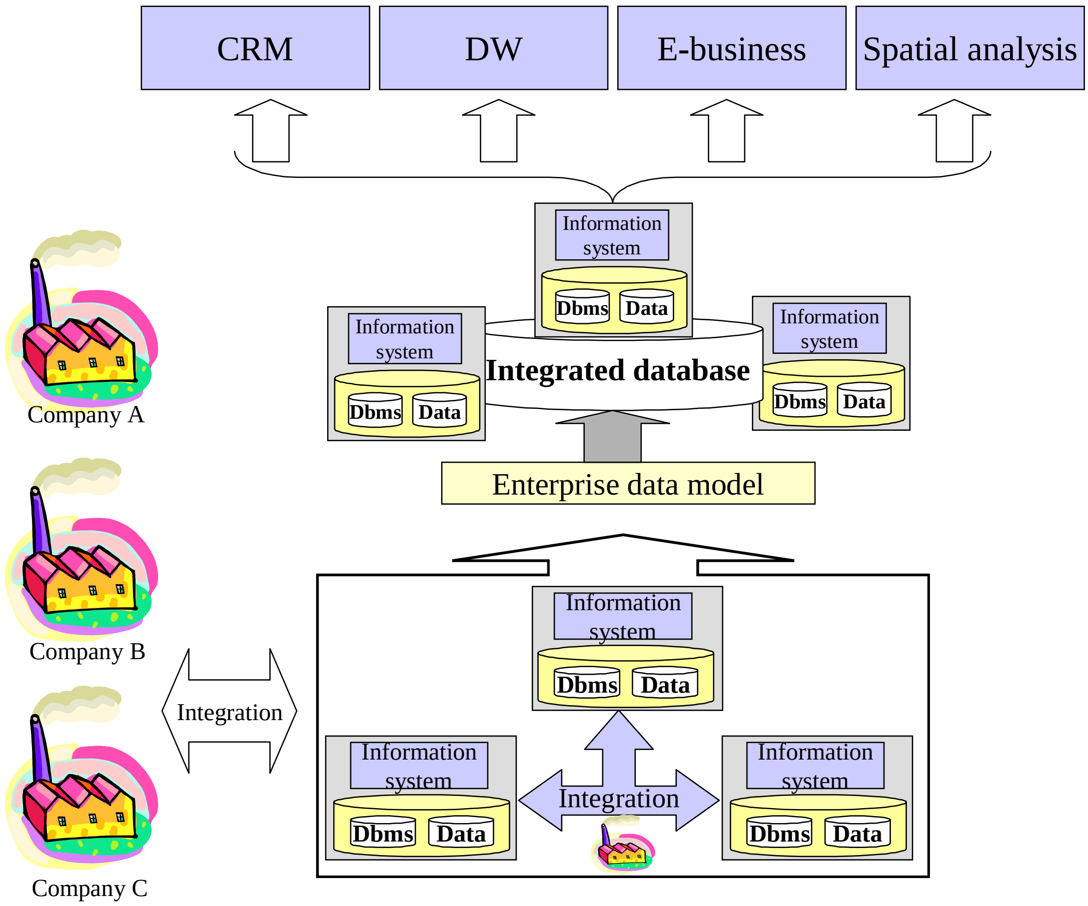


Figure 6.6: Integration opens new possibilities.

6.1.3 Other observations

There are also other issues that support the conclusion that database innovations will diffuse during the next few years. This section presents these additional empirical observations and conclusions, which relate to reporting, browser-based user interfaces and documentation.

6.1.3.1 Reporting needs to be improved

This research showed that the organizations have problems with reporting. Currently, they do not base this on data warehouses, but rather on some special applications that do not answer to the reporting needs adequately. Similar problems have been identified elsewhere as well, since “facilitating analysis and decision making” and “creating an effective environment for sharing relevant information” were ranked as the top two issues that were most constrained by a lack of appropriate technology solutions (CSC et al. 2003). These problems are also associated with problems in business intelligence and knowledge management (CSC et al. 2003). Also, the case organizations hope to make reports available on the web. XML is one natural alternative for producing online reports. Producing XML from the database brings us back to the point where the service of the database management system needs to be analysed.

In future the SOK Corporation needs to build a comprehensive view of its information. Building an enterprise data model and then modifying the existing information systems can achieve this. From the reporting point of view the operational data should be further modelled into a data warehouse solution. Taking these steps should solve the reporting problems.

At Salon Seudun Puhelin Ltd most effort needs to be made to deal with data accuracy and the ways users work with the information systems to improve data consistency. Before this is achieved it is not rational to start exploiting a data warehouse. The company should evaluate the true requirements of establishing a data warehouse for reporting, since other easier solutions might be available.

In the State Provincial Office of Western Finland (WEST) the most important issue is to get a suitable tool for managing the collected data. After that the emphasis can be on reporting, analysing and comparing data in different time-series. Basically, the environment of WEST is ideal for exploiting a data warehouse. The processes and the functions are heavily dependent on the analysis and developments in time-series.

In addition, a data warehouse could solve TS-Group’s reporting problems. This would standardize report formats, something that is currently a problem. It would also solve the report distribution problems at the same time.

In Optiroc Ltd the time it takes to run reports and the difficulties in analysis can also be solved with a data warehouse. One reason for the slowness was that the reports were run directly from the operational databases and transferring the source of the data to a data warehouse might offer improvements. Another problem concerned analysis and whether reports could be generated on an ad hoc basis. Introducing necessary OLAP tools and training the users sufficiently will solve this problem.

At Statistics Finland the interviewees did not mention reporting as a problem. However a data warehouse might still offer extra value in data analysis. At the moment though, data warehousing is not the most acute topic in IT-development in this organization.

6.1.3.2 Heading for browser-based interfaces

Organizations are emphasising browser-based user interfaces as the future way to access information systems and databases. This also touches on XML and the combination of XML and databases. XML needs not only to be produced from the database but also to be stored in the database. Both these issues are confirming the diffusion of the database innovations that can manage XML fluently.

6.1.3.3 We can live with the documentation

The documentation in the case organizations proved to be in a better overall condition than some recent studies (Cook et al. 2002; Briand 2003) might suggest. However, there were also numerous information systems where the documentation needs to be improved. In fact, in 66% of the studied information systems the documentation was not satisfactory and had shortcomings. The interviewees felt that the information systems developed by the organization itself had better documentation than those that had been outsourced. The interviewees also stated quite often that *the documentation is outdated, but we can still manage with it*. This confirms the ideas of Forward et al. (2002), who argued that documentation can be relevant even if it is not up to date.

Only two of the cases had developed and used their own standards and templates to assist the documentation process. However, Cook et al. (2002) wrote that the importance and fundamental role of the key practices is very clear. This research validates the importance and meaning of those (recall case STAT), but it also shows that these guidelines must be updated to correspond the current technological environment (recall case TS).

6.1.4 Observations of the organizational attributes

The orientation value for the adoption of the database innovations varied from 39 to 66 percent (see Table 6.1). As these values indicate, these organizations are not particularly fruitful ground for database innovations – indeed, at the moment they are not very interested in them. If we look at the single attributes of the organizational part of the developed framework (see Table 6.2) we see that interconnectedness had the highest average (2,83) and volatility of business had the lowest score (1,33).

Table 6.1: Ranking of organizations based on orientation value.

Rank	**Case**	**Orientation value**
1	TS-Group Ltd	66 %
2	SOK Corporation	64 %
3	Salon Seudun Puhelin Ltd	57 %
4	Statistics Finland	52 %
5	Optiroc Ltd	43 %
6	State Provincial Office of Western Finland	39 %
	Average	53 %

Table 6.2: Organizational values and attribute averages.

Attribute	**SOK**	**SSP**	**STAT**	**WEST**	**TS**	**OPTI**	**Avg**
Centralization (-)	1	3	2	2	0	2	1,67
Formalization (-)	2	3	1	1	2	3	2,00
Interconnectedness	4	3	2	2	4	2	2,83
IS Portfolio (-)	2	3	2	3	1	1	2,00
Complexity	2	1	4	0	4	0	1,83
Organizational Slack	3	0	4	0	3	2	2,00
Size	4	1	3	2	3	2	2,50
IT Management Strength	3	2	2	2	3	2	2,33
System Openness	3	2	2	4	3	2	2,67
Volatility of Business	1	3	0	0	3	1	1,33
IS/IT Environment	3	4	1	1	3	2	2,33
SUM	**28**	**25**	**23**	**17**	**29**	**19**	

The current CRM-project is one reason for the higher orientation value of TS-Group Ltd. The organization is quite prepared for the change and this also raises its score for other attributes, such as interconnectedness. The complexity value is noteworthy as well. TS-Group Ltd has had a lot of experience in developing and integrating information systems and, together with in-house training in new topics like XML, this contributes towards the very high score. However, the IS portfolio is in good condition which reduces the need for adapting the database innovations.

Because of its size, SOK has a good starting point for adopting database innovations. The size enables more possibilities in using resources thus giving them flexibility in organizational slack: they can use resources flexibly and in different ways. Following the partner model and strategic learning paradigm in IT management is assisting innovation in organization. As a co-operative society, SOK is confronted with a highly complex IS Portfolio, but this is recognized by IT Management and development plans already exist to simplify it.

The smallest organization covered in this research, i.e. Salon Seudun Puhelin Ltd, received the third highest orientation value. The main rationale behind this is its external attributes. The business and the external IS/IT environments pressure Salon Seudun Puhelin Ltd to act and innovate according to their market instincts. As a smaller organization the IS portfolio is also easier to deal with and there are lower barriers to adopting new technology. On the other hand, it is clear that resources are scarce.

Like TS-Group Ltd, Statistics Finland is in the middle of a large development project. As a result, IS-related issues are well resourced at the moment. As the project also relates to data management issues, the necessary employees have gained expertise in the field of database innovations, especially in spatial databases and XML. The business environment in which Statistics Finland operates is a distinguishing feature. Statistics Finland is practically the only actor in this specific field in the country and therefore the business environment does not stimulate the diffusion process.

Optiroc Ltd is an interesting case in this research, because it is the only organization to follow the scalable model in organizing IT management. It seems that the model has worked out well, since the IS portfolio of Optiroc Ltd is in good condition, even though their knowledge of database innovations is very low. At the moment the organization is emphasizing maintenance operations of the existing information systems, meaning that the degree of innovativeness is low.

The State Provincial Office of Western Finland is currently on a waiting list for database surgery. It has identified the shortcomings of its information systems, but at the moment there are neither the resources nor the expertise to begin the diffusion process. It seems also that the organization is somehow tied to the bureaucratic way of doing things. Nor does its business environment initiate any innovation processes. However, this organization is probably the one in which database innovations could be most widely utilized and even without any burden of the old non-existing information systems and their upgrades.

6.2 Theoretical conclusions

This section outlines the theoretical conclusions of the research. It evaluates database innovations and their consequences according to theoretical frameworks. It then identifies critical organizational success factors and presents a modified framework, before describing, an innovation-decision process in an organization.

6.2.1 Evaluation of the database innovations

If we want to understand why and how an innovation spreads and which of its characteristics lead to widespread acceptance, we need to look at its advantages and disadvantages. This section deals with these issues by considering database innovations. We evaluated the attributes of database innovations in section 4.5 (see Table 6.3) and these findings are also presented in Figure 6.7 as a chart.

Table 6.3: Evaluation of the attributes of the database innovations.

	ORDBMS	**Spatial DBMS**	**XML enabled DBMS**
Relative advantage	Moderate	Moderate	Very high
Compatibility	Very high	Very high	Very high
Complexity (-)	Low	High	High
Trialability	High	Low	Moderate
Observability	Very low	Low	Moderate

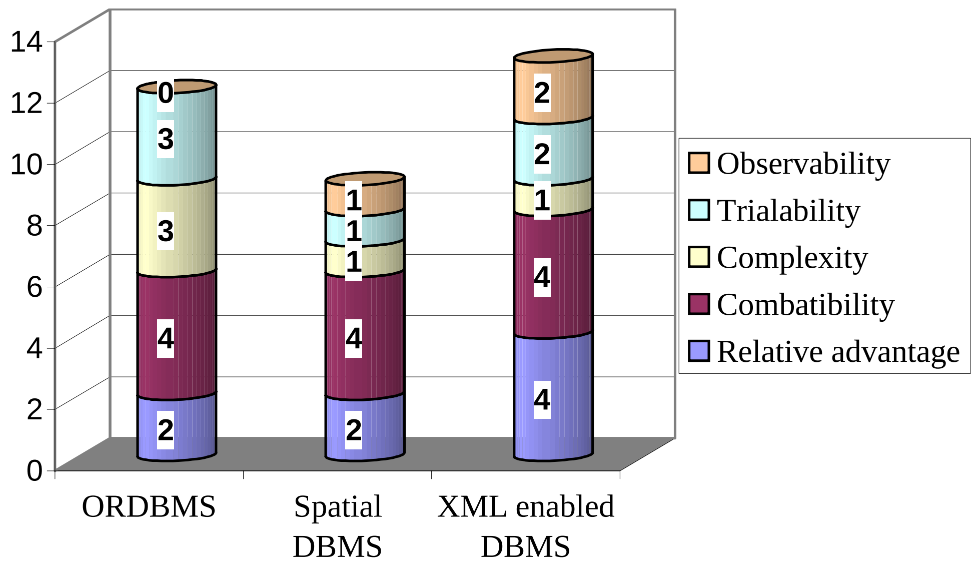


Figure 6.7: Chart of the database innovations' evaluations.

The **relative advantage** of database innovations over current relational databases seems obvious. However, we must also restate the common fact that relational database management systems are perfect if only alphanumeric data is managed and the integration needs are limited. They are also a very good choice when performance and reliability are important. Of course, this does not mean that database innovations cannot offer performance and reliability.

The basic advantage database innovations offer is the wider variety of different information that those database management systems can manage. These new supported types of data are user-defined types, large objects, collections and XML documents. User-defined types make it possible to define totally new types that represent the world closer to its natural format. This means that the database innovations are also a significant alternative to systems that manage mainly numbers and characters. The management of data gets closer to how the actual work is done. Thus the relative advantage of the object-relational DBMS is considered to be moderate, since its main contribution is the extensions to the available data types.

An example of using object-relational database management systems comes from the management of spatial data. When the special spatial storage structures are not used, the necessary analysis tools must be programmed manually. On the other hand, if spatial extensions of the spatial DBMS are exploited, many analysis operations and functions are automatically available. Therefore the relative advantage of the spatial DBMS is higher than in the case of pure object-relational DBMS, but nevertheless scored as moderate. The situation is similar with XML-enabled DBMS. The old relational database management systems do not contain any operators or functions for managing

XML documents fluently. However, new XML-enabled database management systems support SQL/XML (part of the SQL:2003 standard) to enable easier XML data management. In addition, they offer versatile storage alternatives compared to traditional relational DBMS. Consequently, we can say that the relative advantage of an XML-enabled DBMS is very high. The main rationale for this score is that XML-enabled DBMSs:

- extend the available data types,
- enable management of all kind of data in a single DBMS and
- offer a multitude of new tools and functions for managing the data in a DBMS.

The database innovations presented in this research are all based on relational database management systems. Therefore is easy to say that the database innovations are very **compatible** with the relational DBMS. This means in practice that the investments in the database innovations are safe, since the database management systems can still manage the old relational systems. A big advantage is that the organizations can have all kind of data in one database. It can contain well-structured, unstructured and complex data. The data can then be managed with normal SQL commands but also special XML tools like Xpath and Xquery. Actually, relational data can be managed with XML related-tools and data in XML documents can be managed with relational tools and vice versa. The essential message is that all data can be managed together and it is easy to combine different approaches.

The disadvantages of the database innovations are not as easy to identify as the advantages. One of the clear disadvantages is however the rise of **complexity**. In object-relational DBMSs the complexity level raises only little, but in both spatial and XML-enabled DBMSs the degree of complexity is high. This complexity is illustrated clearly by the evolution of the SQL database language. Once so easy and compact, SQL has become much trickier, although the possibilities of what you can do with SQL have increased as well. Database innovations offer new tools that are required to manage the data in the database. Earlier SQL was the only tool you needed to master, but database innovations are making other areas highly important as well.

The complexity is also seen in the storage level since the normalization – the basic requirement in a relational database – is broken when objects and XML documents are stored in the database using these new data types. Of course it is debatable whether this is really a disadvantage, or if in fact it is more valuable to have the data closer to the natural form in the database than divided into number of tables. However, this also pressurised traditional database design stages: how should a conceptual model be turned into a logical model when the goal is – for example – to use user-defined types and collections, not to mention XML schemas.

A disadvantage is also the space needed for storing the new types of data. For example, storing only the element values of an XML document without the elements reduces the amount of characters that need to be stored quite considerably. For example, an XML element <StreetAddress>Helsingintie 1</StreetAddress> is 45 characters altogether, but if only the value of the Streetaddress element is stored just 14 characters are needed. In a large XML document the amount of saved character-space can turn out to be significant.

The **trialability** of the database innovations varies from low to high. The trialability estimation for object-relational DBMS, is high since many of the present database management systems represent the third generation of DBMSs and therefore the object-relational features should be tested and evaluated. However, more preparations are needed before we can test the spatial extensions of the DBMS. Depending on the DBMS, the spatial extensions might be directly available, or alternatively they (or a replacement) could need installing. In all cases the test environment must be designed and implemented and the additional tools and functions needs to be learned. Therefore the trialability of the spatial DBMS is evaluated as moderate. The trialability of the XML-enabled DBMS is low. Testing always requires new installations and probably a change of DBMS, since they all do not yet offer the XML data management features that are available in the enriched third generation of DBMSs. As the empirical evidence shows, present DBMSs are not ready for XML and therefore the readiness to test XML-enabled DBMSs is also low.

The **observability** of database innovations varies from very low to moderate. Observability is moderate for XML-enabled DBMSs, since although the language is discussed everywhere and awareness of it is increasing, pure XML-enabled DBMS solutions are still rare. Economically, XML-enabled DBMS will become increasingly interesting, as the number of XML-capable information systems rise.

The observability of spatial DBMSs is low, since real cases of using them are still uncommon. Spatial data is however managed in many information systems and we can therefore investigate the benefits of this approach. For object-relational DBMSs the degree of observability is very low. Solutions based on an object-relational DBMS are quite impossible to find, for two reasons. First of all the real cases using object-relational features are not typical. Second, the information systems might not show any signs of the use of object features unless the physical database structures are studied.

6.2.2 Consequences of database innovations

Rogers (2003) defines consequences as the changes that occur as a result of the adoption or rejection of an innovation. He further defines three dimensions of consequences: a) desirable vs. undesirable, b) direct vs. indirect and c) anticipated vs. unanticipated. Typically, the desirable, direct and anticipated consequences of an innovation go together, as do those that are undesirable, indirect and unanticipated.

One direct consequence of database innovations is that they smooth the path to e-Business for organizations. Relational database management systems have drawbacks for Business-to-Business and Business-to-Consumer actions. When a relational database management system is used, the data must usually be put into another format to transfer it to another organisation (in B2B transactions) or to publish it on the web (in a B2C situation). Typically, these transformations are examples of impedance mismatch. Object-relational database management systems ameliorate this situation somewhat because they allow related data to be stored as whole objects and all of the necessary data can be accessed easily without many joins. The enriched third generation of database management systems offers the most potential for e-Business. By providing built-in support for XML, it makes performing all kinds of e-Business actions much easier. XML documents can be produced easily and then stored in many different ways. In fact, if we place an estimation of B2B revenues (NFO Infratest GmbH & Co. KG 2003) on the same chart as the diffusion curves presented earlier, we can make an additional observation: the B2B bars form a pattern that is similar to the XML-enabled DBMS curve (see Figure 6.8). Therefore it seems that B2B revenue predictions support forecasts of XML-enabled DBMSs diffusion.

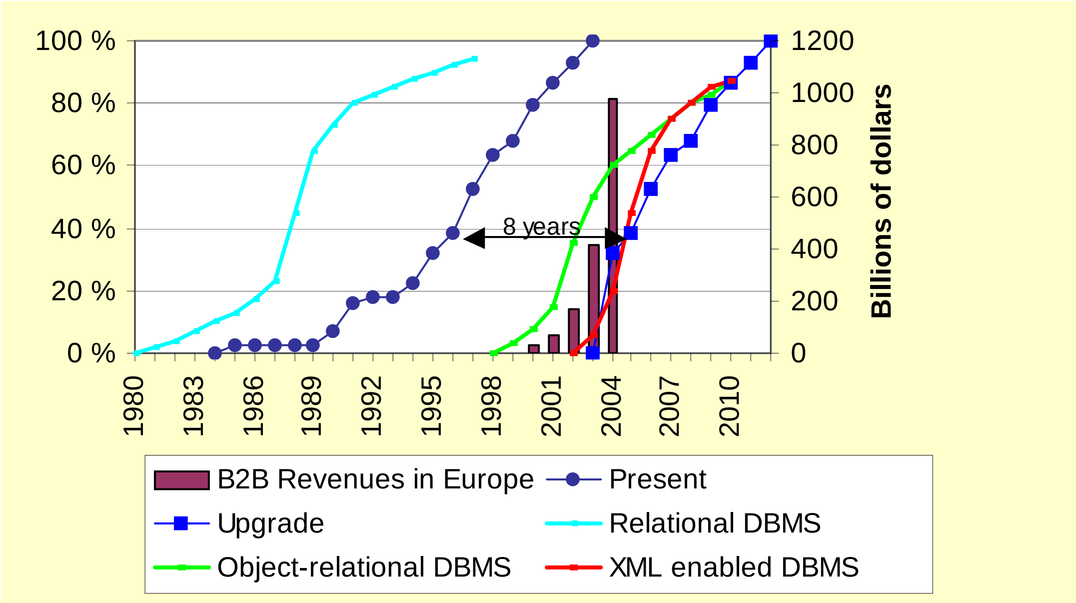


Figure 6.8: B2B revenues and the diffusion of database innovations.

An anticipated and desired consequence of database innovations is that we are seeing the emergence of enterprise-level databases, where many information systems use same information in parallel. At the moment, organizations are thinking about how to integrate information systems and what kind of solutions should be utilized. This research suggests that the aim should be to combine the databases and build larger, enterprise-level databases that hold all kinds of organizational data. This enterprise database could contain all of the data that has been processed in the organization, but there might also be separate enriched third generation databases for managing data that is specific to the information system. In this integrated environment, versatile data can be combined easily and exploited efficiently. It would also be possible to introduce new applications, such as customer relationship management, data warehousing or the use of spatial data. This would require comprehensive education and training in database skills for staff in organizations. The next step, from enterprise databases to a workflow-oriented way of doing things, is also not far away. The creation of an enterprise database already means that the information is held in the enriched third generation of databases and the next step is to provide a business layer that allows the information systems to access all data, wherever it may be located. Having database management systems that enable the management of all kinds of data reduces the level of complexity in this business layer.

For universities, institutes and other organizations offering education in the field of databases, this research produced confusing results. On one hand, it outlined those topics that are becoming important in the forthcoming years. Thus an anticipated and desired consequence is that these topics should be included in the curriculum. Database courses should also cover topics like XML and databases, spatial data management and object-relational databases. On the other hand however, it showed that information systems and databases last for a long time. From this we can conclude that there will be information systems and databases based on relational DBMSs for many years. Thus traditional relational databases should still form the basis of database-related education, but database innovations should also be included in courses and curricula.

6.2.3 The critical organizational success factors for the adoption of database innovations

The original framework for analysing the degree to which an organization might adopt database innovations operated quite well. However, it seems that the attributes could be rearranged to create a more simplified model (see Figure 6.9). Thus three groups of attributes are defined: organizational characteristics, resource-related factors and business environment. The organizational characteristics group contains attributes for describing the organization in the field of information systems. The resource-related factors group contains attributes relating to the resources that are important for adopting new technological innovations. The third group describes the business environment the organization is operating in.

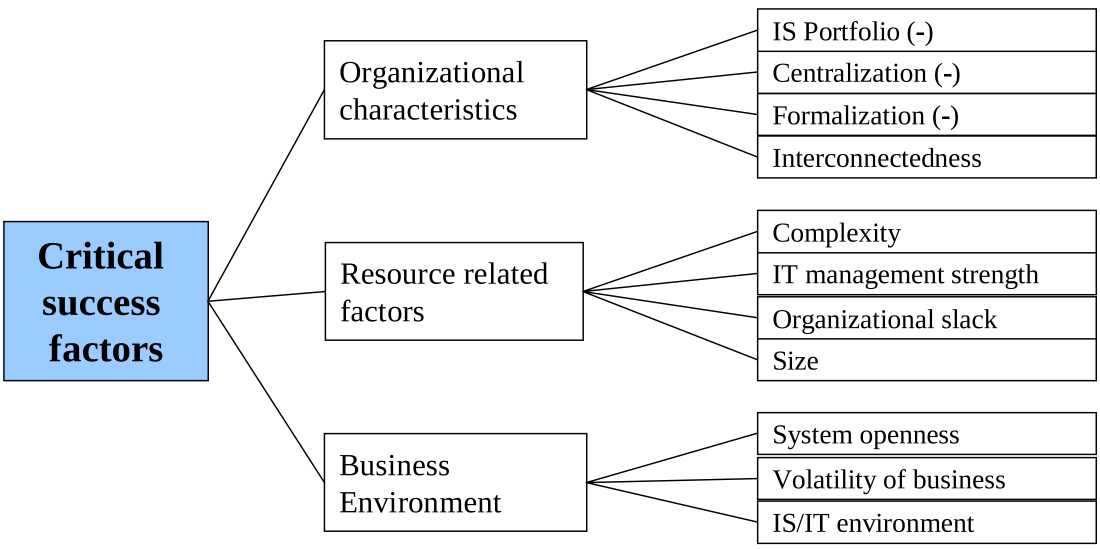


Figure 6.9: Critical success factors.

Table 6.4 presents the original attribute values, grouped and summed according to each critical success factor. The table shows the distinguishing factors between the cases. We see for example that the factors driving Salon Seudun Puhelin Ltd towards innovations are its organizational characteristics and business environment. Similarly, resource-related factors are driving Statistics Finland in this direction.

Table 6.4: Values of critical success factors.

Critical success factor	SOK	SSP	STAT	WEST	TS	OPTI
Organizational characteristics	9	**12**	7	8	7	8
Resource related factors	**12**	4	**13**	4	**13**	6
Business environment	7	9	3	5	9	5

When an organization wants to improve its prerequisites for the adoption of database innovations, it can concentrate either on improving its organizational characteristics or on resource-related factors. On the other hand, it seems very difficult to achieve improvement in the business environment-related factors. It seems impossible suddenly to alter the volatility of the business or to modify its IS/IT environment. Therefore organizations should concentrate on the other critical success factor groups rather than the business environment group, if they are looking to create a successful environment for the adoption of technological developments like database innovations. It is much easier for example to improve employee skills (complexity) or invest in more resources (organizational slack) than to alter the business environment. Similarly, the organization can work on changing its culture to become less centralized and formalized.

6.2.4 Innovation-decision process in an organization

Figure 6.10 summarizes how database innovations diffuse throughout an organization. It also shows the critical success factors defined in the previous section. An organization (the white box in the middle of the Figure) has several information systems and databases. The innovation process is related to the business environment, which has a strategic impact on the organization (Swanson 1994).

The aim and the goal of an organization is naturally that these information systems and databases match to the requirements of its business environment. Usually, the internal environment of an organization is integrated, perhaps through enterprise application integration (EAI). Typically, it would also have an integration solution with its business partners (B2B). However, this set of information systems and databases always has problems. In some organizations these problems are small, but in others they can be significant. An organization may have plans to develop its information system and database environment because of these difficulties, but also because of new business requirements. When a new information system and database are planned, new technological innovations are also discussed. Yet it might also be other way around, i.e. the innovations themselves initiate new plans. In agenda setting, needs and problems are processed and the potential usefulness of database innovations are evaluated. Typically, these so-called performance gaps trigger the innovation process. The evidence from this research confirms that performance gaps start the innovation processes. In all of the cases these performance gaps initiated an internal investigation that set the agenda for the innovation process. These performance gaps are presented in Table 6.5.

Table 6.5: Performance gaps of the case organizations.

Case	Performance gap	Innovation process
SOK	Inefficient exploitation of the data	Data modelling for information systems' integration
SSP	Inefficient exploitation of the information systems	Management of the information systems and the databases
STAT	Inefficient management and analysis of spatial data	Storing and managing spatial data in the databases
WEST	Inefficient management of inquiry data	Managing inquiries with the databases
TS	Inefficient management of customer information	Managing customer information with the databases
OPTI	Management of customisation and maintenance of the information systems	Management of the information systems and the databases

After the agenda setting stage the innovation is tested and the organization attempts to determine whether the solution might solve its problems. The innovation is then analysed to assess the possible benefits and problems of implementation. Key issues here are whether it will provide relative advantage and the degree of compatibility, complexity, observability and trialability.

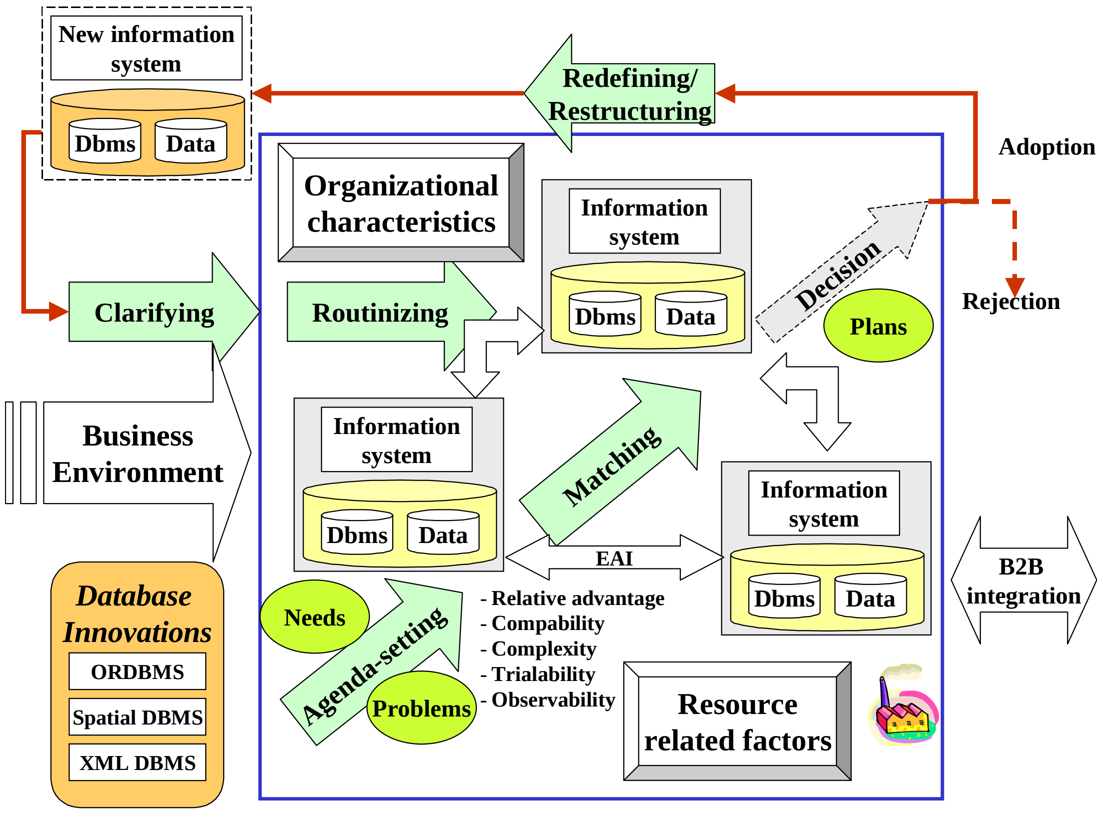


Figure 6.10: Database innovations in an organization's innovation process.

If the organization's decision-makers conclude that the innovation will not solve its problems, bridge the performance gaps or fit with corporate plans, the innovation is rejected. Otherwise the innovation is adopted and it enters the next stage of the innovation process.

If the decision is favourable, then the innovation is adapted to accommodate the organization's needs and the organization's structure is modified to fit with the innovation. Both the innovation and the organization usually change during the innovation process (Swanson 1994; Rogers 2003). The change may for example result in a creation of a new organizational unit, like in TS-Group Ltd, where a new team for the CRM system was created. In practice this restructuring might mean designing a new information system with the new database innovation.

When the new information system and innovation are put into widespread use, *clarifying* occurs. This stage is actually a social process of human interaction when the people in an organization talk about the innovation and gain a common understanding of it.

Finally the innovation becomes a part of the regular activities and loses its separate identity. This final, *routinizing* stage completes the innovation process. An important concept related to routinizing is sustainability, which defines how the innovation continues to be used after the initial innovation process is completed (Rogers 2003). Two factors influencing sustainability are participation and the degree of re-invention. Participation describes the involvement of the members of an organization in the innovation process. The degree of re-invention describes how much the innovation is modified by the adopters in the innovation process. (Rogers 2003) For example, in Optiroc Ltd participation is recognized as critical factor in the sustainability of their information systems and innovations.

This shows that the theory of innovation process in organizations can be used to interpret the diffusion of database innovations. We should also remember that there are four main elements in the diffusion of new ideas (Rogers 2003): the innovation itself, communication channels, time and social system. The first element, database innovations, has already been explained thoroughly, but the three others are largely absent from this thesis. However, it has shown (Figure 6.10) that the meaning of time is easy to understand in the innovation process. Going from the agenda-setting stage to the routinizing stage takes plenty of time, as does the diffusion of database innovations. The social system in this research has been the organization. There are many different departments and units within an organization and the role of the IS/IT department is emphasized in the innovation process of database innovations. Like Swanson (1994) wrote, the success of IS innovations may rest upon an effective partnership between the IS department and its users.

6.3 Reliability and validity

Some criticize a case study approach to research, because it is not possible to generalize the findings statistically for example (Cavaye 1996). There are however ways to generalize the case study research findings, as Walsham (1995) described. Another criticism is that case studies are subject to the influence of the researcher’s characteristics and background (Cavaye 1996). Such a line of argument was turned on its head during this thesis: a researcher must have a certain background to be able to understand and interpret the many solutions utilized in organizational information systems and databases. However, when semi-structured interviews are used, a researcher’s readiness also plays a significant role and might influence the results of the research.

The validity of the research, i.e. the ability of the research methods to measure the right things, seems to be good. The selected research methods produced detailed information about the information systems and databases, as they were designed to do at the beginning of the project. Maybe the interview script could have been even more specific, since in a couple of interviews the interviewees did not understand some terms or the formulations of the themes. These situations were nevertheless solved during the interviews and necessary information was received.

Yin (1994) divides validity into three sub-concepts: construct, internal and external validity, but only two of them are suitable for interpretive case studies like this research. These are described in Table 6.6.

Table 6.6: Construct and External validity concepts by Yin (1994).

Concept	Description
Construct validity	Establishing correct operational measures for the concepts being studied
External validity	Establishing the domain to which a study's findings can be generalized

The researcher can affect construct validity in the data collection process by using multiple sources of evidence and establishing a chain of evidence. In this study, data collection was based on multiple data sources, not just the interviews. This gave a larger view of the phenomena and raised the construct validity of the research. Having a draft case study report reviewed by the key informants also increases the construct validity. In this project the interview transcripts were sent to the interviewees for confirmation. Every case also finished with a final meeting where the case report was published. After this the organization and the people involved had an opportunity to add necessary topics or make corrections to the case report. However, only very small changes were requested at this point in time, probably because the interview transcripts were already checked earlier.

The problem of external validity has been a major barrier in doing case studies (Yin 1994), since it concerns the generalizations of case study findings. External validity can be increased during research design by using replication-logic for multiple case studies. This bases the selection of cases on the idea of literal replication, which means that similar results are expected in every case. This was a reason for limiting the cases to organizations utilizing database technology, rather than those that were mainly developing information systems on top of the databases. If the goal of the research had been to produce contrasting results (but for predictable reasons) then a theoretical replication approach could have been utilized. In this situation, organizations and companies that have IT/IS as their main line of business would have been selected. However, no such companies were involved in this study.

The cases were a heterogeneous group of organizations, which allowed for a more general examination of the phenomena than would have been the case if the organizations had only represented a single line of business. Every case had a specific research focus in addition to the general framework, although these were also studied in the other organizations. For example, spatial data management was viewed in every case, but it was the focus of Statistics Finland. Common trends were investigated during data analysis and general results were described. These were also used when analysing the diffusion of database innovations. The organizations' decision processes and internal structural characteristics were not analysed in detail, but these issues influenced the evaluations made for the orientation value. Both issues were referenced and discussed during the interviews, but maybe a little more effort could have been placed on them to simplify the assessment of orientation value.

The reliability of this research was improved by following certain ways of working. There are two main ways in which the reliability of a case study research can be improved. These are a case study protocol and a case study database. A case study protocol is essential in a multiple case study and guides the researcher in carrying out the data collection. (Yin 1994). It should contain:

- An overview of the case study project.
- Field procedures.
- Case study questions.
- A guide for the case study report.

In this research a case study protocol was maintained, although it was not as detailed as Yin (1994) proposed.

Benbasat et al. (1987) wrote that clear descriptions of case selection criteria and data collection processes should be provided. Also Dubé and Paré (2003) wrote that the data collection and analysis processes must be described thoroughly. Both of these requirements have some similarity to Yin's definition of a case study protocol. To improve reliability, these issues are also presented in section 3, where the whole research approach is described quite thoroughly.

A case study database is about organizing and documenting the data collected for the case studies (Yin 1994). It should contain case study notes and documents, tabular materials and narratives. In this research the case study notes were stored in a special database application to improve ease of access and simplify analysis. The case study documents were all collected in separate folders case by case. Tabular materials produced by the researcher from the case study material were stored in Excel-files. Finally the narratives, i.e. the open-ended answers to the questions in the case study protocol, were produced from the data in the database application. These narratives were also collected in a specific folder. Having all the case study material easily accessible proved to be very advantageous.

The reliability of the research is also strengthened if multiple investigations produce similar results. In this sense the reliability of the research is questionable. However, since the information systems and databases are changing, another study would probably not come to the same conclusions. Repeating the research would no doubt produce different results. The information systems and databases have already developed from the point when the interviews were arranged. However, another study might confirm the results of this research by showing that developments occurred in the case organizations. Reliability can also be assessed technically, i.e. how the research was conducted. Technically the research is reliable, since the research is documented very thoroughly.

Diffusion research has been criticized for different weaknesses. One of the most serious shortcomings is pro-innovation bias. This implies that an innovation should:

- Be diffused and adopted.
- Be diffused more rapidly.
- Not be re-invented or rejected.

One typical reason for pro-innovation bias is that actors whose purpose is to promote the innovation fund the research. (Rogers 2003) No database company funded this project. However, the researcher might have had an overly-positive attitude toward database innovations.

Other mentioned weaknesses of diffusion research are individual-blame bias and recall problem. Both these weaknesses were avoided in this study. Individuals were not the research targets, rather information systems. As the recall problem exists only when diffusion is analysed after adoption, this weakness is also averted, because the research investigated the early stages of diffusion and no adoption of object-relational, spatial or XML-enabled DBMS had actually happened.

6.4 Future research

This project raised many ideas for future study. The first one is to conduct similar research, but within organizations that are mainly developing information systems, i.e. their line of business is IS-development. This kind of research could be understood as studying the early adopters of DBMS innovations. This might give more precise information about the time period in which IS-user organizations will adopt DBMS innovations.

The second idea is to study these same organizations after some years and analyse the changes that have happened in their information systems and databases. This would provide time-series data and make it possible to analyse whether the development had occurred or not. There would also be the possibility for deeper study into the reasons why the organizations developed in particular ways.

The third topic is to study how personnel have coped with the forthcoming change in information systems and databases. Using DBMS innovations requires different and updated skills. It might be interesting to study how staff were prepared for the change and the extent to which preparation actually succeeded. It is also reasonable to study how employees feel about the service the DBMS innovations have delivered. What is the change that has happened from the personnel point of view?

The fourth possible research topic is to concentrate on spatial data and describe how these features are exploited in the information systems and databases. This research claimed that spatial data management is becoming increasingly important and beginning to enter databases. An interesting research topic would be the description of spatial data management in different cases. It would also be interesting to describe the real benefits that have been achieved and the possible problems encountered.

The fifth possible is similar to the fourth, but now the focus is on XML and how it has diffused in the information systems and the databases. This area could also be divided into many further research categories. One possibility is to concentrate on the management of data-centric or document-centric XML documents. Another alternative is to concentrate on native XML databases and how those have diffused in the organizations. However, this topic area might be wiser to postpone for couple of years, since XML has not yet made a great deal of progress in this area.

All of the previous research suggestions would need to be case studies. This research could also be done on a more general level. It would be interesting to do a survey of a large number of organizations and try to get a picture of the state of their databases. The survey could confirm the results of this research, if it is prepared in a way that is comparable with this project. However, an in-depth understanding of the phenomena is achieved with a case study research, but with a survey similar level of depth is not possible.

REFERENCES

ABA Bank (2003). *Customer retention: Getting a CRM program of on the Right Start*. ABA Bank Marketing, 35**:** 8.

Achterberg, J. - Fischer, S. J. - Vinig, T. G. (1993). Identifying different paradigms for managing information technology. In *Proceedings of the Computer personnel research,* St Louis, Missouri, United States, ACM Press, 37 - 55.

Agarwal, R. - Sambamurthy, V. (2002). Principles and models for organizing the IT function. *MIS Quarterly,* 1(1): 1 - 16.

Akoka, J. - Comyn-Wattiau, I. - Lammari, N. (2003). Supporting Database Evolution: Using Ontologies Matching. In *Object-Oriented Information Systems,* ed. by, 2817**:** 284 - 288. Springer-Verlag Heidelberg.

Anderson, W. O. J. (2001). Customer relationship management in an e-business environment. In *Proceedings of the IEMC '01,* Albany, NY, USA, IEEE, 311 - 316.

Arsanjani, A. (2002). Developing and Integrating Enterprise Components and Services. *Communications of the ACM,* 45(10): 31 - 34.

Atkinson, M. - Bancilhon, F. - DeWittt, D. - Diitrich, K. - Mater, D. - Zdonik, S. (1989). The Object- Oriented Database System Manifesto. In *Proceedings of the First International Conference on Deductive and Object-Oriented Databases*, Kyoto, Japan, 40 - 57.

Atwood, T. - Cattell, R. G. (1993). *Object Database Standard: ODMG-93*. Morgan Kaufmann Publishers Inc.

Ball, G. L. - Park, J. - Ram, S. (1999). *Semantic-model support for geographic information systems*. Computer, 32**:** 74 - 81.

Banerjee, S. (2002), Oracle XML DB Technical White Paper, Oracle Corporation, http://otn.oracle.com/tech/xml/xmldb/pdf/xmldb_92twp.pdf, Retrieved on 1.8.2002.

Banerjee, S. - Krishnamurthy, V. - Krishnaprasad, M. - Murthy, R. (2000). Oracle 8i - The XML Enabled Data Management System. In *Proceedings of the 16th International Conference on Data Engineering,* San Diego, CA, USA, 561 - 568.

Barron, T. M. - Chiang, R. H. L. (1995). Quality issues in database reverse engineering: an overview. In *Proceedings of the Engineering Management Conference*, Singapore, IEEE, 185 - 189.

Barron, T. M. - Dey, D. - Storey, V. C. (1999). Improving database design through the analysis of relationships. *ACM Transactions on Database Systems (TODS),* 24(4): 453 - 486.

Barry, D. - Bartels, D. K. - Berler, M. - Cattell, R. G. - Eastman, J. - Gamerman, S. - Jordan, D. - Springer, A. - Strickland, H. - Wade, D. (1997). *The object database standard: ODMG 2.0*. Morgan Kaufmann Publishers Inc.

Barry, D. - Berler, M. - Cattell, R. G. - Eastman, J. - Jordan, D. - Russell, C. - Schadow, O. - Stanienda, T. - Velez, F. (2000). *The Object Data Standard: ODMG 3.0*. Morgan-Kaufmann.

Batini, C. - Lenzerini, M. - Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys (CSUR),* 18(4): 323 - 364.

Becker, J. - Gille, M. - zur Muehlen, M. (2002). Workflow Application Architectures: Classification and Characteristics of Workflow-based Information Systems. In *Workflow Handbook 2002. Future Strategies*, ed. by L. Fischer**:** 39 - 50. Lighthouse Point.

Beech, D. - Bernstein, P. - Carie Brodie, M. L. - Gray, P. - Lindsay, B. - Stonebraker, M. - Rowe, L. (1990). Third-generation database system manifesto. *ACM SIGMOD Record*, 19(3): 31 - 44.

Begg, C. - Connolly, T. (2002). *Database Systems: A Practical Guide to Design, Implementation, and Management*, 3 Ed. Addison-Wesley.

Begg, C. - Connolly, T. (2004). *Database Systems: A Practical Approach to Design, Implementation, and Management*, 4 Ed. Addison-Wesley.

Benbasat, I. - Goldstein, D. K. - Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, 11(3): 369 - 386.

Benbasat, I. - Moore, G. C. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, 2(3): 192 - 222.

Benham, S. E. (2003). IBM XML-Enabled Data Management Product Architecture and Technology. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 91 - 121. Addison-Wesley Boston.

Bennett, K. H. - Rajlich, V. T. (2000). Software Maintenance and Evolution: a Roadmap. In *Proceedings of the Future of Software Engineering*, Limerick, Ireland, IEEE, 73 - 87.

Bergeron, B. (2002). *Essentials of CRM: a guide to customer relationship management*. John Wiley & Sons, Inc.: New York.

Bertino, E. - Catania, B. (2001). Integrating XML and databases. *IEEE Internet Computing*, 5(4): 84 - 88.

Bird, L. - Goodchild, A. - Routledge, N. (2002). UML and XML schema. In *Proceedings of the thirteenth Australasian conference on Database technologies*, Melbourne, Victoria, Australia, Australian Computer Society, Inc., 157 - 166.

Blackwood, P. (2000). Eleven Steps to Success in Data Warehousing. *Business Journal*, 14(44): 26 - 27.

Bonifati, A. - Cattaneo, F. - Ceri, S. - Fuggetta, A. - Paraboschi, S. (2001). Designing Data Marts for Data Warehouses. *ACM Transactions on Software Engineering and Methodology*, 10(4): 452 - 483.

Bourret, R. (2002a), XML and Databases, http://www.rpbourret.com/xml/XMLAndDatabases.htm, Retrieved on 4.8.2002.

Bourret, R. (2002b), XML Database Products: Native XML Databases, http://www.rpbourret.com/xml/ProdsNative.htm, Retrieved on 4.8.2002.

Bourret, R. (2002c), XML Database Products: XML-Enabled Databases, http://www.rpbourret.com/xml/ProdsXMLEnabled.htm, Retrieved on 4.8.2002.

Brancheau, J. C. - Janz, B. D. - Wetherbe, J. C. (1996). Key Issues in information systems management: 1994-95 SIM Delphi results. *MIS Quarterly*, 20(2): 225 - 242.

Brandin, C. (2003). Information modeling with XML. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 3 -17. Addison-Wesley Boston.

Briand, L. C. (2003). Software documentation: how much is enough? In *Proceedings of the Software Maintenance and Reengineering Conference*, Ottawa, Ontario, Canada, IEEE, 13 - 15.

Broadbent, M. - Darke, P. - Shanks, G. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. *Information Systems Journal*, 8(4): 273 - 289.

Brookshaw, C. - Hammond, E. (1998). Add Objects to your world. *InfoWorld*, 20(28): 74 - 75.

Brown, P. - Chaudhri, A. B. - Herbach, M. - Koerner, A. (2001). Requirements for XML Functionality in Relational Database Management Systems. In *Proceedings of the 12th International Workshop on Database and Expert Systems Applications*, Munich, Germany, 55 - 57.

Brown, P. - Stonebraker, M. - Moore, D. (1999). *Object-relational DBMSs tracking the next great wave*, 2 Ed. Morgan Kaufmann Publishers.

Browne, G. J. - Ramesh, V. (1999). Expressing casual relationships in conceptual database schemas. *Journal of Systems and Software,* 45(3): 225 - 232.

Burn, J. - Dhillon, G. - Hackney, R. (2000). Challenging assumptions for strategic information systems planning: theoretical perspectives. *Communications of AIS,* 3(Article 9): 1 - 24.

Busborg, F. - Christiansen, J. G. B. - Tryfona, N. (1999). starER: a conceptual model for data warehouse design. In *Proceedings of the ACM international workshop on Data warehousing and OLAP*, Kansas City, Missouri, United States, ACM Press, 3 - 8.

Busichia, G. - Ferreira, J. E. (1999). Database modularization design for the construction of flexible information systems. In *Proceedings of the Database Engineering and Applications, IDEAS '99,* Montreal, Quebec, Canada, IEEE, 415 - 422.

Calero, C. - Genero, M. - Piattini, M. (2001). Assurance of conceptual data model quality based on early measures. In *Proceedings of the Quality Software,* Hong Kong, China, IEEE, 97 - 103.

Carr, D. F. (2001). XML Native Databases. *Internet World,* 7(14): 54 - 55.

Carson, G. (2000). Spatial standardization. *Computer Graphics,* (August): 38 - 41.

Cavaye, A. L. M. (1996). Case Study Research: a multi-faceted research approach for IS. *Information Systems Journal,* 6: 227 - 242.

Cavero, J. M. - Marcos, E. - Vela, B. (2001). Extending UML for Object-Relational Database Design. In *Proceedings of the UML 2001 - The Unified Modeling Language, Modeling Languages, Concepts, and Tools*, Toronto, Canada, Springer-Verlag Heidelberg, 225 - 239.

Cavero, J. M. - Marcos, E. - Vela, B. (2004). A methodological approach for object-relational database design using UML. *Informatik Forschung und Entwicklung,* (18): 152 - 164.

Champion, M. (2001). Storing XML in Databases. *eAI Journal*, October 2001.

Chaudhri, A. B. - Rashid, A. - Zicari, R., Eds. (2003). *XML Data Management - Native XML and XML-Enabled Database Systems*. Addison-Wesley: Boston.

Chawla, S. - Fetterer, A. - Liu, X. - Lu, C. - Shekhar, S. - Ravada, S. (1999). Spatial databases - accomplishments and research needs. *IEEE Transactions on Knowledge and Data Engineering,* 11(1): 45 - 55.

Chawla, S. - Shekhar, S. (2003). *Spatial Databases - A Tour*. Pearson Education Inc.: Upper Saddle River, New Jersey.

Chen, J. - DeWitt, D. J. - Tian, F. - Zhang, C. (2002). The Design and Performance Evaluation of Alternative XML Storage Strategies. *ACM SIGMOD Record*, 31(1): 5 - 10.

Cheng, J. - Xu, J. (2000). XML and DB2. In *Proceedings of the 16th International Conference on Data Engineering*, San Diego, CA, USA, 569 - 573.

Cheong, K. H. - Haglind, M. (2001). Striving towards strategic IT & business alignment in small and medium -sized electric utilities - a pragmatic approach. In *Proceedings of the Electricity Distribution*, Amsterdam, Netherlands, IEEE, 7 - 12.

Chester, T. M. (2001). Cross-platform integration with XML and SOAP. *IT Professional,* 3(5): 26 - 34.

Chin, W. W. - Marcolin, B. L. (2001). The future of diffusion research. *ACM SIGMIS Database,* 32(3): 7 - 12.

Christensen, B. H. - Gottschalk, P. - Watson, R. T. (2000). Global comparisons of key issues in IS management: extending key issues selection procedure and survey approach. In *Proceedings of the System Sciences*, Hawaii, IEEE, 1 - 10.

Clark, D. - Cook, K. C. - Faber, B. D. - Zachry, M. (2001). The changing face of technical communication: new directions for the field in a new millennium. In *Proceedings of the Systems Documentation*, Santa Fe, New Mexico, USA, ACM Press, 248 - 260.

Coffee, P. (2002). *In pursuit of a CRM process*. eWeek, 19**:** 47 - 50.

Computer Dealer News (2001). *Will that database be native or enabled?* Computer Dealer News, 17**:** 21.

Cook, C. R. - Visconti, M. (2002). An overview of industrial software documentation practice. In *Proceedings of the Computer Science Society Conference*, Valparaiso, Chile, IEEE, 179 - 186.

Córcoles, J. E. - Gonzáles, P. (2002). Analysis of different approaches for storing GML documents. In *Proceedings of the Advances in geographic information systems*, McLean, Virginia, USA, ACM Press, 11 - 16.

Coyle, M. - Goyal, B. - Liu, D.-R. - Sarkar, S. - Shekhar, S. (1997). Data Models in Geographic Information Systems. *Communications of the ACM,* 40(4): 103 - 111.

CSC (2002a), Critical Issues of Information Systems Management, Computer Science Corporation, http://www.csc.com/aboutus/uploads/CI_Report.pdf, Retrieved on 30.7.2004.

CSC (2002b), Technology Issues for Financial Executives, Computer Science Corporation, http://www.csc.com/solutions/managementconsulting/ knowledgelibrary/1022.shtml, Retrieved on 29.6.2004.

CSC - FEI (2003), Technology Issues for Financial Executives, Computer Science Corporation, http://www.csc.com/solutions/managementconsulting/ knowledgelibrary/1328.shtml, Retrieved on 29.6.2004.

Darrow, B. - Rooney, P. (2001). XML Database War. *Computer Reseller News*, (943): 3 - 4.

Date, C. J. (1995). *An Introduction to Database Systems*, 6 Ed. Addison-Wesley: Reading, Massachussetts.

Date, C. J. - Darwen, H. (1998). *Foundation for Object/Relational Databases-The Third Manifesto*, 1 Ed. Addison-Wesley.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3): 319 - 339.

Dedrick, J. - Gurbaxani, V. - Kraemer, K. L. (2003). Information technology and economic performance: A critical review of the empirical evidence. *ACM Computing Surveys (CSUR)*, 35(1): 1 - 28.

DeJesus, E. X. (2000). XML Enters The DBMS Arena. *Computerworld*, 34(44): 80 - 82.

DeLone, W. H. - McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1): 60 - 95.

DeLone, W. H. - McLean, E. R. (2002). Information Systems Success Revisited. In *Proceedings of the System Sciences*, Hawaii, United States, IEEE, 2966 - 2976.

Dodds, D. - Hasan, H. - Hyland, P. - Veeraraghavan, R. (2000). Approaches to the development of multi-dimensional databases: lessons from four case studies. *ACM SIGMIS Database*, 31(3): 10 - 23.

Dodge, J. (2001). XML gets bigger DB 2 role. *eWeek*, 18(49): 1 - 2.

Duchon, D. - Kaplan, B. (1988). Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quarterly*, December 1988: 570 - 586.

Dudé, L. - Paré, G. (2003). Rigor in information systems positivist case research: current practices, trends, and recommendations. *MIS Quarterly,* 27(4): 597 - 635.

Duncan, R. (1972). Characteristics of Organizational Environments and Perceived Environmental Uncertainty. *Administratice Science Journal,* (17): 313 - 327.

Dyché, J. (2002). *The CRM handbook: a business guide to customer relationship management*. Addison-Wesley.

Dyck, T. (2001). XML Database Doubts. *eWeek,* 18(47): 67.

Eberhardt, C. (2001), A CRM Starter Pack: Strategic Steps to Success, http://www.destinationcrm.com/print/default.asp?ArticleID=1816, Retrieved on 23.8.2003.

Edwards, R. (2003). A generic architecture for storing XML documents in a relational database. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 189 - 258. Addison-Wesley Boston.

Eisenberg, A. - Malaika, S. - Melton, J. (2003). Standards for databases on the grid. *ACM SIGMOD Record,* 32(3): 92 - 100.

Eisenberg, A. - Melton, J. (1999). SQL:1999, formerly know as SQL3. *ACM SIGMOD Record,* 28(1): 131 - 138.

Eisenberg, A. - Melton, J. (2002). SQL/XML is making good progress. *ACM SIGMOD Record,* 31(2): 101 - 108.

Elmasri, R. - Navathe, S. B. (2000). *Fundamentals of Database Systems,* 3 Ed. Addison-Wesley: Reading, Massachussetts.

EU Commission (2003), Commission recommendation concerning the definition of micro, small and medium.sized enterprises, EU, http://europa.eu.int/comm/enterprise/enterprise_policy/sme_definition/decision_sme_en.pdf, Retrieved on 5.8.2004.

Feeny, D. F. - Willcocks, L. P. (1998). Core IS capabilities for exploiting information technology. *Sloan Management Review,* Spring 1998.

Finnegan, P. - Sammon, D. (2000). The ten commandments of data warehousing. *ACM SIGMIS Database,* 31(4): 82 - 91.

Florescu, D. - Kossman, D. (1999). Storing and Querying XML Data using an RDBMS. *IEEE Data Engineering Bulletin,* 22(3): 27 - 34.

Fong, A. - Fong, J. - Wong, H. K. (2003). Performance analysis between XML-enabled database and a native XML database. In *XML Data Management - Native XML and XML-enabled database systems,* ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 547 - 570. Addison-Wesley Boston.

Fortune (2000). Keep the 0010001 Happy. *Fortune: Summer2000 Technology Guide*, 141(11): 65 - 66.

Fowler, D. (2000). *The Personal Touch*. netWorker, 4**:** 25 - 29.

Fremantle, P. - Khalaf, R. - Weerawarana, S. (2002). Enterprise services. *Communications of the ACM*, 45(10): 77 - 82.

Friis-Christensen, A. - Jensen, C. S. - Tryfona, N. (2001). Geographic Data Modeling: Requirements and research issues in geographic data modeling. In *Proceedings of the ACM international symposium on Advances in geographic information systems*, Atlanta, Geogria, USA, ACM, 2 - 8.

Gallupe, R. B. (2000). Images of information systems in the early 21st century. *Communications of AIS,* 3(3): 1 - 17.

Gardner, S. R. (1998). Building the data warehouse. *Communications of the ACM*, 41(9): 52 - 60.

Gaudin, S. (2004). Report: 25 % of Critical Data is Flawed. *Datamation*, (May 19).

Gennick, J. (2003). *SQL in, XML out*. Oracle Magazine, XVII**:** 99 - 102.

Gibson, N. - Kelly, S. - Holland, C. P. - Light, B. (1999). Focus issue on legacy information systems and business process engineering: a business perspective of legacy information systems. *Communications of AIS,* 2(1es): Article No. 7.

Glass, R. L. (2002). Predicting future maintenance cost, and how we're doing it wrong. *Software, IEEE*, 19(6): 112 - 113.

Golfarelli, M. - Rizzi, S. (1998). A methodological framework for data warehouse design. In *Proceedings of the ACM international Workshop on Data warehousing and OLAP*, Washington, D.C., United States, ACM Press, 3 - 9.

Gonzales, M. L. (2000). Seeking Spatial Intelligence. *Intelligent Enterprise,* 3(2).

Goodchild, M. F. - Longley, P. A. - Maguire, D. J. - Rhind, D. W. (2001). *Geographic information systems and science*. John Wiley & Sons, Ltd.: Chishester.

Goodhue, D. L. - Kirsch, L. J. - Quillard, J. A. - Wybo, M. D. (1992). Strategic Data Planning: Lessons from the Field. *MIS Quarterly,* 16(1): 11 - 34.

Gorla, N. (2003). Features to consider in a data warehousing system. *Communications of the ACM*, 46(11): 111 - 115.

Gupta, Y. P. - Karimi, J. - Somers, T. M. (1997). Alignment of a firm's competitive strategy and information technology management sophistication: the missing link. *IEEE Transactions on Engineering Management*, 44: 399 - 413.

Guting, R. (1994). An introduction to spatial database systems. *Very Large Databases Journal*, 3(4): 357 - 399.

Ha, S. - Kim, K. (2001). Mapping XML Documents to the Object-Relational Form. In *Proceedings of the IEEE International Symposium on Industrial Electronics, ISIE 2001*, 1757 - 1761.

Harding, E. U. (2001). *Kathy Calta: Focusing on customers*. Software Magazine, 21**:** 12.

Hartono, E. - Santhanam, R. (2003). Issues in linking information technology capability to firm performance. *MIS Quarterly,* 27(1): 125 - 153.

Hasselbring, W. (2000). Information System Integration. *Communications of the ACM*, 43(6): 33 - 38.

Heales, J. (2000). Factors affecting information system volatility. In *Proceedings of the Information systems*, Brisbane, Queensland, Australia, Association for Information Systems, 70 - 83.

Hicks, M. (2002). Oracle 9i to treat XML as native data. *InfoWorld*, 19(12): 9 - 10.

Hoffer, J. A. - Prescott, M. B. - McFadden, F. R. (2004). *Modern Database Management*. Pearson Prentice Hall: Upper Saddle River, New Jersey.

Hohenstein, U. (2003). Supporting XML in Oracle9i. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 123 - 164. Addison-Wesley Boston.

Huang, S. - Tilley, S. (2003). Towards a documentation maturity model. In *Proceedings of the 21st annual international conference on Documentation*, San Francisco, CA, USA, ACM Press, 93 - 99.

IDC (2003), IDC Releases Database Market Share Study, IDC, http://databases.about.com/b/a/016881.htm, Retrieved on 17.10.2003.

Idris, N. B. - Sahibuddin, S. - Sulaiman, S. (2002). Production and maintenance of system documentation: what, why, when and how tools should support the practice. In *Proceedings of the Software Engineering Conference*, Kuala Lumpur, Malaysia, IEEE, 558 - 567.

IEEE (1993). *IEEE Std. 1219: Standard for software Maintenance*. IEEE Computer Society Press: Los Alamitos, CA, USA.

Inmon, W. H. (1992). *Building the Data Warehouse*. John Wiley & Sons, Inc.: New York, NY, USA.

ISO (2000), Draft Business Plan of ISO/TC 211 - Geographic information/ Geomatics, ISO, http://isotc.iso.ch/livelink/livelink/fetch/2000/2122/687806/ ISO_TC_211__Geographic_information_Geomatics_.pdf? nodeid=1002196&vernum=1, Retrieved on 7.4.2003.

ISO (2003), Information technology -- Database languages -- SQL -- Part 14: XML-Related Specifications (SQL/XML), http://www.iso.ch/iso/en/stdsdevelopment/techprog/workprog/TechnicalProgrammeProjectDetailPage.TechnicalProgrammeProjectDetail?csnumber=35341, Retrieved on 9.12.2003.

James, G. (2003). *CRM rises to the top*. Electronic Business, 29**:** 13.

Jazzar, A. - Scacchi, W. (1995). Understanding the requirements for information system documentation: an empirical investigation. In *Proceedings of the Organizational computing systems*, Milpitas, California, USA, ACM Press, 268 - 279.

Johnson, J. - Henderson, A. (2002). Conceptual models: begin by designing what to design. *interactions*, 9(1): 25 - 32.

Jung, H.-H. - Savioz, P. (2003). How can enterprises diagnostically and interactively control their technology-based strategies? An action research case. In *Proceedings of the Management of Engineering and Technology*, Portland, USA, IEEE, 39 - 46.

Kambayashi, Y. - Kumar, V. - Mohania, M. - Samtani, S. (2004). Recent Advances and Research Problems in Data Warehousing. *Lecture Notes in Computer Science*, 1552: 81 - 92.

Kankanhalli, A. - Tanudidjaja, F. - Sutanto, J. - Tan, B. C. Y. (2003). The role of IT in successful knowledge management initiatives. *Communications of the ACM*, 46(9): 69 - 73.

Khan, L. - Rao, Y. (2001). A performance evaluation of storing XML data in relational database management systems. In *Proceedings of the third international workshop on Web information and data management*, Atlanta, Georgia, USA, ACM Press, 31 - 38.

Kim, J. - Lee, W. - Lee, K. (2001). The Cost Model for XML Documents in Relational Database Systems. In *Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications*, Beirut, Lebanon, 185 - 187.

Kimball, R. (2001). *A Trio of Interesting Snowflakes*. Intelligent Enterprise, 4**:** 30 - 32.

Kimball, R. (2002). *Divide and Conquer*. Intelligent Enterprise, 5**:** 20 - 22.

King, P. (2002), The Top 10 Reasons CRM Projects Fail, http://www.crmnet.at/Artikel/TopTenReasons.php, Retrieved on 23.8.2003.

Kontio, J. (2003). RETRO-hankkeen loppuraportti: Tietokantatrendit yrityksissä ja organisaatioissa. *TUCS National Publication,* (No 2, December 2003).

Kontio, J. (2004). Combining XML And Databases - Perceptions In Six Finnish Organizations. In *Proceedings of the Databases and Applications*, Innsbruck, IASTED.

Korth, H. F. - Silberschatz, A. - Sudarshan, S. (1996). *Database systems concepts*. McGraw-Hill: Portland, Oregon.

Kroenke, D. M. (2004). *Database Processing: Fundamentals, Design and Implementation*, 9 Ed. Pearson Prentice Hall: Upper Saddle River, New Jersey.

Krogstie, J. - Solberg, A. (1994). A Software Maintenance in Norway: A survey investigation. In *Proceedings of the Software Maintenance*, Victoria, BC, Canada, IEEE, 304 - 313.

Lake, R. (2001). GML 2.0 Enabling the Geospatial Web. *Geospatial Solutions*, (July 2001): 38 - 41.

Larson, P.-Å. - Florescu, D. - Graefe, G. - Moerkote, G. - Pirahesh, H. - Schöning, H. (2001). XML Data Management: Go Native or Scruce Up Relational Systems? In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, Santa Barbara, California, United States, ACM Press, 620.

Lear, A. C. (1999). XML seen as integral to application integration. *IT Professional*, 1(5): 12 - 16.

Lee, A. S. (2001). Editorial. *MIS Quarterly*, 25(1): iii - vii.

Lee, G. (2002), Simple Strategies for Complex Data: Oracle9i Object-Relational Technology, http://otn.oracle.com/products/oracle9i/pdf/simple_strat_for_complex_rel2.pdf, Retrieved on 19.8.2002.

Lee, J.-N. - Pan, S. L. (2003). Using e-CRM for a unified view of the customer. *Communications of the ACM*, 46(4): 95 - 99.

Lehman, M. (1980). Programs, Life Cycles and the Lawas of Software Evolution. In *Proceedings of the IEEE*, Pistacaway, N.J., IEEE, 1060 - 1076.

Leon, M. (2001), Find a home for your XML data, http://www.infoworld.com/articles/hn/xml/01/04/02/010402hnxml.xml, Retrieved on 1.10.2002.

Levin, R. (1998). The Trickle of Objects. *Information Week*, (688): 95 - 99.

Linthicum, D. S. (2001). *B2B Application Integration: e-Business-Enable Your Enterprise*. Addison-Wesley: Indiapolis.

Loukas, T. - Spencer, T. (1999). From Star to Snowflake to ERD: Comparing Data Warehouse Design Approaches. *Enterprise Systems*, (10/99).

Lyytinen, K. - Rose, G. M. (2003). The Disruptive Nature of Information Technology Innovations: The Case of Internet Computing in

Systems Development Organizations. *MIS Quarterly*, 27(4): 557 - 595.

Länsi-Suomen lääninhallituksen peruspalvelujen arviointiryhmä (2002). *Peruspalvelut Länsi-Suomen läänissä 2001*. Länsi-Suomen lääninhallituksen julkaisusarja. Länsi-Suomen lääninhallitus: 155.

Martyn, T. (2004). Reconsidering Multi-Dimensional schemas. *ACM SIGMOD Record*, 33(1): 83 - 88.

McAuley, J. - Sacks, R. - Waddington, P. (1997). Aligning IT and organisational strategies to avoid information overload. In *Proceedings of the IEE Colloqium on IT Strategies for Information Overload*, London, UK, IEE, 3/1 - 3/4.

Meier, W. M. (2003). eXist Native XML Database. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari: 43 - 68. Addison-Wesley Boston.

Moreira, H. M. - Sousa, M. J. C. (1998). A Survey on Software Maintenance Process. In *Proceedings of the Software Maintenance*, Bethesda, MD, USA, IEEE, 265 - 274.

Morris, H. - Sonnen, D. (2000). *Location in CRM: Linking Virtual Information to the Real World*. An IDC White Paper.

Munch, S. (2000). *Building Oracle XML Applications*. O'Reilly.

Needleman, M. (2001). XML Schema language: The new way of coding XML. *Serials Review*, 27(2): 49 - 52.

NFO Infratest GmbH & Co. KG (2003), Monitoring Information Economics - 6th Factual Report, http://193.202.26.196/bmwi_english/Faktenbericht_6/ main.asp, Retrieved on 26.4.2004.

Nguyen, T. N. (2000). Modeling EAI-based e-business solutions. In *Proceedings of the Systems, Man, and Cybernetics*, IEEE, 286 - 291.

Noah, S. A. M. (2000). Intelligent system for conceptual modelling of databases. In *Proceedings of the TENCON*, Kuala Lumpur, Malaysia, IEEE, 504 - 509.

OpenGIS (1999), OpenGIS Simple Features Specification for SQL Revision 1.1, http://www.opengis.org, Retrieved on 14.3.2003.

OpenGIS (2002), OpenGIS Reference Model, Open GIS Consortium, http://www.opengis.org, Retrieved on 14.3.2003.

OpenGIS (2003a), OGCs Role in Spatial Standards World, Open GIS Consortium, http://www.opengis.org/press/?page=papers&view=20010225.TS. Stds, Retrieved on 15.12.2003.

OpenGIS (2003b), OpenGIS Geography Markup Language (GML) Implementation Specification, Open GIS Consortium, http://www.opengis.org/docs/02-023r4.pdf, Retrieved on 15.12.2003.

Oracle (2002), Oracle XML DB--Uniting Content and Data, Oracle, http://www.oracle.com/features/9i/index.html?xml_feature.html, Retrieved on 20.8.2002.

Pardede, E. - Rahayu, J. W. - Taniar, D. (2004). On using collection for aggregation and association relationships in XML object-relational storage. In *Proceedings of the Symposium on Applied Computing*, Nicosia, Cyprus, ACM Press, 703 - 710.

Parent, C. - Spaccapietra, S. - Zimányi, E. (1999). Spatio-temporal conceptual models: data structures + space + time. In *Proceedings of the ACM international symposium on Advances in geographic information systems*, Kansas City, Missouri, United States, ACM Press, 26 - 33.

Patel, A. - Sim, M. - Weber, R. (1998). Stakeholder experiences with conceptual modeling: an empirical investigation. In *Proceedings of the international conference on Information systems*, Helsinki, Finland, Association for Information Systems.

Premkumar, G. - Potter, M. (1995). Adoption of computer aided software engineering (CASE) technology: an innovation perspective. *ACM SIGMIS Database*, 26(2-3): 105 - 124.

Ram, S. - Ramesh, V. (1998). Collaborative conceptual schema design: a process model and prototype system. *ACM Transactions on Information Systems (TOIS)*, 16(4): 347 - 371.

Ranganathan, C. - Sethi, V. (2000). External IT environment: dimensionality and measurement. In *Proceedings of the International Conference on Information Systems*, Brisbane, Queensland, Australia, ACM, 594 - 600.

Renner, A. (2001). XML data and object databases: the perfect couple? In *Proceedings of the 17th International Conference on Data Engineering*, Heidelberg, Germany, 143 - 148.

Rigaux, P. - Scholl, M. - Voisard, A. (2002). *Spatial Databases with application to GIS*. Morgan Kaufmann Publishers: San Francisco.

Roddick, J. F. - Al-Jadir, L. - Bertossi, L. - Dumas, M. - Estrella, F. - Gregersen, H. - Hornsby, K. - Lufter, J. - Mandreoli, F. - Männistö, T. - Mayol, E. - Wedemeijer, L. (2000). Evolution and change in data management — issues and directions. *ACM SIGMOD Record*, 29(1): 21 - 25.

Rogers, E. M. (1995). *Diffusion of innovations*, 4th Ed. The Free Press: New York.

Rogers, E. M. (2003). *Diffusion of innovations*, 5th Ed. The Free Press: New York.

Rosenthal, A. - Seligman, L. (2001). XML's impact on databases and data sharing. *Computer,* 6(34): 59 - 67.

Sanglu, L. - Ziaofang, Z. - Xiayo, W. (2000). Spatiotemporal data modeling and management: a survey. In *Proceedings of the Technology of Object-Oriented Languages and Systems*, Xi'an China, IEEE, 202 - 211.

Scannell, E. - Sullivan, T. (2001). Tracking down the XML reward. *InfoWorld*, 23(48): 17 - 18.

Scardina, M. V. (2000). *XML Support in Oracle 9i*. An Oracle White Paper. Oracle Corporation.

Schmauch, C. - Fellhauer, T. (2003). Comparison of Database Approaches for Storing XML Documents. In *XML Data Management - Native XML and XML-enabled database systems*, ed. by A. B. Chaudhri - A. Rashid - R. Zicari**:** 519 - 546. Addison-Wesley Boston.

Schöning, H. (2001). Tamino - a DBMS Designed for XML. In *Proceedings of the 17th International Conference on Data Engineering*, Heidelberg, Germany, 149 - 154.

Shanks, G. (1997). The challenges of strategic data planning in practice: an interpretive case study. *Journal of Strategic Information Systems,* 6: 69 - 90.

SOK-Yhtymä (2003). *Vuosikertomus 2002*. Suomen Osuuskauppojen Keskuskunta: Helsinki.

SSP (2001). *Vuosikertomus 2001*. Salon Seudun Puhelin Oy: Salo.

Stal, M. (2002). Web services: beyond component-based computing. *Communications of the ACM*, 45(10): 71 - 76.

Stonebraker, M. (1998). Betting on ORDBMS. *Byte,* 23(4): 91 - 95.

Sutherland, J. - van den Heuvel, W.-J. (2002). Enterprise application integration and complex adaptive systems. *Communications of the ACM,* 45(10): 59 - 64.

Swanson, E. B. (1994). Information Systems Innovation Among Organizations. *Management Science,* 40(9): 1069 - 1092.

Swanson, E. B. (1999). IS “maintainability”: should it reduce the maintenance effort? *ACM SIGMIS Database*, 30(1): 65 - 76.

Swanson, E. B. - Burton, E. (2000). System Life Expectancy and the Maintenance Effort: Exploring Their Equilibration. *MIS Quarterly*, 24(2): 277 - 297.

Talentum (2004), 500 suurinta, Talouselämä, http://www.talouselama.fi/te500list.te, Retrieved on 8.7.2004.

Tehrani, N. (2003). *Custom-built CRM applications*. Customer Inter@ction Solutions, 21**:** 2 - 6.

Tian, F. - DeWitt, D., J. - Chen, J. - Zhang, C. (2002). Data management issues in electronic commerce: The design and performance evaluation of alternative XML storage strategies. *ACM SIGMOD Record*, 31(1): 5 - 10.

Tilastokeskus (2003). *Vuosikertomus 2002*. Tilastokeskus.

Tillet, L. S. (2001). XML Databases Gain Momentum. *Internetweek*, 860.

Treese, W. (2002). XML, web services, and XML. *netWorker*, 6(3): 9 - 12.

TS-Yhtymä Oy (2003). *Toimintakertomus 2002*: Turku.

United Nations (2003). *E-commerce and development report 2003*: New York and Geneva.

W3C (2004a), Extensible Markup Language (XML) 1.0 (Third Edition) - W3C Recommendation 04 February 2004, http://www.w3.org/TR/REC-xml/, Retrieved on 21.8.2004.

W3C (2004b), XML Query (XQuery), Marchiori, M., Retrieved on 3.5.2004.

Walsham, G. (1993). *Interpreting Information Systems in Organizations*. Wiley: Chichester.

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems*, 4: 74 - 81.

Varlamis, I. - Vazirgiannis, M. (2001). Bridging XML-Schema and relational databases. A system for generating and manipulating relational databases using valid XML documents. In *Proceedings of the ACM Symposium on Document Engineering*, Atlanta, Georgia, USA, ACM Press, 105 - 114.

Wedemeijer, L. (2001). Engineering for conceptual schema flexibility. In *Proceedings of the Research Issues in Data Engineering*, Heidelberg, Germany, IEEE, 85 - 92.

Yin, R. K. (1994). *Case Study Research - Design and Methods*, 2 Ed. SAGE Publications, Inc.: Thousand Oaks.

Zaino, J. (2003). *Fleet Sees Payoff in CRM*. Information Week**:** 47.

Zaltman, G. - Duncan, R. - Holbek, J. (1973). *Innovations and organizations*. John Wiley & Sons: New York.

Östensen, O. (2001). The expanding agenda of Geographic information standards. *ISO Bulletin*, (July): 16 - 21.

APPENDIXES

Appendix 1. Presentation of the cases

SOK Corporation

The S Group has organized itself into a co-operation network comprising regional cooperative societies and the SOK Corporation. SOK’s task within the S Group is to help the cooperative societies produce services and benefits competitively for their customer-owners by focusing on providing support and procurement services for all of the S Group’s companies. (SOK-Yhtymä 2003)

The focus for this research within the SOK Corporation was to study the prerequisites and the need for designing an enterprise level data model. SOK’s IT management aimed to develop the utilization of stored data and create common data stores. The enterprise data model would help to recognise points of integration and would also clarify responsibilities.

In this case the research was divided in three stages: a) study the prerequisites and needs of an enterprise data model, b) study the data managed in the SOK Corporation and c) design the enterprise data model. During the interviews (Table A.2) many information systems were studied. The original Finnish position terms are translated to English for this thesis. The essential information systems are listed in
Table A.1.

Table A.1: Essential information systems in SOK Corporation.

Information system	Function	Implemented, updated
REX-Enterprise resource planning	ERP of Sokos chain management unit	1995
Realisti	IS for managing the real estates of S-group	2000
Akseli	Retail sale IS of Maan Auto	1991
AOP	IS for managing customer information of S-Group	1990, 1996
Kauha	IS for managing season products	1994
Intro	ERP of Intrade Partners Ltd	
CRS/CIS	Central room reservation IS of Hotel chain	1998, 2002
Fenix	Personnel administration IS of S-group	2001
Tuhti	Assortment control IS of market chain management	2002

Table A.2: Interviews in SOK Corporation.

Date	Time	Interviewee	Position	Unit
3.12.2002	8:00 - 10:00	Ruuskanen Laura	IT and Logistics Manager	Sokos-ketju
		Mattila Jorma	IS Manager	Sokos-ketju
3.12.2002	10:00 - 11:30	Blomberg Lea	IS Manager	SOK Kiinteistötoiminnot
3.12.2002	12:30 - 14:30	Majamäki Ilkka	IT manager	Maan Auto
9.12.2002	8:00 - 10:00	Muhonen Jari	IT manager	SOK TalousUnit
9.12.2002	10:00 - 11:35	Koskinen Pekka	IS manager	ABC-Ketju
9.12.2002	13:00 - 15:00	Kokkila Jari	Head of technology	Agrimarket
		Kankaala Harri	IT Manager	Agrimarket
11.12.2002	12:00 - 13:35	Seitsamo Jari	IT Manager	SOK Asiakasomistajapalvelut
11.12.2002	14:00 - 16:30	Nikkinen Ari	IT Manager	Intrade Partners Oy
13.12.2002	10:10 - 12:10	Vihavainen Sirkku	IT Manager	Sokotel
17.12.2002	8:10 - 9:35	Hentunen Irma	Head of salary office	SOK Henkilöstöpalvelut
17.12.2002	10:30 - 14:00	Simola Irja	IT Manager	Market-liiketoiminta
17.12.2002	15:30 - 16:30	Ehanti Seija	Head of finance	SOK Rahoitus

Salon Seudun Puhelin Ltd

The vision of Salon Seudun Puhelin Ltd is to be the closest, most reliable, best quality and most useful partner in telecommunications for the Salo region now and in the future. The role of the company is to build and develop networks for telecommunications and services for the people, companies and communities of Salo region. The focus areas of the company are:

- Broadband telephone network and services.
- Mobile network and services.
- Regional network.
- Cable television network and digital services.
- Internet portal (Allu).
- Real estate security and control systems. (SSP 2001)

The focus of this research within Salon Seudun Puhelin was to study the present state of its information systems and databases and to identify possible targets for development. In this case the research was divided in two parts. The first part concentrated on studying the essential information systems and databases. The second part looked the same information systems and the databases from a user perspective. The information systems studied in this case are presented in Table A.3 and the arranged interviews are presented in Table A.4, along with a translation of the interviewees' positions.

Table A.3: Studied information systems in Salon Seudun Puhelin Ltd.

Information system	**Function**	**Implemented**
TTM-system	ERP for managing purchase orders, customers, contracts, calls and products	1999
Tcable	Network Planning System	1994
TTM-reporting	Self-made reporting system	2000
Wintime	IS for financial administration	1999

Table A.4: Interviews in Salon Seudun Puhelin Ltd.

Date	Time	Interviewee	Position
5.12.2002	13:00 - 16:00	Pantsu Jari	IS-designer
		Kannisto Jani	IS-designer
		Uusitalo Jaakko	IT Manager
12.12.2002	8:10 - 9:00	Kannisto Jani	IS-designer
		Pantsu Jari	IS-designer
		Uusitalo Jaakko	IT Manager
12.12.2002	9:00 - 10:00	Vienonen Tommi	Main user
		Uusitalo Jaakko	IT Manager
		Pantsu Jari	IS-designer
12.12.2002	10:00 - 12:00	Pantsu Jari	IS-designer
		Kannisto Jani	IS-designer
		Uusitalo Jaakko	IT Manager
9.1.2003	8:40 – 10:00	Vuorio Pertti	CEO
9.1.2003	13:00 - 14:00	Vuorinen Kari	Sales Manager
10.1.2003	13:00 - 14:20	Meller Heikki	Financial Director
16.1.2003	8:05 - 9:30	Iso-Markku Matti	Technical Manager
16.1.2003	9:45 - 11:30	Järvi Pauliina	Head of e-services
16.1.2003	12:15 - 13:30	Marttila Janne	Head of Accounting

Statistics Finland

The mission of Statistics Finland is to combine collected data with its own expertise to produce statistics and information services for the needs of society, promote the use of statistics and develop national official statistics. Statistics Finland operates administratively under the Ministry of Finance, but is fully and independently responsible for its activities, services and statistics. Statistics Finland has personnel of around 1,100, of whom 200 are employed as statistical interviewers. (Tilastokeskus 2003)

The focus of this research in Statistics Finland was to study the possibilities of storing and managing spatial data in a relational-based database management system. In this case the research was divided into three stages. The first stage concentrated on standards of spatial data management. The second stage focused on spatial data management in different database management systems. Finally, the third stage looked at current spatial data management solutions in Statistics Finland and the goal was to design the first version of a data model of spatial data and its connections to other data.

The most important information systems for spatial data management in Statistics Finland are presented in Table A.5 and the arranged interviews are presented in Table A.6.

Table A.5: Essential information systems managing spatial data in Statistics Finland.

Information system	**Function**	**Implemented, updated**
Classification DB	Contains all of the basic classification and specialities of different statistics	1985, 1993
Business and location register	IS for the management of information of companies and organizations.	1997, 1998

Table A.6: Interviews in Statistics Finland.

Date	**Time**	**Interviewee**	**Position**
6.5.2003	13:30 - 14:30	Asp Jouni	Database specialist
		Nieminen Jari	Project Manager
		Starck Christian	Special Researcher
		Leppänen Maija	Spatial data specialist
		Tammilehto-Luode Marja	Development Manager
20.5.2003	10:00 - 11:50	Asp Jouni	Database specialist
20.5.2003	12:05 - 13:30	Leppänen Maija	Spatial data specialist
20.5.2003	14:00 - 15:10	Tammilehto-Luode Marja	Development Manager
20.5.2003	15:20 - 16:30	Kanerva Liisa	Spatial data specialist
26.5.2003	10:00 - 11:50	Starck Christian	Special Researcher
26.5.2003	12:00 - 14:15	Nieminen Jari	Project Manager

State Provincial Office of Western Finland

The State Provincial Office of Western Finland is a joint regional administrative authority of seven ministries. The agency is divided into eight departments. It promotes governmental and regional goals by taking care of duties in:

- judicial administration
- fire and rescue administration
- education and culture administration
- physical education and juvenile administration
- agriculture and forestry administration
- traffic administration
- consumer, competition and groceries administration
- welfare and health administration
- jurisdictional district and register administration.

Regionality of governmental administration means also that every State Provincial Office is responsible for evaluating basic services regionally and locally. (Länsi-Suomen lääninhallituksen peruspalvelujen arviointiryhmä 2002) The State Provincial Office of Western Finland collects a large amount of data to produce the evaluation reports. The management of this data was identified as the area of study, with an aim to design the first version of a data model. The research focused on basic services in public transport and narcotics.

There are no special database solutions for the management of evaluation data. However some operational information systems were exploited during the evaluation process by collecting data from these information systems (see Table A.7). The arranged interviews are presented in Table A.8.

Table A.7: Essential information systems of State Provincial Office of Western Finland used in evaluation of basic services.

Information system	Function	Implemented
Vallu	IS for managing traffic licences	1998
Riki	IS for management of crimes and criminals	1992

Table A.8: Interviews in the State Provincial Office of Western Finland.

Date	**Time**	**Interviewee**	**Position**
13.2.2003	10:15 - 12:35	Lehtonen Jari	Special Designer
14.2.2003	10:30 - 11:45	Mieskolainen Juha	Province’s Welfare inspector
27.3.2003	12:05 - 13:15	Jääskeläinen Antti	Province’s Forensic pathologist
27.3.2003	14:00 - 15:30	Aaramo Sirpa	Inspector of Education and Culture
3.4.2003	12:00 - 13:30	Hurskainen Raija	Development Manager
3.4.2003	13:35 - 15:00	Salmi Reijo	Inspector of Education and Culture
24.4.2003	9:00 - 10:15	Kymäläinen Marja-Leena	Designer
12.5.2003	9:30 - 10:50	Sihvonen Pertti	Province’s Assistant Chief Director of the Police Force
14.5.2003	Email	Tikkala Vesa	Development Manager
9.6.2003	Email	Antero Harri	Special Designer

TS-Group Ltd

The TS Group is a communications group, which processes and provides information, as well as develops, produces and markets graphics products and related services. The group's parent company, TS-Yhtymä Oy, is responsible for subsidiaries, real estates and the administration of the whole group. In addition, it publish newspapers and the main newspaper is Turun Sanomat (TS-Yhtymä Oy 2003).

In TS-Group Ltd the focus of this study was to study how current information systems and databases support the management of customer data. At the same time it was also hoped to identify possible development points. The research was divided in two stages. In first stage the essential information systems and the databases were studied. In second stage the studied infrastructure was analysed from the viewpoint of CRM.

The essential information systems of TS-Group Ltd are presented in Table A.9 and the arranged interviews are listed in Table A.10.

Table A.9: Essential information systems of TS-Group Ltd.

Information system	Function	Implemented
Centralized customer database	Centralized management of customer data	1990
Marja-Order Entry	IS for managing newspaper orders	1990
ATEX-Announcements	IS for managing announcements	1991
GD 2000	ERP for printing houses	1997

Table A.10: Interviews in TS-Group Ltd.

Date	Time	Interviewee	Position
15.5.2003	10:00 - 11:30	Mäenala Pauli	Accounting Manager
22.5.2003	10:30 - 12:00	Pakarinen Paul	Project Manager
22.5.2003	12:05 - 14:30	Määttänen Markku	Project Manager
22.5.2003	8:30 - 10:00	Lehtiö Jari	Project Manager
23.5.2003	10:30 - 12:00	Landström Heidi	Product Manager
23.5.2003	12:30 - 13:45	Laiho Matti	IT Manager
23.5.2003	8:30 - 10:00	Sinisalo Tero	IS designer
27.5.2003	8:30 - 9:30	Pellonperä Teemu	Financial Manager

Optiroc Ltd

Optiroc Ltd manufactures mineral-based materials for both buildings and other civil engineering projects. The main product categories are Kahi-bricks, Leka-products and maybe the most important outputs, dry products like Vetonit and Serpo. Optiroc is part of HBE-group, which is part of Heidelberg Cement group. Organizationally, Optiroc is a very independent unit.

In Optiroc the research focused on studying the present state of its information systems and databases and identifying possible development points. Also, maintenance and customisation issues were defined as important topics for the study.

Optiroc's essential information systems are presented in Table A.11 and the arranged interviews are listed in Table A.12.

Table A.11: Essential information systems of Optiroc Ltd.

Information System	Function	Implemented
Fina	Financial Management IS	1996
Ufo	Order and Sales IS	1998
Rubu	Budget IS	2003
Intex	IS for Internet/Extranet	2003

Table A.12: Interviews in Optiroc Ltd.

Date	Time	Interviewee	Position
25.6.2003	10:00 - 11:30	Koskinen Ari	IT-Manager
25.6.2003	11:40 - 12:50	Kanerva Pekka	Consultant, Softemis Ltd
14.7.2003	9:00 - 10:50	Suckman Pirjo	Accounting and IS Manager
16.7.2003	8:30 - 10:10	Oksanen Tuija	Main user of Order and Sales IS
22.7.2003	9:00 - 11:40	Pöyry Olli	Consultant, Essaim Ltd
8.8.2003	9:00 - 10:45	Fjäder Mikael	Marketing Director
19.8.2003	9:00 - 10:45	Holm Magnus	CEO

Appendix 2. Interview script

General part

1 Line of business

2 Interviewee’s position

3 Key figures
- turnover
- employees

4 Essential databases
- purpose of information system

5 Database administration and management
- Organization in DB administration and management
- Know-how in DB administration
- Costs (licenses, ...)
- Procurement decisions

6 State of and need for enterprise data model

7 What information is used/managed/produced?
- Essential entities

8 What utilized information is received elsewhere?
- From whom?
- What is the format?
- Interface

9 What information could be useful to others?
- Are others already using information from you
- What is the format?
- Interface

10 What information should be received elsewhere?
- From whom and why?

- In what format?
- Interface

11 Others

Information system/database specific part

1 Purpose of the database?
- Information system
- Purpose of use
- DBMS and version
- When implemented?

2 Owner and master of the database?
- Who owns?
- Who is responsible of the maintenance and administration?

3 Structure of the costs?
- Costs of licences?
- Costs of hardware?
- Costs of personnel?

…

4 Interfaces to information system?
- Client/Server?
- Browser?
- Mobile?

5 Key figures of the database?
- Number of tables
- Number of bytes reserved
- Number of rows
- Number of users
- Number of concurrent users
- Number of transactions

6 Documentation?
- Existence?
- Level of documentation?

7 Logical structures used in database?
- Views
- Triggers
- LOB- Type

8 Interfaces to other information systems and databases?
- Stand-alone system or delivering/receiving data to/from somewhere
- Techniques used in data transfer between information systems

9 Problem areas and critical points
- Which problems have occurred?
- Breaks on usability and reasons for the breaks
- Weaknesses of the present database

10 XML?
- Are XML documents produced from the data in database?
- Are XML documents received elsewhere?
- Are XML documents stored in the database? How?
- Level of XML knowledge in the organization

11 What future plans exist?
- Exploiting object-features
- Exploiting XML
- Storing XML
- Web services
- Other