Opportunities and limitations of using SOA concepts and technologies for building BI applications: a Delphi Study

Author: Sefan Linders

Opportunities and limitations of using SOA concepts and technologies for building BI applications
a Delphi Study

Author:	Sefan Linders
Master:	Business Information Technology University of Twente
Committee:	Dr. Roland M. Müller (University of Twente) Dr. Luís Ferreira Pires (University of Twente) Drs. Niels van der Zeyst (Capgemini Netherlands) Drs. Annemarie Grote Gansey (Capgemini Netherlands)
Date:	June 12, 2008

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

University of Twente
Enschede - The Netherlands

Preface

This thesis marks the end of my studies at the Twente University, where I have studied the past several years to receive both a bachelor and master degree in Business Information Technology. It also marks the end of my life as a student, a life that I have appreciated in many ways.

Service-Oriented Architecture and Business Intelligence, the two research areas that are at the basis of this thesis, were rather unknown concepts to me when I started working on this thesis. Over the past ten months, I have gained an considerable amount of knowledge on both of these research areas, and have tried my best to add some knowledge to the scientific community. During the process that led to this thesis, I have been supported and guided by several people who I would like to thank here.

First of all, I would like to express my gratitude to Roland Müller and Luís Ferreira Pires, both associated with the University of Twente. I have been blessed with their excellent guidance during the entire process, providing me with their insights and opinions whenever was in need. I truly believe they greatly improved the quality of my research.

The research has been carried out at Capgemini Netherlands, at the SAP Netweaver & Intelligence practice. Thanks goes out to my supervisor Niels van der Zeyst, with whom I have had many energetic discussions on the direction of my research, and who always found the energy to support me during my downs. I would also like to thank my second supervisor at Capgemini, Annemarie Grote Gansey, for her knowledgeable insights during our meetings, and her enduring faith in a successful completion of this research. I would also like to thank my practice manager Frank Wammes for inviting me to many joyful and interesting activities, where I've had the chance to meet many of my colleagues.

During my full stay at Capgemini, I have been very happy with the companionship of Chris Jager and Annemaaike Daniel, who started on their thesis about the same time as I did. Wonderful enough, we all managed to graduate in the same month! In my opinion, we have had many interesting and helpful discussions about our research, not in the least to help us manage our thesis frustrations. The less serious discussions might have helped even more. Chris, thanks for teaching me some of your advanced juggling skills. Annemaaike, thanks for pointing me –daily- on the awkward behavior of seagulls.

Apart from the people directly involved in the process of my thesis, I would like to thank my family and friends for their support and the many enjoying moments we shared in the past months. Special thanks goes out to my parents for their care, patience, and unwavering support.

Sefan Linders
June 2008

Participating experts

I owe a debt of gratitude to all the experts who participated in the Delphi study, which was so important for this research. All input to the study, as well as all the other feedback I have received, have been deeply appreciated.

Thanks goes out to the following individuals, who each have participated in the Delphi study:

Michael Bienstein, Manager, Bearingpoint, France

Frans Tolen, Senior Managing Consultant, IBM Global Services, The Netherlands

Rik Laurens, Principal Consultant, SAP Architect, Capgemini, The Netherlands

Liya Wu, BI Architect, Hewlett Packard, United States

Richard Veryard, Industry Analist, CBDI, United Kingdom

Hans-J. Lenz, Professor, Free University Berlin, Germany

Oleg Figlin, Senior NetWeaver Solution Architect, SAP, United Kingdom

Rogier Poelmann, Managing BI Consultant, Capgemini, The Netherlands

Stephen Robson, BI Architect, Capgemini, Canada

Bill Baker, Software Development Manager, Microsoft, United States

Remy van der Kleij, Professional Services Consultant, Informatica, The Netherlands

Mike Bergman, Senior Sales Consultant, Business Objects - an SAP company, The Netherlands

Bert Oosterhof, Director of Technology EMEA, Informatica, The Netherlands

Pascal de Haan, Managing Consultant, Capgemini, The Netherlands

Jürgen Kress, SOA Partner Adoption EMEA, Oracle, Germany

Jack Esselink, Principal Solutions Specialist, Cognos – an IBM company, The Netherlands

Edwin Scheepstra, Data Warehouse Architect, Capgemini, The Netherlands

Maurice van Keulen, Assistant Professor, University of Twente, The Netherlands

Andreas Seufert, Professor, Applied University of Ludwigshafen, Germany

Management Summary

Service-Oriented Architecture (SOA) is gaining increasing attention to improve the flexibility and agility of organizations. However, the current focus is mainly on supporting operational systems, not on supporting Business Intelligence (BI) systems. This thesis aims to identify potential benefits or challenges of using SOA for the organization and use of BI. The thesis is guided by the following research question:

"What are the opportunities and limitations of using SOA concepts and technologies for building BI applications?"

To answer this question, we first performed a literature study on the concepts and technologies of both BI and SOA. After that, we applied a Delphi study to a panel of experts in order to generate and discuss opportunities and limitations. In this study, we used models from our literature study for guidance of the experts.

The result of this Delphi study is a ranked list of opportunities and limitations, which include the arguments and comments from the experts. We have further investigated a small subset of these opportunities and limitations, by using the comments from the expert panel, having private interviews with some of our experts, and consulting additional literature.

Literature study on SOA and BI

We identified BI as the process of gathering and analyzing data, and using the produced information to steer the organization. The process consists of four phases, (1) planning and direction, (2) collection of data, (3) analysis of data, and (4) distribution of data. We also identified the areas of the organization for which BI delivers information, and what this information is about.

We defined SOA as an application architecture within which all applications logic is defined as services, which can be called in defined sequences to form business processes. We discussed how services interact and to what principles they should adhere. The main benefits that would justify the use of SOA are improved possibilities for both reuse and integration, which could result in increased agility and adaptability of the organization as a whole.

Opportunities

The majority of the opportunities that have been identified and rated correspond to the reuse and integration benefits that are also associated with SOA. Our main observations on the identified opportunities are as follows:

- Reuse of services for the collection of data and distribution of information.
- Integration with operational systems for the collection of data. SOA is perceived as offering capabilities for better integration of (operational) systems with the BI systems.
- Integration with operational systems for the distribution of information, to improve the use of BI information in operational processes.
- Integration of the components of BI systems through the use of services could result in a more flexible BI architecture.

For the tactical and operational level more opportunities are identified than for the strategic level, probably because BI is most often used at these levels, and because they have a greater need for a flexible BI organization that can adapt to the more frequent changes at these levels. No considerable differences in the value of opportunities for different focus areas have been identified.

We have further investigated Business Activity Monitoring (BAM), which is one of the identified opportunities. BAM supports the steering of operational processes by providing real-time information on the current state of these processes. BAM systems need to collect data from various operational systems, and SOA provides the means for integrating these systems. Part of this support is found in event-based messaging, which enables real-time data collection.

We also further investigated Master Data Management (MDM). MDM serves to maintain a consistent definition of business entities throughout operational systems, and can also store the data of those entities in a central location. MDM can provide consistent and up-to-date data on entities to the BI system, and can improve the consistency of the data analyses. SOA is identified as a suitable delivery system to integrate the operational systems with the MDM system. Furthermore, transformation and cleansing services could be used to build the MDM system.

Limitations

The most serious and most discussed limitation in the Delphi study is the transportation of large data sets over web services. Large sets of data often need to be transported from the operational systems to the data warehouse, for the collection of data for BI. Web services, often employed for the exchange of rather small messages, currently do not seem suitable for transporting large data sets. Although event-based communication is regarded a possibility for using services for the collection of data, we have no knowledge of viable implementations of this concept.

Further research on handling large data sets in an SOA is needed to find viable approaches for realizing actual benefits of handling this data for the collection of data for BI.

Conclusion

Overall we regard the approach of this research a success. Many opportunities and limitations of using SOA concepts and technologies for building BI-applications have been identified, and some of them indicate several benefits for organizations. This research has touched on many subjects concerning SOA and BI that are only abstractly mentioned, and therefore provides a starting point for further research into those subjects.

Table of contents

1 Introduction

Companies experience an increasing need to obtain information about their internal organization and environment, for supporting the decision making processes necessary for steering the organization. Business Intelligence (BI) provides support for delivering this information, by gathering and analyzing data available in- and outside the organization. To retrieve the needed data, BI systems interact with the internal systems that provide operational information, the so-called operational systems.

In recent years, Service-Oriented Architecture (SOA) is gaining considerable attention in practice and research. The main promise of this architecture is increased flexibility and agility of the organization, while at the same time lowering the costs of IT. Although limited research on the actual realization of SOA and its promises has yet been performed, many organizations are currently adopting the concepts and technologies of SOA.

However, little is known in theory and practice about how SOA could support the organization and use of BI. Now that the diffusion of SOA into organizations is taking place, one could wonder if this presents potential benefits or challenges for the organization and use of BI systems, or if service-orientation principles can be smoothly applied to BI architectures as well.

This thesis aims at providing insight into the relation between SOA and BI, by identifying the opportunities and limitations of building BI applications using SOA concepts and technologies.

1.1 Objectives

The main goal of this thesis is to provide an overview of the opportunities and limitations for building BI applications using SOA concepts and technologies. An opportunity is defined as a concept or pattern that involves the use of SOA and can improve the organization (design, development, maintenance) and use of BI. A limitation is defined as a constraint of using SOA for the organization and use of BI.

Since the academic literature on this subject is rather limited, this research is meant to be a useful source of information to those practitioners and academics working and learning in the field of BI. Organizations using BI, and vendor organizations developing products for BI, could benefit from the information this research has produced.

1.2 Research problem

This thesis is guided by a main research question, which has been formulated as follows:

What are the opportunities and limitations of using SOA concepts and technologies for building BI applications?

To answer the main research question, a number of sub-questions have been formulated as follows:

1. What are the concepts and technologies of Business Intelligence?

BI has a broad perspective on the organization: information is produced for many different persons steering a part of the organization. Furthermore, the process of delivering BI consists of a number of steps. To classify the different areas of BI, the following questions are formulated:

a. What are the main steps of a BI process?

b. What are the different BI functions?

When describing new ways of organizing BI, we should analyze the traditional way of organizing BI first, to see how these can be replaced by or combined with new technologies and concepts. Therefore we have formulated the following question:

c. What constitutes the common architecture of traditional BI systems?

To gain insight in SOA, we have formulated the following questions:

2. What are the concepts and technologies of a Service-Oriented Architecture?
3. What are the benefits of using a Service-Oriented Architecture?

The main research question has been answered by answering the following questions:

4. Which opportunities exist for using SOA concepts and technologies for building BI applications?
5. Which limitations exist for using SOA concepts and technologies for building BI applications?
6. How do these opportunities and limitations rank in importance?
7. How can the opportunities be technically realized?

1.3 Scope

This research covers the whole BI area within an organization. However, we have limited the research to a technology perspective, and have not investigated organizational consequences of using SOA concepts and technologies.

1.4 Approach

The approach taken to answer the research questions is depicted by Figure 1. Research question 1 has been answered by performing several literature studies. For describing the BI concepts and technologies, the book "The organization of Business Intelligence" by Den Hamer (2005) has been guiding in this thesis. This book presents a BI function model and describes the main steps of the BI process. For details on the architecture of traditional BI systems, we have used Kimball & Ross (2002) and Inmon (2002), which focus on the architecture of BI systems.

The concepts and technologies of SOA have been reviewed by performing a literature study. Guiding in this thesis is the book "Service-Oriented Architecture" by Erl (2005), which provides detailed information on both the concepts and technologies of SOA, and on the SOA benefits with respect to traditional architectures.

Research questions 4-6 have been answered through empirical research. For identifying opportunities and limitations to build BI applications using SOA concepts and technologies, we have invited a number of experts that have knowledge of SOA and/or BI, to take place in a discussion panel. This panel of experts has been guided to propose and discuss opportunities and limitations for using SOA concepts and technologies for building BI applications. This discussion has been organized in the form of a Delphi study (Linstone & Turoff, 1975), which is a method for structuring a group discussion. The discussions have taken place through questionnaires. The preceding literature study provided us with appropriate models to guide the group discussion.

After completing the Delphi, we have selected two opportunities and one limitation for further investigation. The selection of these subjects has been based on the amount of input provided by the expert panel during the Delphi study, the amount of information available in literature, and our observation that these three subjects are related to several other identified opportunities. Our investigation of only these items implies that research question 7 has been answered only for a small subset of the total set of identified opportunities and limitations. Researching all of them would not have been feasible within the time frame of this research.

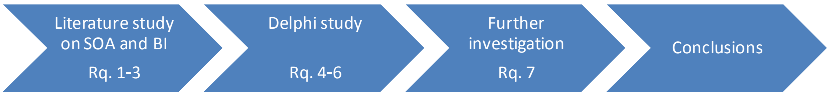


Figure 1: Approach overview

1.5 Structure

The structure of this thesis reflects the order in which the research questions are formulated. The thesis is structured as follows:

- Chapter 2 gives an overview of the concepts and technologies of Business Intelligence. The need for BI is described in the context of organizational steering. The BI process is described, the BI functions model is presented, and the architecture of traditional BI systems is described.
- Chapter 3 presents an overview of the concepts and technologies of the Service-Oriented Architecture. The concept of a service is described, which is the basic component of an SOA. Furthermore we discuss how services are composed to support a business process, what benefits can be realized by adopting SOA, and the most popular set of technologies for implementing SOA.
- Chapter 4 presents the details of the Delphi study, which we used to structure the group discussion between the experts. The chapter describes the Delphi study and its validity for this research, the basic criteria that apply to the selection of experts, and the statistical methods that have been employed.
- Chapter 5 describes the process of the Delphi study as it has taken place. The goals, structure and general outcome of each Delphi round are also discussed.
- Chapter 6 presents the results of the Delphi study. An overview is given of all identified opportunities and limitations, which are categorized according to the models that have been used throughout the study. Statistical analyses of the significance of the results of the study are also discussed.
- Chapter 7 discusses a selection of the results in more detail. Two opportunities and one limitation are further investigated by consulting literature and a small number of interviews with some experts.
- Chapter 8 presents the final conclusion, recommendations, and topics for future research.

2 Business Intelligence

Information is needed to gain insight in the state of an organization and the environment of an organization. The availability of the right information at the right time allows people in making decisions, to steer their organization in a proper way at the right time. Business Intelligence supports the information needs for steering organizations.

This chapter discusses Business Intelligence. Section 2.1 defines Business Intelligence. Section 2.2 discusses why information is important for steering organizations, and explains the support of BI to produce this information. Section 2.3 discusses the BI-cycle, which constitutes the main steps of the BI process. Section 2.4 introduces the BI functions, which cover the different kinds of information for different people of the organization. Section 2.5 presents the architecture of traditional BI systems, after which Section 2.6 presents the conclusion of this chapter.

2.1 Definition

In BI, information is produced by analyzing data, which is often stored in the operational systems of the organization. There is a clear distinction between data and information. Within the context of decision-making, we define *data* as raw, not interpreted facts that have no meaning in themselves. Information is produced when data is interpreted, by placing the data in a certain context, filtering it on relevancy, and correlating it with other data (Agnar & Nygard, 1995; Philips & Vriens, 1999; Den Hamer, 2005). For example, when a store manager needs an overview of the total sales per week over the last year, the data consists of all the individual customer transactions at the counter of that store. By analyzing all these transactions on their date and price, the information of total sales per week can be produced.

In most definitions, BI is seen as a process in which data is used to produce information (Philips, 2004). In this research, we use the following definition, based on the definition of Den Hamer (2005):

“Business Intelligence denotes the process, supported by corresponding facilities, of gathering and analyzing data, and using the produced information to steer the organization.”

In this chapter we describe the BI process for producing information, as well as the different kinds of information that can result from this process. BI covers the production of many different kinds of information, meant to serve many different people. When there is information available that supports making decisions, this information is called “intelligence”. This explains the term Business Intelligence (BI), which denotes intelligence for the business (Philips & Vriens, 1999).

2.2 *Steering of organizations*

As BI is meant to deliver information, it is necessary to explain *why* information is so important for steering an organization. De Leeuw (1982) defines *steering* as *any form of directed influence*, like changing, educating, automating, motivating, and managing. Effective steering is required for an organization to properly adapt to the environment in which it operates. Figure 2 depicts a basic model of a steering situation, consisting of a steered system, a steering entity, and the environment. The steering entity and steered system are abstractions that can correspond to any steering situation in an organization. The steering entity receives information from the steered system and its environment, and influences the steered system by taking steering actions. While the environment can also directly influence the steered system, the steering entity, as well as the steered system, can influence the environment.

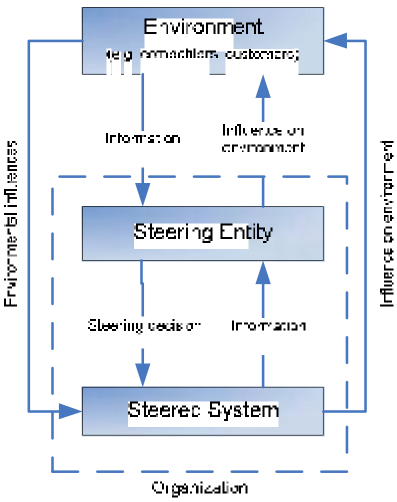


Figure 2: Steering of organizations (De Leeuw, 1982)

De Leeuw (1982) states the importance of information for effective steering. Among others, *sufficient information concerning the state of the environment and the steered system* is needed, because the need for changes should be identified, and because the future outcome of a steering action depends on the current state of the system and the environment. Also needed is sufficient *information processing capacity* to have the right information on time, namely before the steering action needs to be taken.

So, to be able to take the appropriate steering decisions at the right time, enough information processing capacity should be available. However, due to growing pressures of competition, the time to take steering actions is decreasing; organizations have to adapt increasingly faster to the changing environment. At the same time, the amount of data that needs to be processed is increasing, caused by the increasing amount of data that resides in the organization. Therefore, the needed information processing capacity is increasing. An *information gap* emerges as soon as this capacity is too low to deliver the required information at the right time (Den Hamer, 2005).

Consequently, to close or prevent an information gap, the information processing capacity needs to be extended. By efficiently structuring and automating the process for producing information, the time to deliver the right information can be reduced. BI provides the means to achieve this, leaving room for an earlier and better informed steering decision.

2.3 BI-cycle

To produce information out of data, a number of steps need to be taken that apply to all different forms of BI. Several authors describe these steps in a *BI-cycle (*Figure 3), which consists of the main steps taken in the production of information (Bernhardt, 1994; Den Hamer, 2005; Philips & Vriens, 1999; Kahaner, 1996).

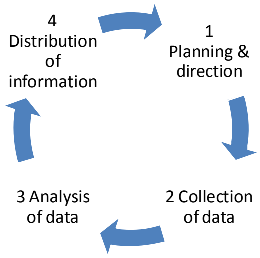


Figure 3: BI-cycle

The cycle consists of four phases:

1. *Planning and direction.* The first phase consists of designing the whole cycle. First the needed information is determined, then the needed data to produce this information is identified (Philips & Vriens, 1999; Kahaner, 1996);
2. *Collection of data.* The second phase consists of identifying the needed sources of data, after which the data can be stored or distributed in a structured format that is suitable for analysis (Den Hamer, 2005). For example, when data is collected from multiple sources, the data needs to be converted to a common structure in order to be combined;
3. *Analysis of data.* In the third phase information is produced, by providing meaning to the data, e.g., by searching for patterns and connections in the data collection, and aggregating data (Den Hamer, 2005);
4. *Distribution of information.* In the fourth phase, the produced information is distributed to the right people or departments in the organization, presented in a suitable format (Philips & Vriens, 1999).

These four phases can be repeated continuously. In the first phase of each new cycle the success of the previous cycle can be evaluated, and taken into consideration when designing the new cycle. The planning and direction phase does not have to be repeated each time; phase 2-4 can be repeated over and over again, resulting in a continuous and automated process of collecting and analyzing data, and distributing the produced information. Some authors (Den Hamer, 2005; Bernhardt, 1994) mention only these last three phases in their BI-cycle, and ignore the planning and direction phase.

2.4 BI functions

There are many different situations in organizations in which decisions have to be made requiring different kinds of information in different formats. To structure the broad perspective on BI, we use the BI function model of Den Hamer (2005), which covers the generally known variations of BI. The model, which is depicted in Figure 4, aligns the BI functions to two axes: the organizational layers and the focus areas. The organizational layers consist of the strategic, tactical and operational layer. The focus areas consist of the internal organization and its environment: the consumers, the suppliers, the market and the competitors of the organization. The internal organization consists of different people and processes, residing in different parts of the organization. The internal organization consists of several steered systems and steering entities.

The steering entity needs information on the state of the steered system, which can be any part of the internal organization or the environment, to support the decision making process. We discuss below the BI functions by describing the information they produce and deliver.

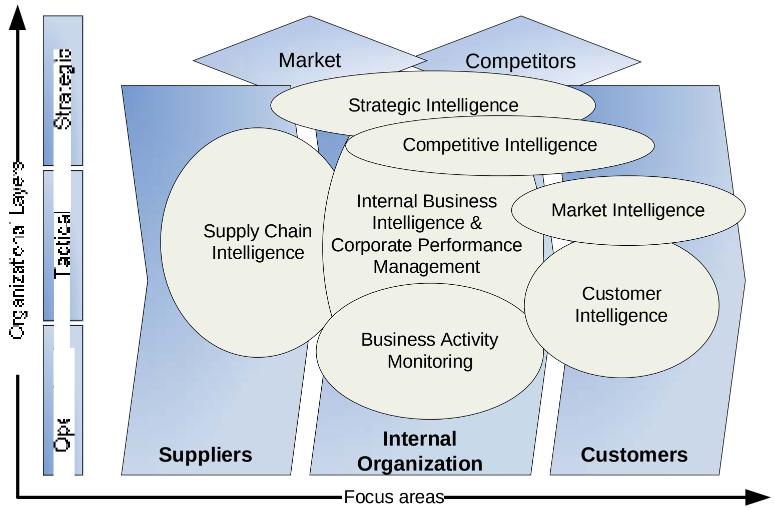


Figure 4: BI Function Model (Den Hamer, 2005)

Internal BI, *Corporate Performance Management,* and *Business Activity Monitoring* are BI functions that focus on information about the internal organization. The currently most used function of BI is *Internal BI*, which can serve all organizational levels, but most often serves the tactical level, by providing management information for steering operational processes. This often happens on weekly or monthly basis. *Business Activity Monitoring* (BAM) provides (near) real-time information on the state of the operational process. This enables timely steering actions by the people controlling the operational process directly and informs higher management of the most current state of the operational processes. *Corporate Performance Management* (CPM) focuses on managing the performance of the organization as a whole, using performance measures that follow from the strategy, to monitor the performance of the whole company in the context of the strategy (Melchert et. al. 2004).

Customer Intelligence and *Market Intelligence* focus on information about (groups of) customers and markets respectively. This information supports decisions about the sales of products, customer retention and personalized marketing efforts. *Market Intelligence* focuses on providing information from a general market perspective, e.g., on certain customer groups, regions or market developments. *Customer Intelligence,* also known as *analytical CRM,* focuses on the individual customer, e.g., by developing a profile of each individual, slicing through all communication channels and products that binds the customer to the organization.

Competitive Intelligence and *Strategic Intelligence* focus mainly on information about the world outside the company, and is used in making strategic decisions. *Competitive Intelligence* informs about competitors and markets, and *Strategic Intelligence* extends this by providing multiple year overview on, e.g., macro-economic trends, technological developments, and government policies. Both BI functions serve to enable fast response on competitors, market developments, and social trends in the environment in which an organization operates.

Supply Chain Intelligence focuses on information about purchasing, logistics, and inventories. This information supports efficient planning of logistics and production processes, and minimization of inventories. The information often comes *from* suppliers, or is provided *to* suppliers.

2.5 Traditional BI Systems

The data needed for producing information is often stored in several *operational systems,* which are transactional systems that are used by the operational business. Queries against these systems are often narrow, one-record-at-a-time queries that are part of the normal transaction flow. The operational systems generally maintain little historical data, and the data format and contents are often optimized for performing transactions (Kimball & Ross, 2002).

The *data warehouse* is often the heart of the BI infrastructure (Den Hamer, 2005). A data warehouse can be described as a database that contains the (historical) data of one or more operational systems or systems external to the organization, in which the combined data serves solely the analytical processes. In this section we further discuss the traditional setup of BI systems, which center around the data warehouse. Kimball & Ross (2002) describe the main components of a traditional BI system, which are the operational systems, the data staging area, the data presentation area, and the data access tools, as displayed in Figure 5. Each of these components is discussed below.

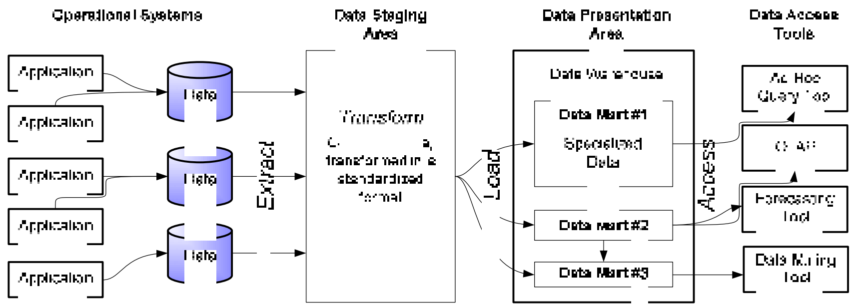


Figure 5: Basic elements of a BI system (derived from Kimball & Ross, 2002)

The operational systems support the operational business, and consist of the applications and their databases. The main priorities of these systems are processing performance and availability for the operational business. Getting the data out of those systems for BI purposes should therefore not reduce the performance of these systems in case they are needed for transactional purposes by the business.

The data staging area consists of both a storage area and a set of processes that collect data from the operational systems and store it in a common format. These processes constitute three main steps, known as *extract-transform-load* (ETL): first data is extracted from the various operational source systems into the *data staging area,* then this data is transformed into a common format, after which the data can be combined, and finally, the data is loaded into the *data presentation area*.

The data presentation area is where data is organized, stored, and made available for access from the data access tools. The data presentation area consists of *data marts*, which contain specialized data, designed for a specific group of users (Den Hamer, 2005). In the data warehouse community, there are two different views on how the data presentation area should be organized. Figure 6 depicts the structure of the data presentation area proposed in Inmon (2002) versus the structure proposed in Kimball and Ross (2002). The Inmon approach consists of an integrated database containing all data, which then serves as a source for *independent* data marts. The Kimball & Ross approach consists of a series of *integrated* data marts, which are directly fed from the data staging area. Both approaches offer a viable alternative for modeling a data warehouse. The most appropriate choice, or even mixture of these two models, depends on the needs of the organization (Jukic, 2006).

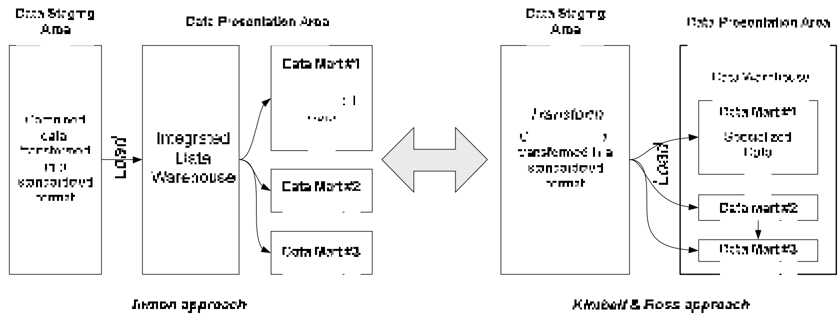


Figure 6: The Data presentation area - Inmon (2002) vs. Kimball & Ross (2002)

Data access tools are the applications that actually use the data that is in the presentation area to deliver the information for which the system is built. Examples of these are reporting tools, which process standardized reports, forecasting tools, which use the data to identify trends, or Online Analytical Processing (OLAP), which can quickly provide answers to analytical questions.

2.6 Conclusion

For people steering an organization, it is important to have information about the internal organization and the environment. BI supports the process of producing this information at the appropriate time. This process of consists of planning the production process, collecting the needed data, analyzing this data to produce information, and distributing it to the right people. BI supports all parts of the organization, at the operational, tactical and strategic level, providing information on the internal organization and its environment, like suppliers, customers, and competitors. The traditional approach to BI systems is centered around a data warehouse, in which all data from operational systems is collected, after which the combined data can be analyzed to produce information.

3 Service Oriented Architecture

Over the past four decades IT systems have grown exponentially, increasing the complexity of software architectures that companies have to manage. The software industry has gone through multiple computing architectures, resulting in reduced implementation time and increased integration possibilities (Channabasavaiah, 2004).

Service-Oriented Architecture (SOA) is currently promoted as the next step to support IT organizations to meet their increasing complex challenges (Channabasavaiah, 2004). Through applying service-orientation principles to the IT organization, the alignment of IT to the business can be improved, and the adaptability and agility of the organization as a whole can be improved. This chapter introduces the commonly identified concepts and technologies of SOA, and discusses its potential benefits for organizations.

In Section 3.1, we first discuss some common terminology used throughout this chapter. Section 3.2 discusses the concept of a service, which is the basic components of an SOA. In Section 3.3, the design principles of SOA are described. Section 3.4 discusses how application logic can be composed using services. Section 3.6 introduces web services, which consist of a collection of technologies popular for implementing an SOA. In Section 3.7, the benefits of SOA are discussed, and finally Section 3.8 contains the conclusion of this chapter.

3.1 Definition

The collective logic that defines and drives an organization is referred to by Erl (2005) as *enterprise logic*. Enterprise logic is an entity constantly changing in response to external and internal influences. From an IT perspective, this enterprise logic can be divided into *business logic* and *application logic*. According to Erl (2005), business logic is generally structured into processes that express requirements of the business, along with any associated constraints, dependencies, and outside influences. Application logic is an automated implementation of business logic into various technology solutions. Through purchased or custom-developed systems, the application logic supports the business processes within the constraints of technology and the organization.

Application logic is often divided into several *applications*. According to Pressman (2001), an application is a specifically designed set of elements of application logic, organized to reach a certain goal by processing information. Ideally such an application relates to a specific process or part of the organization.

An *application architecture* is the standardized definition of a baseline application that can act as a template for other applications (Erl, 2005). It describes the technology, boundaries, rules, limitations, and design characteristics that apply to all solutions based on this template. An organization can have several application architectures that represent distinct solution environments.

SOA applies to application architecture, or to the collection of application architectures of an organization. Channabasavaiah et al. (2004) defines SOA as follows:

"SOA is an application architecture within which all application logic is defined as services, which can be called in defined sequences to form business processes".

3.2 Services

An important feature of service-orientation, as stated by Erl (2005), is the approach for separating concerns, in the context of decomposing large problems into a collection of smaller, more manageable pieces. Each piece addresses a specific part of the larger problem. From an architectural perspective, this means that a system is decomposed into smaller units of logic. Within an SOA, these units of logic are known as *services*, and represent the basic components of the architecture.

Logic is decomposed into different services depending on its context (Erl, 2005). We take a business process as an example for which a system exists that consists of services. As displayed in Figure 7, each service can encapsulate the logic of an individual step of the business process, or of a sub-process comprised of a set of steps. A service can also encapsulate other services, forming one resulting service. By combining services, the whole business process can be supported by application logic.

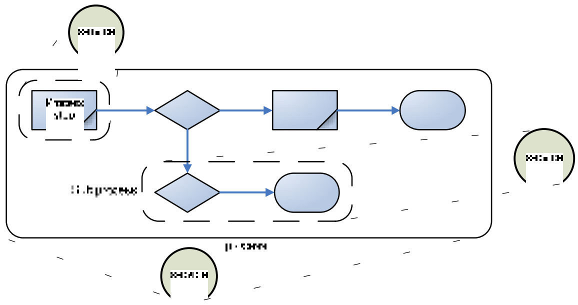


Figure 7: Services can encapsulate varying amounts of logic (Erl, 2005)

Two key roles exist in every relation between services, which are the *service requester* and *service provider* (Colan, 2004b; Papazoglou, 2007a). By sending a message, the service requester invokes the service provider, who processes the request and responds by returning a message. To enable this interaction, services must be aware of each others' characteristics. This awareness is achieved through the use of *service descriptions* (Erl, 2005; Colan, 2004b). Service descriptions contain information such as the service inputs, outputs, and associated semantics. To determine and locate the most suitable service provider to perform the function for the requester, a *service broker* can be used. A service broker is an intermediary between a service provider and a service requester, and maintains an index of available service providers, which includes information like their function, quality and location (Papazoglou 2007a, Colan 2004b). Figure 8 illustrates the relation between the service requester, provider, and broker.

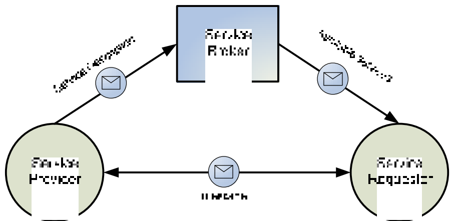


Figure 8: Communication between service provider, requester and broker (Erl, 2005; Papazoglou & Heuvel, 2007)

3.3 Principles

Ranging from public IT organizations to vendors and consulting firms, there are many opinions about what constitutes an SOA. Although an official set of service-orientation principles does not exist (Erl, 2005, p290), there is a common set of principles that is most associated with service-orientation. These principles do not come automatically with the implementation of an SOA. Realization of these principles requires a conscious modeling and design effort (Erl, 2005). We describe the principles below.

Services are autonomous. The range of logic governed by a service exists within an explicit boundary. This allows the service to manage all its processing itself, and makes the service independent from other services. Service autonomy is a primary consideration when deciding how application logic should be divided up into services (Channabasavaiah et al., 2004; Erl, 2005).

Services hide underlying logic. The only part of a service that is visible to the outside world is what is exposed via the service definition. The internal structure of the service is invisible and irrelevant to service requesters. Components using the service should not know or care about the implementation logic of a service, but just want the expected result to be returned (Channabasavaiah et al., 2004; Erl, 2005).

Services are loosely coupled. A service requester should be loosely coupled to a service provider. This means that the service requester has no knowledge of the technical details of the provider's implementation, such as the programming language and the deployment platform. The loose coupling allows the internal structure of requester or provider components to change, without impacting the other, as long as the message format and semantics stay the same. Loose coupling is achieved through the use of service contracts that allow services to interact through messages rather than through the use of API's or file formats (Erl, 2005; Colan, 2004a).

Services are self-contained. Service requesters require a persistent state between service invocations (the service is then *stateful)*, but the service provider should not be required to maintain state information over a longer period of time. This means that service requests should consist of self-contained messages, with all the information the service provider needs to process a request (Erl 2005, p290; Colan 2004a, Papazoglou 2007a).

Services are well-defined (Erl, 2005, p290; Colan, 2004a). A service should have a well-defined interface, described in its service definition. A service definition, also called service contract, provides information on the service endpoint, the operations of the service, and the messages supported by each operation. This information is needed for a service requester to connect to a service provider and invoke the service.

3.4 Service Composition

An important consideration in the design process is which services should be designed, and how these services should be composed to support a business process. In this section we discuss the details of this composition, guided by Figure 9, which depicts the relation between a business process and the automation logic of services.

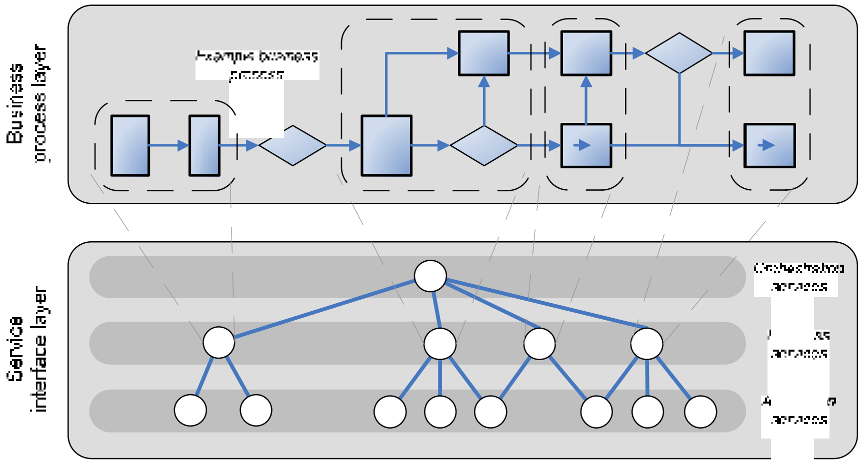


Figure 9: Service orchestration (adapted from Erl (2005) and Channabasavaiah et. al. (2004))

3.4.1 Business process layer

Business logic is a documented implementation of the business requirements that originate from the business areas of an organization (Erl, 2005). Business logic is generally structured into processes that express these requirements. Figure 9 represents the abstraction of a business process layer, in which is a large piece of work is drawn as a composition of smaller units of work (Erl, 2005).

When using service-orientation to support parts of a business process with application logic, it is important to consider how the logic should be supported by different services, as this determines the autonomy of each service. In Section 3.3 we already discussed service autonomy as one of the main principles of service-orientation.

3.4.2 Service interface layer

Figure 9 also depicts the service interface layer, in which the composition of services with respect to a business process is illustrated. The composition can be organized through three layers:

- The application service layer. Services in this layer serve to provide reusable functions that represent common enterprise resources and capabilities (Erl, 2005). For example, they can be used to expose functionality from legacy applications, or build services from scratch that provide new functionality.
- The business service layer. Services in this layer provide functionality closely related to the business process they support. They can use the application services to implement their functions.
- The orchestration service layer. Services in this layer manage how services interact, in which order, and according to which constraints. An orchestration service alleviates the need for other services to manage all their interaction details (Erl, 2005; Papazoglou 2007b).

A benefit of abstracting business and automation logic of services in layers is that a loosely coupled relationship between the business and automation domain can be realized. This allows each domain to evolve independently and adapt to changes imposed by the other (Erl, 2005).

3.4.3 Organizational service-orientation

Some authors (Erl, 2005; Bieberstein et al., 2005; Cherbakov, 2005) argue that besides applying service-orientation to application logic, service-orientation should also be applied to organizational structures. This would result in services in which core tasks and activities are considered as units of business services in which each business service has a unique purpose, and provides one or more services for consumption by other services (Bieberstein et al., 2005). Individual business services can then be orchestrated to form a business process (Bieberstein et al., 2005; Erl, 2005; Cherbakov, 2005).

When an organization consists of business processes composed out of business services, it becomes straightforward to support a business process by mapping its business services onto IT services. This can significantly improve the flexibility and agility with which processes can be remodeled (Erl, 2005).

Although some authors regard organizational service-orientation as a promising concept, in this work we do not further focus on this concept.

3.5 Enterprise Service Bus

To further support integration of services, a connectivity layer can be inserted between service requesters and service providers (Schmidt et al., 2005; Papazoglou 2007a). A version of such a layer often associated with SOA is the Enterprise Service Bus (ESB) (Papazoglou, 2007a). Papazoglou defines the ESB as follows:

"An ESB is an open-standards-based message backbone designed to enable the implementation, deployment, and management of SOA-based solutions."

Papazoglou (2007a) further describes the ESB as a set of infrastructure capabilities implemented to enable an SOA and alleviate disparity problems between applications that run on heterogeneous platforms and use diverse data formats. The ESB promotes loose coupling of systems taking part in integration and can break up integration logic into distinct easily manageable pieces (Papazoglou, 2007a).

The ESB functions as both transport and transformation facilitator to allow distribution of services over disparate systems and computing environments (Papazoglou, 2007b). When messages are transported from service requester to service provider (and vice versa), the ESB is responsible for routing the message to the designated service. The ESB can also provide functionality for the translation of messages to foster interoperability between services, or provide delivery through another protocol.

The ESB also provides the means to enable event-based messaging in an SOA. Examples of events are the arrival of a shipment, the payment of a bill, or an error that occurs in an operational process. An event publisher typically sends a message through the ESB, and the ESB publishes the event notifications as messages to the services that have subscribed to the events (Papazoglou, 2007a). The publisher and consumers of these event notifications can be fully decoupled. The only relationship is indirect, through the ESB, to which requesters and providers are subscribed as subscribers and publishers of events (Papazoglou, 2007a; Keen et al., 2004).

3.6 Web services

For implementing an SOA, a suitable technology is needed that can support the principles of service-orientation. A currently popular collection of technologies that provide this support is ***web services***. However, web services is not a synonym of SOA. Other technologies can be employed as well for implementing an SOA (Channabasavaiah et al., 2004; Papazoglou, 2007a), which is illustrated by the existence of the term "service-oriented" in the software industry long before the introduction of web services (Erl, 2005). However, no technology has been more suitable and successful in manifesting SOA than web services (Erl, 2005; Colan, 2004b)

Web services are based on open standards that are independent from any implementation platform (Papazoglou, 2007a). All major vendor platforms currently support the creation of service-oriented solutions with the understanding that the SOA support provided is based on the use of web services (Erl, 2005).

The web services framework consists of a collection of technologies that apply to the use of services. Below, we introduce the technologies that support the basic SOA concepts we have mentioned earlier:

- The basis for all web services technologies is the Extensible Markup Language (XML), which provides a standard for formatting messages (Moitra and Ganesh, 2005). XML documents are written in plain text, forming a common data representation that can be used as the medium for data exchange between systems from different implementation platforms (Moitra and Ganesh, 2005; Colan, 2004b).
- The Simple Object Access Protocol (SOAP) is a protocol that defines the use of XML-formatted messages for communication between a service requester and service provider. The request is written in XML and transported in a SOAP *envelope* (Colan, 2004b).
- Service definitions are written in the Web Services Description Language (WSDL). A WSDL document contains the details required by the service requester to use a particular service, like a description of the functionality it offers, how it communicates, and where it is accessible.
- As mentioned in Section 3.2, a service broker can serve as an intermediary between the service requester and service provider, by keeping track of published services. The Universal Description, Discovery, and Integration (UDDI) registry is the web services implementation of a service broker.
- Another web service technology often mentioned in relation to SOA, is the Web Services Business Process Execution Language (WS-BPEL) (Erl, 2005). WS-BPEL is an orchestration language that serves to model business processes in terms of composition of web services.
- To support event-based messaging, the ESB supports WS-Notification, providing support for publish/subscribe mechanisms (Papazoglou, 2007a).

3.7 Benefits

Several benefits are mentioned in literature that justify the use of SOA. This section discusses the commonly identified benefits.

3.7.1 Improved integration

When communication standards like web services are applied to the design of services, the functionality of services becomes independent of the implementation platform. This means that all services, although implemented on different platforms, can communicate using the same protocol. In theory, when all services are designed according to the same standards, a service requester from any device, using any operating system, in any programming language, can access another service (Papazoglou, 2007a).

Using standards can therefore result in the creation of solutions that consist of inherently interoperable services. When application logic is represented by standardized services, creating interaction between them requires less effort since the communication proceeds using the same protocol. SOA can therefore significantly reduce the efforts of application integration over traditional methods (OASIS, 2006; Erl, 2005). The benefit of interoperability does not only apply to services that are built from scratch, but also applies to legacy systems. Functionality from these older systems can be exposed by wrapping their functions with standardized services, enabling further integration with other systems.

3.7.2 Reuse

Service-orientation promotes the design of reusable services. Creating a library of services that support reuse, provides increased opportunities for leveraging existing application logic (Erl, 2005; Papazoglou 2007a). When new application logic is built, the time for designing, developing, testing, and deploying the application can be reduced when the required logic is (partly) available in existing services, enabling composition of services, rather than developing all application logic from scratch (Channabasavaiah et al., 2004).

The principles that we discussed drive the opportunities for reuse of services. According to Erl (2005), the principles that should be applied to the design of services foster reuse in the following ways:

- Autonomy and loose coupling of services results in independency of services, which broadens the applicability of its reusable functionality;
- Self-containedness of services maximizes the availability of a service;
- Hiding underlying logic fosters reuse, because service requesters are presented a generic public interface;
- A well-defined service promotes reuse, because it allows (developers of) service requestors to search and discover reusable services.

3.7.3 Agility and adaptability

The agility of an organization represents the speed with which an organization can adapt itself to changes in the environment (Erl, 2005). The agility of an organization depends in part on the agility of the application logic that supports the business processes of the organization. Through the opportunities for both reuse and integration, SOA can increase the ability of the organization to cope with changes (Channabasavaiah et al., 2004). A standardized IT environment comprising composable, interoperable and reusable services establishes a more adaptive organization, in which automation solutions can be delivered faster, with less effort involved (Erl, 2005; OASIS, 2006). Therefore, cost reductions could be realized for building applications.

3.7.4 Validity of benefits

The benefits that an SOA can deliver are promising, but we have to keep in mind that experience with SOA is scarce, and so is the academic research on the results of this experience. Driezen (2008) provides an overview of the benefits attributed to SOA, identified from the relevant articles published in the last ten years in the top-25 information systems (IS) journals as ranked by Mylonopoulos and Theoharakis (2001). Noteworthy is that, according to Driezen (2008), none of the identified benefits in these articles were based on empirical findings.

3.8 Conclusion

We defined SOA as an application architecture within which all applications logic is defined as services, which can be called in defined sequences to form business processes. Services take the role of either requester or provider, and can be defined at application or at business level. We discussed how services interact, and to what principles they should adhere. Enabling the implementation, deployment, and management of services is the Enterprise Service Bus (ESB). Currently, web services are the most popular technology for implementing an SOA. The main benefits that would justify the use of SOA are improved possibilities for both reuse and integration, which could result in increased agility and adaptability of the organization as a whole.

4 Delphi Study

Now that both BI and SOA are introduced, we turn to the main research question which relates these two concepts: ***What are the opportunities and limitations of using SOA concepts and technologies for building BI applications?***

As the previous chapters show, sufficient literature has been published on SOA and BI separately. However, literature on the relation between the two concepts is very limited, and the information that does exist is often restricted to vendor-specific products. Therefore we have decided to gain knowledge from people working with SOA and/or BI, who have the ability to provide us with valuable information on this subject. From now we refer to these people as ***experts***. To collect and analyze information from the experts, we have identified the ***Delphi study*** (Linstone & Turoff, 1975) as the most suitable method.

This chapter discusses the Delphi study in general, and how we have designed a Delphi study that suited our resources and goals. In Section 4.1, we describe the basics of a Delphi study. In Section 4.2, we discuss the validity of using the Delphi study in our research. In Section 4.3, we discuss how we have selected our experts. In Section 4.4, we discuss the communication methods that we employed during the study. In Section 4.5, we discuss the statistical methods that we have applied in order to produce meaningful results. Finally, in Section 4.6, the conclusion is given.

4.1 Basics

A Delphi study enables experts to discuss a complex problem through a structured communication process (Linstone and Turoff, 1975). In this process, individual experts submit their ideas to the organizing researchers, defined as the ***monitor group***, who collate the responses from the whole panel of experts into one overview. In turn, this overview is provided to the experts, who can revise their submitted ideas on basis of opinions of other experts. In the end, this process should produce the collective thought of the group (Linstone & Turoff, 1975). According to Scholl et al. (2004), a Delphi study lends itself especially well to exploratory theory building on complex, interdisciplinary issues, which often involve a number of new or future trends.

A Delphi study constitutes several ***rounds***, as depicted in Figure 10. In each round, a questionnaire is used to collect opinions from the expert panel. The questionnaires are designed to focus on problems, opportunities, solutions, or forecasts. Each subsequent questionnaire is developed based on the results of the previous questionnaire. The process stops when the research question is answered, for example, when consensus is reached, theoretical saturation is achieved, or sufficient information has been exchanged (Skulmoski & Hartman, 2007).

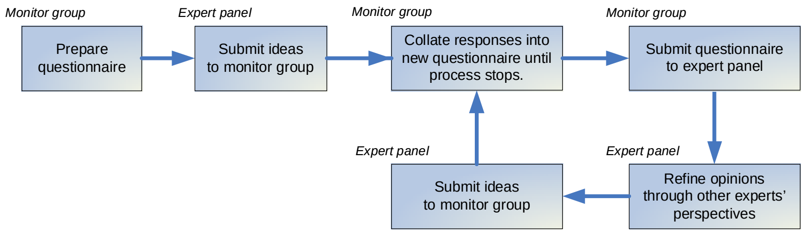


Figure 10: Process of the Delphi study

Rowe and Wright (1999) have identified the four common characteristics of a Delphi study:

1. *Anonymity of Delphi participants.* This allows the experts to freely express their opinions without experiencing social pressures to conform with others in the group. Because of the anonymity of participants, the ideas proposed in a study are evaluated without a bias towards the participant that proposed the idea.
2. *Iteration.* This allows the participants to refine their views in light of the progress of the group's work from round to round.
3. *Controlled feedback.* The participants are informed of the other participant's perspectives, on basis of which they can clarify or change their views.
4. *Statistical aggregation of group response.* This allows for quantitative analysis and interpretation of data.

4.2 Suitability for this research

We already discussed the need for involving experts on SOA and/or BI to provide us with their ideas for opportunities and limitations of SOA concepts and technologies for building BI-applications. We have found the Delphi study as a very suitable approach to this empirical research, because of the interdisciplinary nature and new trends our research concerns. As we expected the participants of this research to have different opinions on the subject and to come up with a small set of different opportunities and limitations, the Delphi study provided us the means to share all these opportunities and limitations with the whole experts panel, and have the participants widen their scope and provide their opinion again. This resulted in an informative and educational process, in which the participants could learn from each other's ideas and evaluate them as a group. By implementing a rating system in the questionnaires, which is common for a Delphi study (Skulmoski & Hartman, 2007), we were able to generate an overall ranking of opportunities and limitations, enabling us to select the ones regarded most important.

4.3 Expert selection

For composing the expert panel, the requirements for selecting experts as described by Adler & Ziglio (1996) have been considered. According to the authors, experts should have:

1. Knowledge of and experience with the issues under investigation;
2. The capacity and willingness to participate;
3. Sufficient time to participate, and
4. Effective communication skills.

Considering the first requirement, experts have been selected who have knowledge of and experience with SOA, BI, or both. Because experts with extensive knowledge of both were difficult to find, some participants mainly had knowledge of BI *or* SOA, but were still regarded to be able to provide useful input from their perspective. To measure the knowledge and experience of the experts, a self assessment section has been added to the questionnaire, which has been derived from a comparable Delphi study by Schmidt et al. (1997). The results of this self-assessment are presented in Chapter 5.

Considering the other three requirements, all experts have been invited on a voluntary basis, which has led to the assumption that all participants had the capacity, willingness, and sufficient time to participate in the study. At the time of invitation, all experts were working for a company or research institute, where the nature of their work required them to communicate. Therefore we have assumed that each participant possessed the required communication skills for this study.

To generate a broad spectrum of opinions throughout the study, we have invited experts working from different perspectives on the relevant concepts and technology. The invited experts were either working for a consulting firm, for a research institute, for a BI-vendor, or for the client company of a consulting firm, that uses BI products in their organization.

For determining the needed panel size in order to produce valid results, we have studied previous articles on Delphi studies. As identified by Skulmoski & Hartman (2007) and Rowe and Wright (1999), there is a large difference in panel size from study to study. A common method for determining a suitable size has not been found. Therefore we have based our needed panel size on a comparable study. Akkermans et al. (2003) organized a Delphi study for identifying supply chain management issues, and the ERP support for these. The panel consisted of 23 experts on supply chain management. The authors reported that at least 20 participants were needed for this Delphi. Therefore we have tried to get at least 20 experts to participate in our study.

4.4 Communication Method

Two main alternatives exist for the setup of a Delphi study. In a ***real-time study***, the experts are gathered in the same location at the same time, enabling a real-time communication process in which the whole study can be carried out in a limited time frame. In an ***asynchronous study***, participants are sent a questionnaire individually, which they can answer at a location and time suitable for them. Our study has been organized as the latter, using the Internet for communication between the experts and the monitor group. This provided us with three advantages over a real-time study:

1. Enough experts could participate, as it would have been difficult to get the required number of participants together for a real-time Delphi study, because of high costs and time constraints involved.
2. Participants were not bound to any geographical region. This enabled us to invite experts from distant locations.
3. We could adjust new questionnaires as a function of the group response without major time constraints (Linstone and Turoff, 1975), because the next round would start when the design of the new questionnaire was finished. In a real-time study, there would be little time to analyze the results from the previous round to change the setup of the next one.

In the first round, in which qualitative answers have been given, the questionnaire was designed as a word processing document. In subsequent rounds, the questionnaire was designed as a spreadsheet, which suited the quantitative questions and answers. The questionnaires have been sent and received by e-mail.

4.5 Data analysis

In the second and third round of our study, each item that experts had provided in the first round has been rated by the whole expert panel. All items of the overview have been rewritten as hypotheses, so that the level of agreement to each statement could be measured. As a rating method, we used a 7-point scale, ranging from ***strongly disagree (-3)*** to ***strongly agree (+3)***. Such a scale is commonly referred to as the Likert scale, and is often employed within a Delphi study (Scholl et al., 2004; Scheibe et al., 1975, Shields et al., 1987; Pérez and Schüler, 1982). Employing this scale enables statistical aggregation of the input of the experts and expressing the group opinion in statistics.

Another popular method for rating a list of items is by ranking them, so that the relative position of one item with respect to the others determines its importance (Scheibe et al., 1975). We have preferred the Likert scale over this method for two reasons. First, ranking a large list of items residing in different categories, which is the case in our study, requires quite some time from the experts. Second, rankings only measure importance of an item with respect to other items. If an item is important in itself may not be identified. In this study, however, this is important to know.

The quantitative group opinion is presented by the central tendency and level of dispersion of the ratings. A high level of dispersion characterizes a low level of consensus. In calculating these two measures, we have considered that Likert scales do not fall within the interval level, but the ordinal level of measurement. Some authors claim that calculating the means and standard deviation is illegitimate, since that would require interval data (Jamieson, 2004; Shields et al., 1987; Scheibe et al., 1975). Their advice is to use the ***median*** as the measure of central tendency, and the ***inter-quartile range*** as the level of dispersion. However, during our analysis we have experienced that many items have the exact same median and inter-quartile range values, making it impossible to distinguish among some of them, while there seems to be a difference in the ratings. Therefore we ***did*** use the mean and standard deviation for our statistical analysis, because they produce more precise results.

We have not found commonly accepted criteria for determining consensus among the participants, when a Likert scale is used. According to Miller (2006), consensus on a topic can be decided if a certain percentage of the votes falls within a prescribed range. One criterion recommends that consensus is achieved by having 80 percent of participants' votes fall within two categories on a seven-point scale (Ulschak, 1983). However, as Shields et al. (1987) and Hsu & Sandford (2007) argue, a common definition of consensus cannot be derived from the literature, as many studies employ arbitrary criteria. We have used the standard deviation for measuring the consensus, so that a high consensus is represented by a low standard deviation. However, we could not continue the study until consensus was reached, for several reasons. First, as we mentioned before, there is no agreed definition of consensus in literature. Second, we did not have enough time to continue the Delphi study until consensus had been reached. Third, this was the most we could do considering the availability of the participants.

4.6 Conclusion

For identifying the opportunities and limitations of using SOA concepts and technologies for building BI applications, we needed an approach to communicate with experts, who have knowledge of SOA and/or BI. The Delphi study provided us with a suitable setup for our empirical research, enabling an anonymous group discussion between the participating experts, split up in several rounds. We considered all invited participants to be suitable experts. The asynchronous version of the Delphi study that we employed increased the possible number of participants, and also provided us with enough time to adjust questionnaires between rounds. The questionnaires of the second and third round contained a rating system, which we used to statistically analyze the results in order to rank the items and identify consensus between the experts.

5 Rounds setup & participation statistics

The objective of our Delphi study was to produce an overview of opportunities and limitations of using SOA concepts and technologies for building BI applications. On basis of the requirements as described in Section 4.3, we have invited a number of candidate experts for participating in this study. All invitees were informed on the goal and structure of the study, and on the estimated time the study would consume. A total of 26 experts agreed to participate. The entire overview of the identified opportunities and limitations can be found in Chapter 6. In this chapter we discuss the setup and results of each round.

Section 5.1 discusses the participation of the experts, Section 5.2, Section 5.3, and Section 5.4 discuss the setup and response of the first, second and third round of this Delphi study, respectively. in respectively. Section 5.5 contains the conclusions. The original questionnaires can be found in Appendices E, F and G.

5.1 Participating experts

Table 1 presents the background and response rates of the experts, which comprise a total of 26 experts that agreed to participate. For the second and third round, the original group of experts was invited, disregarding their participation in the previous rounds.

In the first round, 16 experts participated. There has been a significant decline in the participation rate over subsequent rounds. According to Adler & Ziglio (1996), this is a known phenomenon in Delphi studies. In our study, the short time frame of two weeks in which the questionnaires needed to be answered has sometimes coincided with experts being out of office. This possibly led to a total absence or an increasing busy schedule of the experts during the time frame of a round, making him or her unable to participate in our study. One expert has not taken part in any of the three rounds, but has provided valuable input on some of the subjects in the third round. We have not included this expert in the participation statistics. All rounds together, 18 experts participated in the study.

	Invited	**Round 1 Part.**	**Round 2 Part.**	**Round 3 Part.**
BI-vendors	8	7	5	3
Clients	3	0	0	0
Consultants	9	5	4	5
Researchers	6	4	4	3
Total	*26*	*16*	*13*	*11*
Response rate		*62%*	*50%*	*42%*

Table 1: Response rates

The invited experts reside in different countries. Figure 11 shows the countries of residence of the experts that participated in at least one round. As the figure shows, most of the participants reside in the Netherlands, which is also the residence of the author of this research.

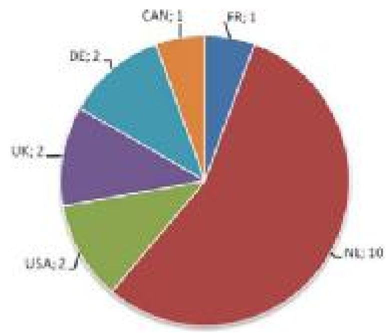


Figure 11: Residing country of participants

To measure the experience of the participants on the topics discussed, we designed a short self-assessment in which the participants were asked for their years of experience with BI and/or SOA, and the number of projects related to BI and/or SOA they had participated in. Table 2 shows the results of this self-assessment. It is clear that most participants have had considerably more experience with BI than with SOA, for both the number of years and number of projects. This is easily explained from the relatively short period of existence of SOA comparison with BI.

BI Experience				**SOA Experience**			
No. Years		***No. Projects***		***No. Years***		***No. Projects***	
0	2	0	3	0	5	0	7
1-5	2	1-10	8	1-2	4	1-3	6
6-10	10	11-20	2	3-5	7	4-10	4
>10	4	>20	5	>6	2	>10	1

Table 2: Experience participants with SOA and BI according to self-assessment

5.2 First round setup & response

In the first round, experts were presented open questions on which opportunities and limitations they could identify for the use of SOA for BI applications. This round can be characterized as a brainstorming session: experts were encouraged to write down any ideas that came to mind. Some structure was in place though. The questions were guided by three BI models: the BI cycle as described in Section 2.3, the BI functions model as described in Section 2.4, and the traditional BI systems architecture model as described in Section 2.5. The traditional BI model was meant to be referred to when identifying an opportunity or limitation related to the traditional BI architecture, but the questions were not based on this model. At the end of the questionnaire a *catch-all question* served to identify further opportunities and limitations that could not be identified based on any of the models.

The questionnaire contained an introduction to the study and to the concepts of SOA and BI, and contained a self-assessment section for measuring the expertise of the participants. The questionnaire was sent out to the 26 experts that had agreed to participate in the study. As displayed in Table 1, 16 experts have sent back a completed questionnaire. The other experts responded that they were too busy to participate, or did not respond at all.

Some experts focused more on certain parts of the questionnaire than others. Also, the answers differed in length. While some participants answered with a few words, others wrote several sentences. We have collated all answers into one overview, which was designed as a list of short descriptions of an opportunity or limitation. Answers that were identical or that largely overlapped have been combined. Answers that were too ambiguous, or simply irrelevant to the study, have been omitted from the overview. The resulting overview contained a total of 51 answers, of which 40 opportunities, 10 limitations, and 1 general remark.

5.3 Second round setup & response

The list of opportunities and limitations that was generated in the first round, has been rated and commented on by the experts in the second round. In the questionnaire, all items of the list were rewritten as hypotheses, for which each expert could select their level of agreement on a seven-point Likert scale. Some hypotheses also included a rating for technical feasibility, for distinguishing the value of the idea from their current implementation feasibility; some opportunities can be valuable, but are difficult to implement using current technology.

The experts were encouraged to provide arguments for their rating, which were used in the subsequent round to inform the group of the different views on each item. New opportunities or limitations could also be added. In case a participant did not comprehend the hypothesis fully, e.g., because he did not have the necessary knowledge on the subject, a *no opinion* option could be selected. This prevented them from choosing “neutral”, which would have corrupted the results.

The questionnaire was sent out to the same 26 participants as in the first round. As displayed in Table 1, 13 experts sent back a completed questionnaire. The other experts responded that they were too busy to participate, or again did not respond at all.

For each hypothesis, the mean, standard deviation, and percentage of “no opinion” has been calculated. Table 3 presents the distribution of the ratings in six levels, each level stating the number of hypotheses that falls within this range. A high mean represents a strong agreement to the hypothesis. Table 4 presents the consensus over all items by means of the standard deviation, split in six levels, each level stating the number of hypotheses that falls within their range of consensus.

	M <0	*0<M ≤0.5*	*0.5<M ≤1.0*	*1.0<M ≤1.5*	*1.5<M ≤2.0*	*2.0<M ≤2.5*
No. Hypothesis’s	2	7	6	11	15	8
Percentage	4%	14%	12%	22%	31%	16%

Table 3: Means (M) of ratings (high M = strongly agree)

	SD <0.75	*0.75<SD <1*	*0.75<SD <1.25*	*1.25<SD <1.5*	*1.5<M <2*	*2<M <2.25*
No. Hypothesis’s	3	11	8	14	10	3
Percentage	6%	22%	16%	29%	20%	6%

Table 4: Consensus in Standard Deviation (SD) (Low SD means high consensus)

We have used the means and standard deviations of the general ratings to select the opportunities and limitations for further research in the third round of the Delphi study. In this process, we have not considered the means and deviations of the technical feasibility ratings, because not all opportunities, and none of the limitations were measured on this variable.

5.4 Third round setup & response

For the third round, we have selected the opportunities and limitations with a high average level of agreement, or with a very low consensus. The opportunities for which the experts already provided us with enough information in the second round were left out of the selection process. Both the items with a high average general rating and with a low consensus were interesting for further research, to determine why some opportunities and limitations received high ratings, and why some the ratings were very different rated.

Thirteen items of the previous round were presented: nine identified opportunities with a high general rating in round 2, three opportunities with a low consensus in round 2, and one limitation with a low consensus in round 2. We also added one opportunity that was not yet rated. A questionnaire with the following information was provided to each participant of the third round:

- The selected hypotheses from round 2;
- For each hypothesis, the average general rating and the standard deviation of the general rating from round 2;
- For each hypothesis, the rating on technical feasibility and the standard deviation of the technical feasibility rating of round 2, if measured in the previous round;
- The rating of this particular participant the questionnaire was sent to, that were provided in round 2, if this expert participated in that round.
- The comments from round 2;

For each item, the experts were asked to provide arguments for their rating, and they were also allowed to adjust their rating, e.g. on basis of new insights provided by comments of others. They were asked why they assigned this specific rating, and which characteristics of SOA were important for this opportunity or limitation.

Again, we sent out the questionnaire to the 26 experts, from which eleven experts sent back the questionnaire, of which most of them provided the arguments for their rating for each item of the questionnaire. The information that they delivered provided us the needed insight for explaining the ratings. This information has been used for analyzing the results, which are presented in the following two chapters.

5.5 Conclusion

In this chapter we have discussed the setup of the Delphi study in terms of the process and the constitution of the expert panel and the statistics on their participation. In the first round, each of the experts that agreed to participate was sent a questionnaire with open questions, based on the BI models that we presented in Chapter 2. From the questionnaires that were sent back, we have created an collated overview of opportunities and limitations. We distributed this overview in the second round to be rated and commented on by the expert panel. A selection of these opportunities and limitations were again presented in an overview, including the average ratings and comments of round 2, to the expert panel in the third round, to provide the opportunity to change the rating and to provide further argumentation on the rating.

After an initial response rate of 62% (16 participants), the response rate declined to 42% (11 participants) in the last round. Although this decline is common for a Delphi study, we do realize that the number of active participants has been lower than the number of participants we aimed for, as we identified a need for 20 participants in Chapter 4. Having said this, we do believe that the study provided us with valuable results. These results are discussed in the following chapter.

6 Results

After three rounds of identifying, rating, and commenting on opportunities and limitations by our group of experts, we have produced a final overview of these opportunities and limitations. In this chapter we present these results, by listing the exact hypotheses as they have been used throughout the Delphi study, together with the statistics of the rating they received. The results are presented separately for the items of the BI-cycle, the items of the BI functions model, and the catch-all question.

Section 6.1 serves as introduction to the detailed results, and discusses some of our general findings from the results analysis. Section 6.2, 6.3 and 6.4 discuss the results related to the BI-cycle, the BI functions model, and the catch-all question respectively. Section 6.5 discusses the statistical analysis, Section 6.6 contains the conclusion.

6.1 General observations

In total, 39 opportunities and 11 limitations have been identified and rated in this research. One additional general hypothesis has been formulated and rated as well. For 16 opportunities, the technical feasibility was also rated. In this section we discuss the distribution of these hypotheses over the categories, the use of the “no opinion” option, and the presentation of the results overview.

6.1.1 Distribution of items over categories

Table 5 provides an overview of the distributions of the items over the categories. However, the distribution itself does not necessarily reflect the importance of each category in terms of value of the opportunities or limitations, because of the following reasons:

- The BI-cycle focuses more on technology than the BI functions model, and since most opportunities are technology related, they have been identified along the lines of this model;
- The BI-cycle and BI functions model overlap. Some opportunities could have been mentioned in both models, and because the BI-cycle appeared first in the questionnaire, there was no need to repeat opportunities or limitations in the BI functions model;
- The granularity of the opportunities and limitations varies. Some cover only a small part of a category in comparison to others;
- The items have different ratings, and therefore differ in the extent to which an opportunity is really an opportunity. The same accounts for the limitations.

Table 5: Distribution of items over categories

	Total	**BI-cycle**	Phase 1	Phase 2	Phase 3	Phase 4	General	**BI-functions**	Strategic	Tactical	Operational	General	[illegible]	**Catch-all**
Opportunities	***39***	**23**	-	11	2	6	4	**16**	1	3	5	2	2	**3**
Limitations	***11***	**4**	1	1	2	-	-	**8**	1	1	1	-	3	**2**
Technical feasibility	***16***	**13**	-	6	1	5	1	**5**	-	2	2	1	-	**-**
General													1	

Based on the technological orientation of an opportunity, we have decided whether or not the technical feasibility of this opportunity should be rated. For the opportunities based on the BI-cycle, a larger percentage has been rated on technical feasibility than in the BI-function model, purely because these opportunities were more technology-oriented.

6.1.1 "No opinion" percentages

The second and third round questionnaires contained an "no opinion" option that could be selected for each hypothesis, instead of providing a rating. The option served to prevent experts from choosing "neutral" in case they had no opinion, which would have corrupted the results. As displayed in Table 6, this option has been used rarely. From the total amount of 51 hypotheses, 15 hypotheses have one or more "no opinion" selections from one or more experts regarding the general rating. For most hypotheses it involved only one expert who selected this option. For four hypotheses it involved two experts, and for one hypothesis it involved three experts.

For the 16 hypotheses for which the technical feasibility has been measured, 11 of them have one or more "no opinion" selections from one or more experts. This is a higher percentage than for the general rating, indicating that knowledge on the technical feasibility is a more difficult subject for some of the experts.

A reason for selecting the "no opinion" option can be a lack of knowledge of the expert about the subject of the hypothesis, ambiguity of the hypothesis, or something else. Because we do not know exactly what caused some experts to select the "no opinion" option, we do not draw any further conclusions from the numbers in Table 6.

Table 6: Overview of selections of "No opinion" for hypotheses

	7-8% (1 expert)	**13-17% (2 experts)**	**23% (3 experts)**
	General rating		
Number of hypotheses	10	4	1
Hypotheses involved	1, 2, 13, 15, 23, 25, 28, 34, 40, 45	11, 17, 48, 51	14
	Technical feasibility rating		
Number of hypotheses	5	4	2
Hypotheses involved	19, 22, 34, 38, 39	2, 10, 18, 23	11, 14

6.1.2 Presentation of overview

In the following three sections, we present the results according to the phases of the BI-cycle, the categories of the BI functions model, and the catch-all question. The corresponding tables contain the hypotheses that have been used in the questionnaires, and summarize the results with the following variables:

- Number of the item (#). An asterisk has been added when the item has been rated in both round 2 and 3;
- Number of times mentioned in the first round by different experts (R1);
- Mean of the general ratings ($\bar{x}$);
- Standard deviation of the general ratings (s);
- Mean of the technical feasibility ratings (tf);
- Standard deviation of the technical feasibility ratings (tfs).

We regard the mean of the general rating ($\bar{x}$) as the most important rating, to which the items in the tables are ordered accordingly in each category.

The tables that summarize the results also contain a box plot, which graphically depicts the smallest observation, median, lower quartile, upper quartile, and largest observation. The box plots are representing only the general ratings. The value axis that numbers from -3 to +3 can be found at the bottom of each table. They serve to display differences between the ratings without making any assumptions of the underlying statistical distribution.

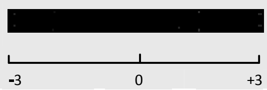
6.2 BI-cycle

As described in Section 2.3, the BI-cycle consists of four phases: (1) planning and direction, (2) the collection of data, (3) analysis of data, and (4) distribution of information. Most of the identified opportunities are in phase 2, the collection of data, and phase 4, the distribution of data.

One item has been identified for the first phase of the BI-cycle: a limitation of using SOA in the planning & direction phase. Because the phase is merely a preparation phase, it does not need a lot of technology, and that is why it is probably hard to come up with ideas. Concerning the limitation itself, it received a low rating (x=0.1), but also low consensus (s=2,2). From the comments we can conclude that some experts strongly disagree with the hypothesis; they argue that the planning and direction phase can be structured while being creative at the same time, and that preparing the use of SOA is very important before implementing it. The possibility of using SOA in *supporting* this phase is unclear: no specific opportunities are mentioned.

Table 7: Results for BI-cycle Phase 1: planning & direction

#	Limitation	R1	$\bar{x}$	s
1	An important limitation of using SOA in the planning and direction phase is that this phase mainly consists of creative processes that are not very well structured and requires a lot of human interaction.	1	0.1	2.2



The second phase of the BI-cycle, the collection of data, yielded a large amount of items: 11 opportunities and 1 limitation (Table 8 and Table 9). Using Master Data Management (MDM) based on SOA (#3; $\bar{x}$=1,9) received a high rating and high consensus, although it was mentioned only by one expert in the first round. Implementing transformational services (#3; x=2,1) also received a high rating and high consensus. These two opportunities are both about standardizing data formats, indicating that this is a hot topic in the business intelligence community. They also received a rather good rating for the their technical feasibility. We further analyze these two opportunities in Chapter 7.

Most opportunities mentioned in this phase concerned using services as a source for data: services that wrap data access and data feeds instead of directly accessing the data. The use of services as data source for BI is likened by the experts (#4; x=1,9). Several benefits of using services as data sources are mentioned. Experts mentioned easier access of data sources outside of the organization (#5; $\bar{x}$=1,7) as an important benefit, probably because SOA enables the standardized use of services over organizations. Enabling reuse of data sources is mentioned as another important benefit (#6; x=1,7), because this would improve the agility of implementing and changing BI solutions. Other identified benefits are the possibilities SOA offers for receiving data from its source in real-time on event-basis (#7; $\bar{x}$=1,5) and having an improved overview of the available data for BI (#8; $\bar{x}$=1,2). A rather low rated benefit concerns separation of concern for people responsible for different kinds of data ((#9; $\bar{x}$=0,3), which is stated to reduce the risk that the data is wrong.

It is clear that the experts regard services as data source for BI a good opportunity. However, the technical feasibility of using services as a data source received a low rating (tf=0,5). Most probably this low rating is related to the only limitation mentioned in this phase: the large data size involved in BI (#13; x=1,4). The argumentation of some experts is that the high data volumes often involved in the collection of data for BI, is not suitable for transportation through services. The consensus on this hypothesis is remarkably low (s=1,9), which indicates that opinions on this hypothesis are sharply divided. This hypothesis is discussed in more detail in Chapter 7.

For the other opportunities mentioned (#10, #11, and #12) we do not have additional comments.

Table 8: Results for BI-cycle Phase 2: collection of data

#	Opportunities	R1	$\bar{x}$	s	$\bar{tf}$	tf s
2*	A good opportunity for applying SOA to the collection of data is by transformational services that fulfill the role of a transformation step of the conventional BI architecture. These services can increase the use of standard data formats.	4	2.1	0.7	1.0	1.4
3*	A good opportunity for applying SOA to the collection of data is by providing access to Master Data Management (MDM) based on SOA. Every layer of the BI cycle can provide input to and/or use MDM information.	1	1.9	1.4	1.1	1.2
4	A good opportunity for applying SOA to the collection of data is to use services as a data source.	5	1.9	1.6	0.5	1.6
5	An important benefit of using services as a data source is that external data sources are easier accessible through services.	5	1.7	1.7		
6	An important benefit of using services as a data source is that it enables reuse of those data sources. This makes the BI organization more agile in response to changing information needs.	1	1.7	1.7		
7	An important benefit of using services as a data source is that it is then easier to access data in a (near) real time, event driven model.	1	1.5	1.6		
8	An important benefit of using services as a data source, is that by having an overview of the services for BI (e.g. though a services library) you have insight in the data available for BI, instead of the data being scattered and 'hidden' across the application landscape in non-SOA environments.	4	1.2	1.4		
9	An important benefit of using services as a data source is that collecting data from published services ensures that there is a separation of concern and that the risk that the data is wrong is reduced. The responsibility for the quality of data is taken by a different project team who must have thought about it and dealt with it to have been able to publish a service definition.	1	0.3	1.7		
		-3		0		+3

Table 9: Results for BI-cycle Phase 2: collection of data(continued)

#	Opportunities	R1	x̄	s	tf̄	tf s
10	A good opportunity for applying SOA to the collection of data is by wrapping components like data validation and data cleansing as services, so multiple data integration and BI projects can reuse those components. The input should provide the source data needed to be processed and filter conditions that the service should apply to this source data. The output will be pre processed data or a subset of changed data.	1	1.6	1.8	0.3	1.6
11	A good opportunity for applying SOA to the collection of data is by using a distributed/virtual/federated data warehouse. Raw data can be distributed across a grid, and are accessed for analysis via grid services.	2	1.3	1.2	0.9	1.6
12*	A good opportunity for applying SOA to the collection of data is by performing bulk synchronization between the operational systems and the BI systems, using conventional methods (which are able to handle the voluminous size of the data), and receiving regular updates via services. More up to date information is then available.	1	1.2	1.4	1.4	0.9
	Limitations					
13*	An important limitation for applying SOA to the collection of data is the large data size involved, which is usually not suitable for transportation through services.	3	1.4	1.9		
		3		0		+3

The third phase of the BI-cycle, analysis of data, has yielded a rather small amount of items, as presented in Table 10. Using SOA for communication between the different components of the BI solution is regarded as a good option (#14; x̄=1.8) and rated technically feasible, which can be explained by the support of SOA for integration purposes, as described in Section 3.7.1. The other opportunity in this phase is that SOA would provide the possibility to offer specific analytic functionality through services (#15; x=1.2). Offering functionality in a standardized format, e.g. through web services, would mean the functionality can easily be consumed by others.

The two limitations that are mentioned for this phase received a rather low rating (#16; x̄=0.3) (#17; x=0.1), but the low consensus also shows that on both hypotheses the opinions are very diverse. Unfortunately no further comments have been given on why experts did or did not agree with these hypotheses.

Table 10: Results for BI-cycle phase 3: analysis of data

#	Opportunities	R1	x̄	s	tf̄	tf s
14	A good opportunity for applying SOA to the analysis of data is for communication between different components of the BI solution.	1	1.8	0.6	1,3	1,1
15	A good opportunity for applying SOA to the analysis of data is that providers of specific analytic functionality are provided a low friction way of deploying their specific expertise.	1	1.2	1.0		
	Limitations					
16	A limitation of using SOA in the analysis phase is that this phase is often very diverse and specific, limiting the role of services to pre defined queries and standard analysis (row count, null values, etc.)	3	0.3	1.8		
17	A limitation of using SOA in the analysis phase is that analysis through services is only suitable for lower organizational levels, where decisions are structured more clearly.	1	0.1	1.9		
		-3		0		+3

The fourth phase of the BI-cycle, the distribution of information, seems to provides some good opportunities for the use of SOA (Table 11). The expert panel highly agrees that exposing information through services is a good opportunity (#19; x=2.3), providing a comfortable way to access BI information. Especially the exposure of information for use in other applications, known as *closed-loop BI*, is considered a good opportunity (#18; $\bar{x}$=2.5). Closed-loop BI is referred to by Schiefer et al. (2003) as “the enhancement of the operational system with BI”.

Mashups are identified as another opportunity (#20; x=1.9). *Mashups* are build by combining data from several web services into one single tool, and is often associated with web services available on the Internet (Wikipedia, 2008). Because existing examples often focus on visualization of information, this opportunity is mentioned in this category. As commented on by several of the experts, mashups are just an example of what can be done with services.

Another identified opportunity is known as *BI-for-all*, in which SOA could support the distribution of information to many different types of users who can employ the available services the way they like (#21; x=1.4). However, the rather low consensus (s=1.7) illustrates disagreement with this hypothesis as well. Comments of experts indicate they are afraid of the inability of some users to use the services, and anticipate an uncontrolled wild grow of services across the organization.

Using configurable services as data provider to the OLAP query process (#23; x=0.5) shows low ratings on the level of agreement to representing a good opportunity for SOA. However, from the low consensus (s=2.3) we can conclude that opinions on the opportunity itself are divided. Most experts do agree that the opportunity is currently not really technically feasible (tf= -0.4), making it the single opportunity in this research with a negative rated technical feasibility. This opportunity is also mentioned in research by Wu et al. (2007), in which they propose a service-orientation approach for BI.

Table 11: Results for BI-cycle phase 4: distribution of information

#	**Opportunities**	**R1**	**$\bar{x}$**	**s**	**$\overline{tf}$**	**tf s**
18*	A good opportunity for applying SOA to the distribution of data is by exposing information for use in other applications (so called closed loop BI).	6	2.5	0.6	2.1	1.0
19*	A good opportunity for applying SOA to the distribution of data is by exposing the produced BI information through services.	4	2.3	0.7	1.8	1.0
20	A good opportunity for distributing information by services is by creating *mashups*: combining information sources by using 3rd party providers. E.g. visualizing sales or stock levels using Google Maps.	2	1.9	0.7	1.6	0.7
21	A good opportunity for applying SOA to the distribution of data is the concept BI-for-All: using services to provide BI functionality and information to many different types of users throughout the whole organization, in an easy way.	3	1.4	1.7		
22*	A good opportunity for applying SOA to the distribution of data is by wrapping complex queries and other BI functionalities, demanded by users, as services.	1	0.5	2.3	0.5	1.6
23	A good opportunity for applying SOA to the distribution of data is by using configurable services as data provider to the OLAP query process. The input can be a cube and filter conditions, the outputs can be a cube or a data mart.	1	0.8	1.8	-0.4	0.9
		-3		0		+3

Some opportunities have been proposed that cover more than one phase of the BI-cycle, presented in Table 12.

The highest rated opportunity is about componentization of the whole BI architecture (#24; $\bar{x}$=1.9), with the reasoning that the use of services could support the building of loosely coupled BI systems. Comments from experts point again to possible issues with large data sizes involved, which should be kept in mind.

Opportunity #25 ($\bar{x}$=1.3) is about using data from services and data from the data warehouse both to provide more timely information. A possible realization of this opportunity would be to collect timely data from services that do not add up to large amounts of data, and combine this with data from the data warehouse.

One expert mentioned the opportunity to provide insight in process information to SOA, through the Business Process Management layer of SOA (#26; $\bar{x}$=1.1). The average agreement level is rather low. A comment provided by an expert indicate that it is not clear what is exactly meant by the hypothesis. Another comment made, which could also explain the low rating, is that process information can still be scattered in an SOA environment. As is described in more detail in Chapter 7, Business Process Management (BPM) is often related to Business Activity Monitoring, but is not necessarily part of an SOA.

Another proposed opportunity is the support of SOA for better fault prevention in data, and more flexibility in reconfiguring complex information flows. The rather low rating (#27; $\bar{x}$=0.5) cannot easily be explained, because the hypothesis contains several parts on which can be agreed or not. The only comment provided pointed out a weak agreement level because it would require the services to be rigorously defined and implemented, within a robust data architecture.

Table 12: Results for BI-cycle phase: general

#	Opportunities	R1	$\bar{x}$	s	$\bar{tf}$	tf s
24*	A good opportunity for applying SOA to the BI process is that the use of services enables componentization of the BI architecture, to support loosely coupled BI systems.	2	1.9	0.8	1.2	1.3
25*	A good opportunity for applying SOA to the BI process is by Information Integration based on SOA, which can decrease the latency of data (e.g. caused by overnight processing), by collecting, combining, analyzing and distributing information together with the information from the traditional data warehouse.	1	1.3	1.6		
26	A good opportunity for applying SOA to the BI process is by modeling and executing processes in the Business Process Management layer of SOA. This provides insight into process modeling and process execution information available in the application landscape and available for BI, instead of the process information being scattered and 'hidden' across the application landscape in non SOA environments.	1	1.1	1.3		
27	The data/information flows in organizations from the operational source systems to data warehouse and from data warehouse to the end user are often very complex. Moreover, errors and mistakes in the components in the path of the information flows are easily hidden due to aggregations and statistical analysis. A good opportunity for applying SOA to the BI process is the self containedness of services, which allows better fault prevention and tolerance, and more flexibility in reconfiguring the information flows.	1	0.5	1.7		

-3 0 +3

6.3 BI Functions model

In this section we present the results based on the BI Functions model. Most opportunities have been identified for the operational and tactical level. Comments from the participants indicate that BI information in general is more often delivered to the tactical and operational level. We cannot conclude from this that there are less opportunities of using SOA for building BI applications for the strategic level, but it would explain why less opportunities are identified.

For the strategic level, only one opportunity and one limitation has been identified, as displayed in Table 13. The identified opportunity that using SOA for BI on the strategic level can lead to cost reduction and a decrease in development time in building BI solutions (#28, x=0.6) indicates a rather low average agreement to this hypothesis, but also a low consensus between experts. Comments indicate the relation of this opportunity to the limitation mentioned for the strategic level, which draws on the large data sets issue that was identified earlier, and according to the hypothesis would especially apply to the strategic level. The rating for this limitation is rather low as well (#29, x=0,6), and also points out low consensus among the experts. Experts who disagreed with the limitation, and agreed with the opportunity, commented that large data sets are not a problem for data that is already aggregated; services could be used very well to distribute the produced information and could be useful in combining and reconfiguring the many sources of information that the strategic level often requires.

Table 13: Results for BI functions: strategic level

#	Opportunities	R1	$\bar{x}$	s	tf	tf s
28	A good opportunity for applying SOA to the strategic level is that using SOA for this level can save costs and reduce development time of BI solutions.	1	0.6	1.6		
	Limitations					
29	An important limitation for applying SOA to BI on the strategic level, is that BI on this level often involves large amounts of data, which is not suitable for use with services.	1	0.6	1.7		
		-3		0		+3

As displayed in Table 14, opportunity #30 ($\bar{x}$=1,7) states that using SOA for BI at the tactical level can result in a flexible architecture, which is useful at this level in which changes often occur. Agreement to this hypothesis is rather high, as well as the consensus. The flexibility achieved can be related to the hypothesis #6, which received a high rating for reuse possibilities of data sources of BI.

The opportunity for integration of data sources (#32, $\bar{x}$=1,2) can be related to the high rating that has been given to hypothesis #4 on the use of services as data sources. However, the rather low consensus indicates that not all experts agree that this is especially an opportunity for the tactical level. The limitation mentioned for this level (#33, x=0,1) states that integrating multiple sources of data, on the contrary, would lead to more work. However, the low level of agreement to this statement in comparison to opportunity #32 indicates that on average the experts agree that SOA would improve integration possibilities for data sources of BI.

For opportunity #31 we have no further information available as to why SOA could shorten the time frame between demand and delivery of information.

Table 14: Results for BI functions: tactical level

#	Opportunities	R1	x̄	s	tf x̄	tf s
30	A good opportunity for applying SOA to BI on the tactical level is by creating a flexible architecture, as changes often happen at this level.	2	1.7	0.8		
31	A good opportunity for applying SOA to BI on the tactical level is that it enables shortening the time frame between demand and delivery of information.	3	1.5	0.8		
32	A good opportunity for applying SOA to BI on the tactical level exists for integrating data from multiple sources, which is easier through (platform independent) services.	2	1.2	1.5		
	Limitations					
33	An important limitation for applying SOA to BI on the tactical level, is that this often requires combining multiple sources. Processing these data sources would result in large amounts of programming.	1	0.1	2.0		
		-3		0		+3

Several opportunities have been identified for the operational level, which are displayed in Table 15. The highest rated opportunity concerns Business Activity Monitoring (#34, x=2.1), a BI function that has been introduced already in Chapter 2, and is about monitoring business processes in a (near) real-time fashion. This opportunity is discussed in more detail in Chapter 7.

Opportunities #35-37 are defined almost the same as #30-32 were defined for the tactical level. These opportunities were mentioned in both categories in the first round, and are therefore also separately rated. The ratings for these opportunities on the operational level are almost the same as for the tactical level, which leads us to the conclusion that the experts consider these almost equal opportunities for both levels.

Opportunity #38 (x̄=1.5) on closed-loop BI relates strongly to opportunity #18, which is about closed-loop BI in a general sense, not specific to any organizational level. For the operational level, this opportunity is rated somewhat lower, but no further comments suggest why this is the case.

Table 15: Results for BI functions: operational level

#	Opportunities	R1	$\bar{x}$	s	$\bar{tf}$	tf s
34*	A good opportunity for applying SOA to BI on the operational level is in Business Activity Monitoring, to monitor business processes in (near) real time fashion.	4	2.1	1.2	1.4	1.0
35	A good opportunity for applying SOA to BI on the operational level is by creating a flexible architecture, as changes often happen at this level.	2	1.8	0.9		
36	A good opportunity for applying SOA to BI on the operational level is that it enables shortening the time frame between demand and delivery of information.	2	1.6	1.4		
37	A good opportunity for applying SOA to BI on the operational level exists for integrating data from multiple sources, which is easier through (platform independent) services.	2	1.5	1.6		
38	A good opportunity for applying SOA to BI on the operational level is in closed loop BI, transporting produced information directly into other applications and processes.	1	1.5	1.5	0.8	1.6
		-3		0		+3

In Table 16, we listed the opportunities that apply to more than one organizational level. Opportunity #39 (x=1.8) describes a BI architecture that enables a more dynamic approach to deliver source data to the ETL process. Services can accept messages with information on the source data that should be approached and information on how this should be transformed, and how often the data should be refreshed. Part of this opportunity can also be found in the article of Wu et al. (2007).

Opportunity #40 ($\bar{x}$=1.1) is about business process improvement (BPI): improving the efficiency, effectiveness, and adaptability of business processes (Bhatt, 2000). No comments further explain this hypothesis. Research by Bhatt (2000) indicates that better integration of IT benefits BPI. Since integration is one of the goals of SOA, the link can be made to BPI, though this does not prove that BPI *based on BI information* is easier through SOA.

Table 16: Results for BI functions: general

#	Opportunities	R1	$\bar{x}$	s	$\bar{tf}$	tf s
39	A good opportunity for applying SOA to BI in general, is that through SOA-application integration services, BI can turn into operational BI as one part of the IT systems of the organization. The input data should be source specification and refresh rate and subscribed data structure. The output can be CDC data (Change Data Capture, data changed in specific time period) for BI near-real time ETL processes for operational level and tactical level.	1	1.8	1.4	1.1	1.3
40	A good opportunity for applying SOA to BI in general is that on all organizational levels, business process improvement based on BI information is easier through SOA.	1	1.1	1.4		
		-3		0		+3

The other axis of the BI-functions model outlines the focus areas for BI, like the internal organization, suppliers, customers, and competitors. Several opportunities and limitations are mentioned for these areas, and are listed in Table 17. However, a general statement is made as well (#46, x=1.6), which summarizes the ratings of the identified opportunities and limitations: “There is not much difference between the focus areas in opportunities for applying SOA to BI.

Everywhere where data is transferred from one system to another this might be done through services. This can be within an organization but also between organizations." Though the opportunity of SOA for BI for the internal organization is rated slightly higher (x=1.8), it seems that it is the opinion of the experts that the opportunities of SOA for BI are almost equally applicable to each focus area.

Table 17: Results for BI functions: focus areas

#	Opportunities	R1	x̄	s	tf x̄	tf s
41	Good opportunities for applying SOA to BI exist within the internal organization.	3	1.8	0.9		
42	Good opportunities for applying SOA to BI exist in the supply chain, when the processes are complicated and has a high volume of orders.	2	1.6	1.3	[redacted]	
	Limitations					
43	An important limitation for applying SOA to BI for market & competitor analysis is that this is often an unstructured process and therefore mainly a human process, not offering opportunities for SOA.	1	0.2	2.0	[redacted]	
44	An important limitation for applying SOA to BI for market & competitor analysis is often delivered now and then, therefore processing batches would be more suitable than using services.	1	-0.5	1.7	[redacted]	
45	A good opportunity for applying SOA to BI for customer intelligence only exists in an industry with a large number of customers.	1	-0.8	1.9	[redacted]	
	General					
46	There is not much difference between the focus areas in opportunities for applying SOA to BI. Everywhere where data is transferred from one system to another this might be done through services. This can be within an organization but also between organizations.	2	1.6	1.2	[redacted]	
		-3		0		+3

6.4 *Catch-all question*

The BI-cycle and the BI-function model were meant to structure the answers of experts and encourage to think about several areas in which opportunities and limitations could be identified. At the end of the questionnaire of the first round, the experts were asked a *catch-all* question, for identifying opportunities and limitations that did not fit in one of the models. Table 18 lists the answers on this question.

According to Gile et al. (2005), existing BI systems are composed of mostly tightly coupled components, and have a proprietary architecture meant to satisfy the needs of only one system rather than fitting in with all IT components. Opportunity #47 (x̄=2.4) is about improved integration possibilities of BI (components) with other components. As SOA indeed promotes improved integration, most of the experts believe this improvement can be realized for BI as well. The comments from experts indicate that the current developments around SOA will lead to (more) open standards, which are required for improvement of integration.

The integration and reuse possibilities that were already mentioned in previous opportunities, return in hypothesis #48 on pervasive or ubiquitous BI, which, according to Mittereder (2005), empowers everyone in the organization, at all levels, with analytics, alerts and feedback mechanisms. The high rating of this opportunity (x=2.1) points out that most experts think that SOA can certainly support this concept.

Table 18: Results for BI functions: strategic level

#	Opportunities	R1	x̄	s	tf	tf s
47*	A good opportunity for applying SOA to BI is in easier integration of other components with BI, like the data access tools.	1	2.4	0.6		
48*	A good opportunity for applying SOA to BI is pervasive or ubiquitous BI, which means that BI will be used everywhere and in every process throughout the organization. Just like SOA can add business value by disconnecting the business process from the underlying applications, BI can be a component of this.	1	2.1	1.1		
49*	A good opportunity for applying SOA to BI is in the creation of a more agile organization, which can react faster to changes in the business	1	1.8	1.6		
	Limitations					
50	An important limitation for applying SOA to BI is that the (Business) Process Focus is not very common in the BI community, but is important for working with SOA.	1	0.6	1.6		
51	An important limitation for applying SOA to BI is that SOA is often vendor specific, which poses limitations to the use of BI and SOA	2	0.1	1.9		
		-3		0		+3

Developing an agile organization is the ultimate goal of using SOA, as we described in Chapter 3. It is no wonder that this opportunity is mentioned for the organization of BI as well, and receives a rather high rating (#49, $\bar{x}$=1.8). However, the statement is very general and does not provide us much additional information on the reuse and integration opportunities already mentioned, which are *supposed* to improve agility of the organization.

Two limitations are identified, and both get a rather low rating, with low consensus. Limitation #50 states that in the BI community, the business process focus is not very common, which could be a problem for working with SOA, which is often process focused. This non-technical limitation is regarded by some experts as a serious limitation, though the low consensus shows that the opinion about this is diverse. The second limitation (#51, x=0.1) concerns the proprietary products of vendors that are not always compatible. This limitation can be related to #47, for which experts commented that they think the increasing adoption of SOA will drive the use of open standards.

6.5 Further statistical analysis

Using the quantitative data collected during the 2nd and 3rd round of the Delphi study, we have performed several statistical analyses to test significance and patterns in the data.

6.5.1 Significance of ranking

In the overview of the results that has been provided in this chapter, the means indicate the average agreement to each hypothesis, and the standard deviation indicates the consensus between the experts. A high value for standard deviation is caused by a high spread of ratings, and indicates a low consensus between the group of experts.

In statistics, the *Duncan's new multiple range test* (MRT) is used to compare sets of means, and analyzes the statistical significance of differences between those means. When the means in a set are not significantly different from each other, the means in this set should have the same ranking. Duncan's MRT has a relation to the standard deviation: when from a set of means, the standard deviation of each mean is high, and the means are very close to each other, applying Duncan's

MRT to this set points to an insignificant difference between the means. This relation can be found in the description of Duncan's MRT in the Encyclopedia of Statistical Sciences (2006).

We have applied Duncan's MRT to the set of opportunities and the set of limitations of our research. The results of this test, calculated with a significance level of $p < 0.10$, are displayed in Table 19 (opportunities) and Table 20 (limitations). From these results we can conclude that there is a large overlap in means: rather large sets of hypotheses have insignificantly different means. For example, subset 4 from Table 19 includes means from 27 opportunities. That these opportunities are all in one set means that they are not significantly different from each other in terms of their ranking. We have also applied the Duncan's MRT to the ranking within each category, for example, for all opportunities categorized in BI-cycle phase 3. The results of these tests are in Appendix A. The results indicate that also within categories, the ranking is insignificant for the majority of the cases.

The results of the test can be explained by the often low consensus in the expert panel, which can also be concluded from the rather high standard deviations. We think the causes of the low consensus lie in the difficulty and novelty of the subjects under consideration, and in the smaller than planned panel of experts. In Chapter 4 we stated the need for twenty participants, though ratings (which have been given in the second and third round) have only been given by 15 different participants.

We conclude to say that the Duncan's MRT test indicates that the order of the items in the overview should not be perceived as a definite ranking in each case. More discussion on the items would be needed to come closer to an agreement between experts, and more participants would be needed, in order to produce a more significant ranking.

6.5.2 Rank order correlation between groups

We have performed further analysis on the correlation between the different groups that participated. Table 21 shows the correlation statistics for the opportunities, the limitations, and the technical feasibility, between the researchers and vendors, the consultants and vendors, and the consultants and researchers. For calculating the correlation we have used the ***Spearman rank correlation coefficient***, also employed by Scholl et al. (2004) for the same purpose. When we take significance as $p < 0.10$, only five of the nine comparisons are rendered significant. Considering that a positive correlation should fall between 0 and 1, 1 representing the strongest correlation possible, we have to conclude that although the correlation between the groups is positive, it is rather low. Further conclusions on relations between the groups are difficult to draw, because the significant correlations do not show clear patterns between the groups.

Table 19: Duncan's new multiple range test (MRT) for all opportunities after third round

Hypothesis Number * = rated in round 2 + 3	**N**	**Subsets for p = 0.10**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
9	12	0,33						
*22	15	0,47	0,47					
27	13	0,54	0,54	0,54				
23	12	0,75	0,75	0,75	0,75			
26	12	1,08	1,08	1,08	1,08	1,08		
40	12	1,08	1,08	1,08	1,08	1,08		
32	13	1,15	1,15	1,15	1,15	1,15		
8	12	1,17	1,17	1,17	1,17	1,17		
15	12	1,17	1,17	1,17	1,17	1,17		
*12	15	1,20	1,20	1,20	1,20	1,20		
*25	14	1,29	1,29	1,29	1,29	1,29	1,29	
11	10	1,30	1,30	1,30	1,30	1,30	1,30	
21	13	1,38	1,38	1,38	1,38	1,38	1,38	
37	13		1,46	1,46	1,46	1,46	1,46	1,46
7	13		1,54	1,54	1,54	1,54	1,54	1,54
31	13		1,54	1,54	1,54	1,54	1,54	1,54
38	13		1,54	1,54	1,54	1,54	1,54	1,54
10	13			1,62	1,62	1,62	1,62	1,62
36	13			1,62	1,62	1,62	1,62	1,62
42	13			1,62	1,62	1,62	1,62	1,62
5	13				1,69	1,69	1,69	1,69
6	13				1,69	1,69	1,69	1,69
30	13				1,69	1,69	1,69	1,69
35	12				1,75	1,75	1,75	1,75
39	13				1,77	1,77	1,77	1,77
41	13				1,77	1,77	1,77	1,77
14	10				1,80	1,80	1,80	1,80
*49	15				1,80	1,80	1,80	1,80
*3	15				1,87	1,87	1,87	1,87
*24	15				1,87	1,87	1,87	1,87
20	11					1,91	1,91	1,91
4	13					1,92	1,92	1,92
*2	4					2,07	2,07	2,07
*48	13					2,08	2,08	2,08
*34	14					2,14	2,14	2,14
*19	15						2,33	2,33
*47	15						2,40	2,40
*18	15							2,53
Sig.		0,11	0,11	0,11	0,10	0,12	0,10	0,12

Table 20: Duncan's new multiple range test (MRT) for all limitations after third round

Hypothesis Number * = rated in round 2 + 3	**N**	**Subsets for p = 0.10**	
		1	2
45	12	-0,83	
44	13	-0,54	
33	13	0,08	0,08
1	11	0,09	0,09
17	11	0,09	0,09
51	11	0,09	0,09
43	13	0,23	0,23
16	13	0,31	0,31
29	13	0,62	0,62
50	13	0,62	0,62
*13	14		1,36
Sig.		0,11	0,15

Table 21: Rank order correlations (Spearman-rho) of means after third round

	Opportunities (39)	*sig.*	Limitations (11)	*sig.*	Techn. Feas. (16)	*sig.*
RS/VD	0,12	*0,45*	-0,30	*0,37*	0,45	*0,08*
CS/VD	0,49	*0,00*	0,07	*0,84*	0,41	*0,12*
CS/RS	0,26	*0,11*	0,57	*0,07*	0,66	*0,01*
RS = researchers, CS = consultants, VD = vendors, sig. = significance (p)						

6.5.3 Change analysis between round 2 and 3

We have also analyzed the change in ratings between the second and third round for the 13 hypotheses that were rated in both of these rounds. For this analysis, we only used the ratings provided by the nine experts that participated in both rounds, which is why these ratings are different from the ratings in the overview.

For some hypotheses the consensus increased, for some the consensus decreased from the second to the third round. No overall increase or decrease in consensus can be detected. We also analyzed the significance in the change of the ratings for each individual hypothesis. For this analysis, we used the ***Kruskal-Wallis one-way analysis of variance*** test, of which the results are in Appendix B. According to this test, none of the changes were significant for $p < 0.10$. Therefore we conclude that there were no significant changes in ratings from the 2^{nd} to the 3^{rd} round.

We also performed a ***opinion stability analysis (***Scheibe et al., 1975), which analyses the ***amount*** of changes in ratings made by the experts. A detailed example of how these changes are calculated can be found in Appendix C. The method does not analyze change in consensus, but to what extent changes are made at all. The method can be used to determine if additional Delphi rounds would change the results, regardless of the consensus achieved. According to Scheibe et al. (1975), a change less than or equal to 15% is considered to represent stability between two rounds.

Table 22 presents the results of the analysis for the 13 hypotheses that were rated in the second and third round. Changes from or to the option of “no opinion”, have not been analyzed. From the 13 hypotheses, 8 reached stability in the general rating. From the 8 hypotheses measured on technical feasibility, 4 reached stability in the ratings on the technical feasibility. From this analysis we can conclude that more stability could have been reached by additional rounds. Time constraints limited us to actually organize a third round.

Table 22: Results of stability analysis between round 2 and 3

# Hypothesis		**General rating**			**Technical feasibility**	
	N	Percentage change		N	Percentage change	
2	8	13%	stable	7	0%	stable
3	8	0%	stable	8	25%	
12	8	13%	stable	8	25%	
13	8	38%				
18	9	22%		7	29%	
19	9	0%	stable	8	25%	
22	9	22%		7	14%	stable
24	9	11%	stable	9	33%	
25	8	13%	stable			
34	9	22%		9	0%	stable
47	9	0%	stable			
48	8	25%				
49	9	11%	stable			

Our analyses on the differences between round 2 and 3 show that changes did occur, but that the overall consensus did not significantly increase or decrease from round 2 to round 3. We do realize that the sample in this analysis is rather small, and might not be large enough to make strong claims about any of the results of the analyses.

6.5.4 Correlation general rating and technical feasibility rating

For the sixteen opportunities for which also the technical feasibility has been rated, we have analyzed the correlation of these two ratings. Although experts were allowed to rate the two variables independently, we did expect that an increased technical feasibility would lead to a higher general rating of an opportunity as well. We have analyzed the correlation for the aggregated data set (the means of the ratings), and for all individual ratings of the experts. Using the Pearson correlation test, the result is a Pearson correlation for the aggregated data of 0.699, with a probability value of $p = 0.003$, which is well below the conventional threshold of $p < 0.05$.

For the individual ratings, the result is a Pearson correlation for the aggregated data of 0.467, with a probability value of $p = 0.02$. Both Pearson correlations indicate a rather strong correlation, which is also illustrated by Figure 12 in which the aggregated correlation is depicted. In Appendix D the figure for the individual ratings of showed as well. This confirms our expectation. Especially towards the higher ratings, the general rating and technical feasibility rating show strong correlation. Another conclusion that can be drawn from the figure is that for each opportunity, the technical feasibility rating is almost always lower than the general rating. We believe that this can be explained by the general assumption that technology lags behind the ideas that can be realized *with* technology.

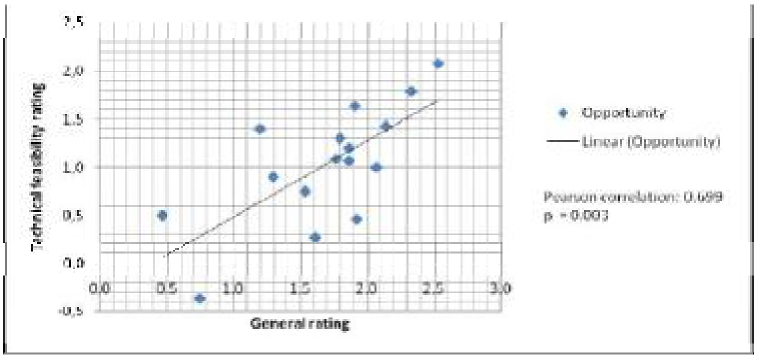


Figure 12: Correlation general rating and technical feasibility rating of an opportunity

6.6 Conclusion

In the Delphi study, 51 hypotheses on opportunities and limitations of using SOA to build BI-applications have been identified and rated. The study has produced different kinds of opportunities and limitations, varying in granularity, area of application, and relation to technology. From the overview, we have drawn a number of conclusions that we discuss below.

The majority of the identified opportunities have a strong relation to the characteristics and benefits of SOA. The two main benefits of SOA, reuse and integration, play a key role in the results:

- Integration of data collection for BI can be improved, as SOA provides the means to connect to a large variety of (operational) systems, internal and external to the organization.
- Integration of BI applications with operational systems for distribution of information can be improved, increasing the possibilities to distribute information to more people in the organization, participating in different business processes.
- Components of BI-applications are provided means for improved integration, and therefore also opportunities are created for better componentization of BI systems.
- Reuse of data sources, once exposed as services, can be facilitated. The same accounts for reuse of information produced by BI.

So SOA can be used for the collection, analysis and distribution of information. The use of an SOA, or building on existing components of an SOA, can improve the speed and ease of building BI-applications through improved reuse and integration opportunities.

The results indicate that more opportunities apply to the operational and tactical level than to the strategic level, for which the explanation of our experts is that BI is more often used for these two levels. Concerning the focus areas, the overall opinion is that opportunities are not significantly differently valued across the areas.

An important identified limitation is that, with current technology, using services for the collection of data can be difficult when large amounts of data are involved. Another limitation that has been mentioned and agreed on by several experts, is that currently not enough standards exist for connecting different systems. On the other hand it is believed that further adoption of SOA will drive the developments of these standards and collaboration between vendors.

From the statistical analysis we can conclude that for many of the identified opportunities and limitations, opinions on their value and technical feasibility vary strongly. We think this is mainly caused by the novelty of and little experience with the subjects under consideration. We also think the smaller amount of people participating in the research than planned for have caused the high deviation in ratings, because a smaller group sample often results in a lower consensus. More discussion is needed between experts to understand each other’s disagreements and to come closer to an agreement.

7 Further investigation

The overview of opportunities and limitations that we presented in Chapter 6 contains 51 items. All but one of these items were identified in the first round, and evaluated in the second round. A small subset of these items has been selected for further evaluation in the third round, on basis of their high general rating, or low consensus. From the results of the third round, we have selected two opportunities and one limitation for further investigation, which are displayed in Table 23. The table shows the rankings (R) in the total list of opportunities and in the total list of limitations, the number (No.) of the hypothesis as in the overview of Chapter 6, the average general rating after the third round ($\bar{x}$), and the standard deviation (s).

For the two opportunities, Master Data Management (MDM) and Business Activity Monitoring (BAM), we further describe the concepts and technologies, and how these are related to SOA and BI. For the limitation, which is transportation of large data sizes through services, we discuss the details of this limitation, and to which kind of opportunities it applies.

The selection of the two opportunities is based on the amount of input provided by the expert panel during the Delphi study, the amount of information available in literature, and on our observation that these three subjects are related to many other identified opportunities. We have selected the limitation of large data sizes for transportation through services, because it proved to be the most serious limitation throughout the Delphi study.

The information that we use in chapter is derived from the Delphi study itself, by consulting additional literature, and by one-on-one interviews with a small number of our experts. One interview has been held with a consultant on the large data sizes limitation. An interview has been held on the same subject with a BI-vendor representative. Two interviews have been held with BI-vendor representatives on all three subjects.

Table 23: Selected opportunities and limitation for further investigation

R	**No.**	**Opportunity/Limitation**	**$\bar{x}$**	**s**
		Master Data Management (MDM)		
9	3	A good opportunity for applying SOA to the collection of data is by providing access to Master Data Management (MDM) based on SOA. Every layer of the BI cycle can provide input to and/or use MDM information.	1.9	1.4
		Transportation large data sizes through services		
1	13	An important limitation for applying SOA to the collection of data is the large data size involved, which is usually not suitable for transportation through services.	1.4	1.9
		Business Activity Monitoring (BAM)		
4	34	A good opportunity for applying SOA to BI on the operational level is in Business Activity Monitoring, to monitor business processes in (near) real-time fashion.	2.1	1.2

In Section 7.1, we discuss Master Data Management. In Section 7.2, we discuss the limitation of large data sizes for transportation through services, and in Section 7.3, we discuss Business Activity Monitoring.

7.1 Master Data Management

As we mentioned in Chapter 2, organizations own several different systems to support their operational business, and these systems often serve as data sources in the BI process, when data from multiple systems is integrated in the data warehouse.

Some operational systems keep track of the same data about the same entities. An example of such entities are products, customers, or employees. Data about these entities can be stored in systems used by, e.g., the finance department, the HR department, or the marketing department. Although these systems record data about the same identities, they often have different definitions on what constitutes such an entity. Therefore the format and contents of the data can differ per system for the same entity.

Sometimes data about a single entity is needed that is stored in different systems, for example to provide a full profile on a customer. In such cases, the relations between the data of those systems should be known to identify the same customer on each system. Also, when an entity is changed in one system, e.g., an case of an address change of a customer, this change needs to be reflected in systems that store this data as well (Dyché and Levy, 2006).

In this section we discuss Master Data Management (MDM), which is an approach to managing these data entities across an organization. We also discuss how SOA can support MDM, and how MDM can support BI.

7.1.1 Definition

Master data uniquely identifies a product, customer, or other business entity. For instance, a customer's first name and last name both represent reference data. Reference data can be generated by many different systems, which often keep redundant or contradictory versions of that data (Dyché and Levy, 2006). Master data is different from transactional data, but closely related. Master data represents the business entities around which the transactions are executed (Oracle, 2007). For example, when a customer buys a product, the transaction is managed by a sales application. The entities of the transaction are the customer and the product. The transactional data is the time, place, price, discount, payment methods, etc. used at the point of sale.

Organizations can put an infrastructure in place to support the sustainability of their master data, in order to ensure that throughout the organization a single version of each business entity is available. The practice of realizing and managing this infrastructure is called Master Data Management (MDM). Dyché and Levy (2006) define MDM as follows:

"Master Data Management is the set of disciplines and methods to ensure the currency, meaning, and quality of a company's master data within and across various data subject areas."

MDM is about managing the data within an organization, by setting standards and synchronizing the data between the systems that record this information. MDM allows to uniquely identify entities by reconciling data across different systems, promising a single, trusted source of master data (Dyché and Levy, 2006).

7.1.2 Functionality

To explain how an MDM system works, we take the entity “customer” as master data example. Figure 13 depicts the connection of the MDM system to the operational systems that manage data of a customer, such as a finance system that keeps track of the bills for a customer, a CRM system that keeps track of the service calls of the customer, and the sales system that keeps track of the purchased products of the customer.

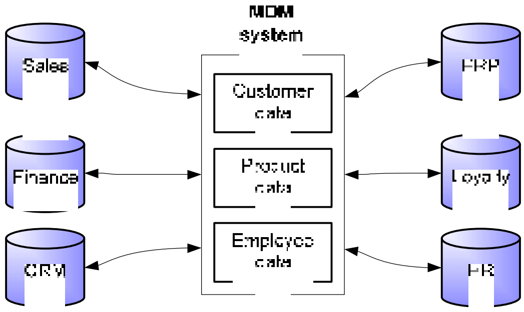


Figure 13: MDM system example for customer data

MDM simplifies the integration functionality necessary to support movement of master data between these systems, since it forms a central system to which all systems are connected to exchange this data. MDM keeps a record of the details about the data each operational system stores, and how this data maps to the master data record. When, for example, the address of a customer is changed in the sales system, this change is forwarded to the MDM system, which updates its master data. Then, the MDM system forwards these changes to the other operational systems that store the address of this customer.

According to Dyché and Levy (2006), the core functionality of an MDM system provides:

- A single point of data retrieval;
- Consistent value representation;
- An accurate and repeatable means of merging data;
- A repository of clean, reliable master data;
- Support to multiple data sources.

This functionality can be offered in three main types of MDM systems, each differing in the amount of master data they record. We describe these types, as identified by Dyché and Levy (2006) below.

Registry style. Within this type of MDM, the master data remains on the source systems, and the MDM systems only keeps records of the meta data of operational systems, and how this data relates to the definition of the master data. The actual data is only stored in the operational systems, not in the MDM system. This type of MDM system is advocated for its ease of implementation and high data integrity, because only a virtual master data record exists. The downside of this style is the performance of the system: transformation, cleansing, and integration of the data must occur for each master data management request.

Persistent style. Within this type of MDM, all data of an entity is copied into the MDM system, which then serves as a centralized storage platform for master data. Because all data can be acquired and merged before requests on the systems are done, this type of MDM is much faster than the registry style MDM. On the other hand, more data needs to be transported, more storage is needed to store the data, and the system requires rigorous data synchronization and quality checking processes.

Hybrid style. The hybrid style of MDM represents a combination of the registry and persistent style, in which some master data remains on the source systems, while other data is copied in the MDM system. On basis of the importance of certain data elements it can be decided which data needs to be stored in the MDM system, and which data only needs to be referenced.

7.1.3 MDM and SOA

Dyché & Levy (2006) state that SOA is very suitable for implementing MDM, since SOA provides the means for connecting all involved systems. Most of the current MDM products support SOA, and are designed for integration into a company's existing SOA architecture, by providing web service interfaces.

Karel & Fulton (2008) concludes that 64% of the enterprises that have an enterprise-level strategy around SOA have also considered MDM either as a priority (39%) or a critical priority (25%). According to Karel & Fulton (2008), SOA can support MDM in two ways:

- SOA can be a delivery system for MDM. Through services, master data can be collected from and exposed to different systems, providing the means to connect these systems.
- SOA provides an integration model for building the MDM system itself. The system can be built by composing data cleansing, updating, and matching services. This is also where the *transformational services* that were identified in the Delphi study come into play.

MDM can also support SOA, by addressing data quality and integration issues from an organizational point of view (Karel & Fulton, 2008). As we mentioned in Chapter 3, one of the SOA principles is that services should be represented by a well-defined interface. Among others, such a definition describes the exposed data. Inconsistent data definitions across several systems and poor data quality of (some of those) systems poses challenges to a well-defined interface, because multiple definitions of data exist.

MDM emerges from the need to improve data quality (Karel & Fulton, 2008). A requirement for successful MDM is to get stakeholders together to agree on data definitions of business entities across the organization, which improves the quality of the delivery of data through services.

7.1.4 MDM and BI

The value of MDM for BI is in the source data it can provide to BI. According to Karel (2006), accessing master data for BI is often a primary driver for early-phase MDM initiatives. MDM can provision integrated data to a data warehouse in a more reconciled and timely manner than standard data collection techniques, since (part of) the needed data for BI can be collected from a single source that is constantly updated.

MDM can also benefit BI by improving data quality. Data not recognized by BI tools as belonging to the same entity can lead to misleading results. As we mentioned, MDM serves to improve the data quality of entities within an organization. Master data represents the business entities around which transactions are executed, but also represents the key dimensions around which BI analyses are performed. To have a single definition of entities throughout the operational systems means that the BI system can also identify the transactional data related to these entities and improve the quality of analyses (Oracle, 2007).

However, two restrictions of using MDM for BI apply:

1. The design of a data warehouse differs from an MDM system. An MDM system should not simply be regarded as a faster server designed for real-time reporting on complex, integrated data. Because it is designed to serve the operational systems it is ill suited as a full BI platform (Dyché & Levy, 2006). Data still needs to be extracted from the MDM system to the data warehouse.
2. MDM only stores master data, not transactional data. Because many BI analyses are based on transaction history, this data still needs to be extracted from operational systems directly.

Figure 14 illustrates how the MDM system could communicate with operational and BI systems. MDM systems generally support several types of data transportation (Dyché & Levy, 2006), but as we mentioned, services can be very suitable.

When considering an MDM system as a data source for BI, the choice for the type of MDM system can be important. If fast data access is needed, a persistent style of MDM would be preferred.

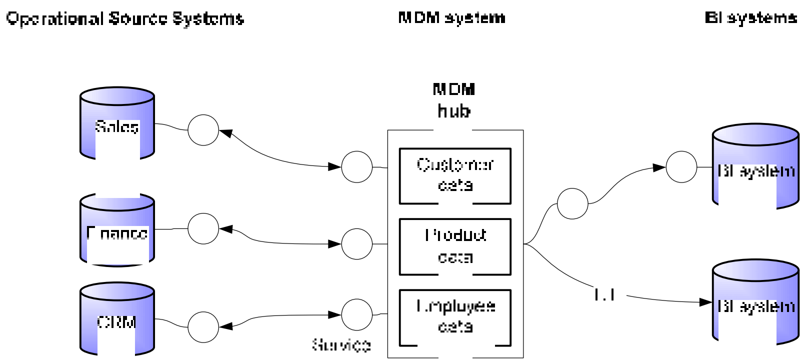


Figure 14: example of how MDM serves both BI and SOA

7.1.5 Conclusion

In our Delphi study, MDM has been rated as a good opportunity for using SOA technologies and concepts to build BI applications. From our investigation we can conclude that the relation between SOA and BI for this opportunity is indirect.

SOA can benefit MDM systems by providing a suitable delivery system for communication of the operational systems with the MDM system, and by providing services to build the MDM system. MDM can provide BI with integrated and timely master data. Since MDM serves to improve data quality within an organization, it can also improve the consistency of BI analyses. The improvement of data quality also benefits SOA in defining the interfaces of its services.

7.2 Large data sizes

In the Delphi study, several opportunities were identified for using services as data sources for the data collection for BI. The general opportunity of using services as data sources received a high rating, as well as some of the benefits, like the following:

- External data sources are easier accessible when exposed as services;
- Services as data sources can more easily be reused;
- Services support data to be accessed in a (near) real-time, event-driven model.

However, the technical feasibility of using services as data source for BI has been rated much lower. An important limitation to the use of services as data sources, as identified by our experts panel, is caused by the large data sets often involved in the collection of data for BI. Although some experts did not agree with this limitation, the majority agreed it is a serious limitation for the use of services for collecting data for BI.

According to our expert panel, this limitation applies to the current technology most associated with SOA, namely web services. Therefore, in our assessment of this supposed limitation, we focus on the web services technology.

7.2.1 Message exchange vs. data loading

Commonly, web services are described as providing a messaging framework (Erl, 2005). The messages transferred between service requester and provider consist of rather small amounts of data. For example, when a service requester needs to know the checking account balance of a certain customer, the request consists of a small-sized message describing the customer. The response from the service provider will also consist of a small-sized message containing the account balance for that specific customer (Colan, 2004a).

Collecting data for BI, however, often involves large amounts of data. For example, at the end of the day, a data warehouse might need the data of all the transactions that have taken place in a certain operational system. The request may be to get all data of the customer transactions at a supermarket, including customer information, products purchased, time of purchase, etc. This can involve millions of transactions, which results in large sets of combined data.

Traditional methods of loading data from operational systems into a data warehouse are designed to cope with these large data sizes. However, services are commonly not designed with these goals in mind. When services come into play for loading operational data into the data warehouse, two main approaches can be taken:

1. Loading the data from the operational databases where the transactional data is stored, or;
2. Loading each transaction as they take place – the so called event-based approach.

In the following two sections, we discuss each of these approaches.

7.2.2 Loading sets of data

In Figure 15 we illustrate the collection of data through services in comparison with the collection of data the traditional way. The traditional approach to collecting operational data is by making a direct connection from the data warehouse to the databases of the operational systems using a stream transfer, or by transferring a file using the File Transfer Protocol (FTP) (Kimball & Ross, 2002).

When web services are used for transferring this data to the data warehouse, the database should expose its data as a service, and the data warehouse should be able to make requests on this service. When data is transported from the operational database to the data warehouse, the data should be wrapped in an XML message, transported to the receiving service of the data warehouse, and then unwrapped and stored in the data warehouse.

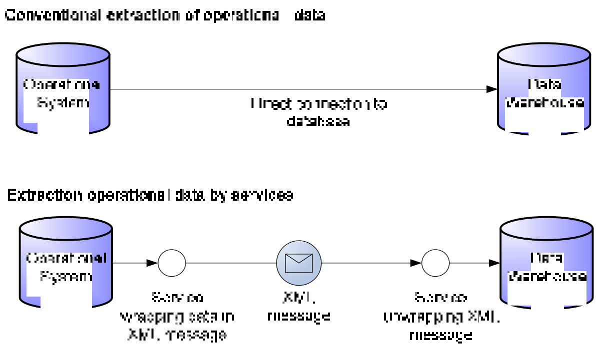


Figure 15: Comparison of extraction operational data by conventional methods and services

Commonly, web services in an SOA communicate through messages, which contain plain text formatted in structured XML. Taking this approach in transporting large amounts of data can lead to performance issues because of two reasons:

1. Wrapping and unwrapping of the data into XML messages requires considerable additional processing power (Bosworth et al., 2003; Ying, 2005).
2. Wrapping data into XML messages as plain text, considerably increases the size of the data, and therefore requires more bandwidth for transportation of this data (Mertz, 2001; Ying, 2005).

Solutions to these performance issues could be established by transferring the data as an attachment to an XML message, instead of including it as plain text inside the message. The data would then be packaged as one or more binary files, which are smaller in size. Several standards dealing with attachments have been proposed for the purpose of transporting large data sets, such as WS-Attachments (Bosworth et al., 2003) and SOAP with Attachments (Bosworth et al., 2003). According to the respective authors, using these standards can considerably improve the efficiency of transporting large files through services. However, two issues apply to this approach:

1. Although these standards are official, they are not yet widely accepted by all major vendors (Bosworth, 2005; Oracle, 2005), which can lead to other transportation issues.
2. When data is packaged in an attachment, not represented by a plain text XML message, it is more complex for a service requester to read the data. The service that needs to be designed to expose the data from the operational database will be different from the commonly used services in an SOA, which are used in a small message sized manner.

Services built for the purpose of transporting large amounts of data from the operational database to the data warehouse, are probably not suitable for purposes other than that. This approach is merely a replacement of transferring a large file through FTP. A standardized service design for this purpose could improve integration opportunities, but probably more suitable web services technologies for transferring large amounts of data would be needed and generally accepted. However, we have not investigated this limitation in enough detail to make strong claims about feasibility and benefits of this approach, and are of the opinion that further research into this limitation is needed.

7.2.3 Event-based loading

Another approach to data collection is to collect data in real-time, as it has been proposed by several experts in our Delphi study. Data involved in a transaction is then stored in the data warehouse just after the transaction takes place. Figure 16 illustrates this approach, which involves the use of an Enterprise Service Bus (ESB), as we mentioned in Chapter 3. Each time a transaction takes place, an operational system can publish event notifications on this ESB. These notifications are formatted as messages. Other systems or services can subscribe to these events, so that messages are pushed to these *event consumers* each time an event takes place.

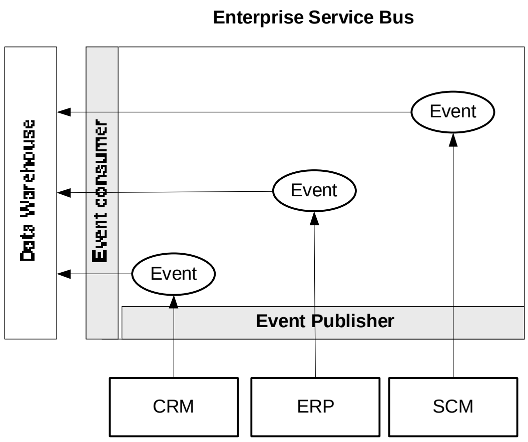


Figure 16: Event-based data collection

The data warehouse could be one of those event consumers. Each time a transaction takes place, the application that handles the transaction can publish a message on the ESB that contains the transaction details. The data warehouse that is subscribed to messages from this application can immediately process the information involved in the transaction.

Although several experts mentioned this approach for overcoming the limitation of services to handle large data sets, not much experience has been gained yet with the approach. One of the experts mentioned a project for a large telecom provider, in which on a daily basis billions of transactions take place. Because the conventional method of batch-wise loading of all this data into the data warehouse would lead to performance problems, it was decided to load the data of each individual transaction event-based. However, this specific approach was not implemented using web services. Experience with viable approaches through web services has not been shared by any of the participating experts.

Benefits can be identified however for this approaches once successfully applied. First of all, operational data is transported to the data warehouse in a (near) real-time fashion, resulting in an up-to-date data warehouse. Secondly, the operational systems are not burdened anymore with the performance sensitive requests of the data warehouse, because the data warehouse does not have to connect to these systems anymore.

7.2.4 Conclusion

As proposed by our panel of experts, using services as a data source can have several benefits, like easier reuse of these services, easier access to data external to the organization, and data access in an event-driven model. However, the same panel also identified a limitation to the use of services as data source. When large sets of data have to be transported, web services might not be well suited for this task.

Using services to transport data from an operational database to a data warehouse increases performance requirements when the data is transported as plain text in structured XML. Using XML attachments, which is a better approach to handle large data sets, does currently not seem to provide benefits over the traditional file transfer approach, because services that need to be designed for this purpose will be different from the services commonly associated with SOA. This implies that the data access for BI should for now be performed like it is being done now, for example, via a direct connection to operational databases, until more suitable web services technologies for transferring large amounts of data are available and generally accepted.

Another approach to loading large amounts of data into the data warehouse is event-based messaging, in which the data involved in transactions can be loaded (near) real-time into the data warehouse. This could would have benefits over the traditional batch-wise approach. The data warehouse will be up-to-date, and the operational systems will be relieved from the performance sensitive data loading process. However, we are not aware of any successful implementations of this approach.

7.3 Business Activity Monitoring

As we mentioned in Chapter 2, *Business Activity Monitoring* (BAM) provides (near) real-time information on the state of the operational process. This enables fast steering actions by the people directly controlling this process and informs higher management of the most current state of the operational processes (Den Hamer, 2005). In our Delphi study, BAM received a high general rating as an opportunity for the operational process that can be supported by SOA. Also, the overall opinion is that this opportunity is technically feasible.

7.3.1 Definition

To be able to steer an operational process, information must be available about that operational process. The faster information is available, the faster it is possible to respond to certain events or trends in the operational process. According to Khoshafian (2007), Business Activity Monitoring (BAM) is the software that monitors, correlates, and allows users to respond to operational events in an organization. The general objective is to help continuously improve process productivity and operational response, by alerting operational employees on situations requiring their attention, providing them the information required to assess the situation, and allowing them to take corrective action (Peyret, 2005).

The presentation of the current state of operational processes is often realized through so-called *dashboards*, in which the information is presented in constantly updated graphical reports, in which users can drill-down to more detailed information (Khoshafian, 2007). An example of such a dashboard is depicted in Figure 17. Although dashboards are often associated with BAM (Eckerson, 2006; Leymann, 2002), also other forms of output can be used, like e-mail or cell phones, e.g., to alert an employee about a new situation in the operational process (Folinas 2007).

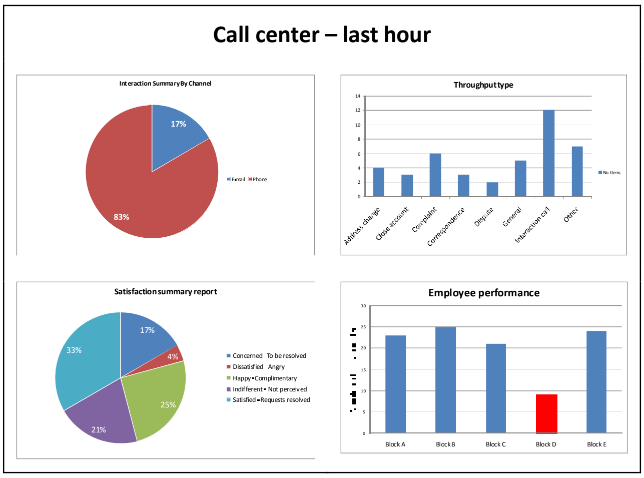


Figure 17: Example of a BAM dashboard (based on Khoshafian, 2007)

7.3.2 Business Process Management

BAM is often associated with Business Process Management (BPM), which takes a broader perspective on business processes than BAM. Aalst et al. (2003) define BPM as follows:

"Supporting business processes using methods, techniques, and software to design, enact, control, and analyze operational processes involving humans, organizations, applications, documents and other sources of information."

Figure 18 illustrates the lifecycle of BPM, which consists of four phases:

1. In the design phase, processes are (re)designed;
2. In the configuration phase, designs are implemented by a process aware information system;
3. In the enactment phase, the processes are executed;
4. In the diagnosis phase, the operational processes are analyzed to identify problems and points for improvement.

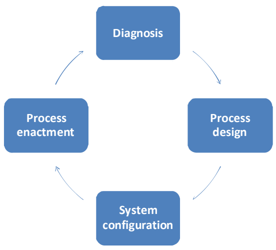


Figure 18: BPM lifecycle (Aalst et al., 2003)

BAM supports the diagnosis phase, for monitoring the newly designed processes (Koshafhian, 2007; Aalst et al., 2003). Therefore, BAM is often offered as a built-in component of BPM products. However, BAM also extends the functionality of BPM for monitoring, and can be used as a separate product (Koshafhian, 2007; Peyret, 2005).

7.3.3 Components

A BAM system is typically organized according to the BI-cycle. We discuss the components of BAM according to the collection, analyzing, and distribution phase.

Since the processes that need to be monitored can span several operational systems, collecting data from these systems involves integration of this data. Opposed to traditional BI, the integration of data should proceed in real-time, to be able to provide real-time information. Instead of integration at the data level, like a data warehouse that at intervals connects to a database to collect the data, integration at the application level is much more amenable to real-time purposes (Brobst, 2002). Integration at the application level is performed through message exchange, which takes place directly after an event has taken place (Eckerson, 2006; Khoshafian, 2007).

Figure 19 illustrates the Enterprise Service Bus, to which several systems can publish their events. The BAM system can subscribe to messages from these producers, and receives a message each time these producers publish a message to the ESB. Other viable options exist to manage events, like built in support in a BAM system (Khoshafian, 2007), but we regard the ESB as an appropriate concept to explain the management of events.

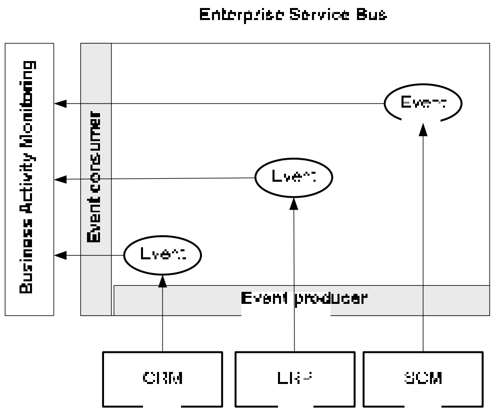


Figure 19: Enterprise Service Bus (ESB) supporting BAM

The event-driven message system required for BAM is well supported through the SOA message-based communication. This has been mentioned by several experts in our Delphi study. This is also supported by Nguyen et al. (2005) and Eckerson (2007), who argue that real-time BI is often associated with integration techniques like SOA. According to Peyret (2005), a number of vendors are currently working on opening up their BAM products to SOA.

During the analysis of data, the BAM system correlates the events that are received from the ESB, and generates information on the current state of the monitored processes. Analysis also includes correlating the events to certain *business rules* that can be defined by users in the form of alerts, targets, and thresholds (Eckerson, 2006). For example, notifications can be sent when certain critical events have occurred, such as an inventory stock that reaches a certain limit.

The information produced can be sent to several destinations, of which the operational dashboard is the best known example. The communication with dashboards or other devices can be implemented by connecting directly to the BAM system, but BAM can also act as an event producer itself by publishing events to the ESB, which can then be consumed by other applications.

Some experts mentioned the use of traditional data loading techniques in some current implementations of BAM, but with very short intervals of checking for new data at the operational systems. However, the event-based approach is also supported by most of these products, and could promise increased benefits for the future.

7.3.4 Current support

The overall opinion from the Delphi study that realizing BAM through SOA is technically feasible, might be related to the amount of information that needs to be transported and stored. Because the focus of BAM is on monitoring the current situation, only the current events need to be transported from the operational systems to the BAM application, and keeping a large history of data is not necessary. According to Eckerson (2006), BAM systems save about two weeks of data at the most.

Integration of the BAM system with operational systems, however, does pose some difficulties to the implementation of BAM. The BAM system should be able to communicate with operational systems by receiving their event-based messages. Consequently, BAM systems should be able to receive these messages, and operational systems should be able to send them.

Unfortunately, not every operational system can communicate at the message level. Implementing such an interface is possible, but involves a large amount of coding to build (Brobst, 2002). Furthermore, according to Peyret (2005), few applications that support event-massaging are actually able to publish all events that BAM needs.

Another issue is the compatibility of the BAM products themselves. Peyret (2005) states that most BAM products can only collect information from BPM tools from the same vendor, and not from other operational systems. However, according to Peyret (2005) some vendors already introduced new integration capabilities of their products, and by now these capabilities should have improved even more. This has been confirmed by some of our experts as well.

7.3.5 Conclusion

BAM has been identified in our Delphi study as one of the better opportunities for SOA to support the building of BI applications. In this section, we have described BAM in more detail by a short literature review. We confirm that the event-based messaging often associated with BAM is well supported by SOA, for both the collection and distribution of data. To have an SOA in place in an organization, will therefore support the implementation of a BAM system, which can be built based on the SOA in place.

The issues identified for BAM, such as the lacking support of some operational systems for a message-based interface, could already been solved when an SOA is already in place that integrates these applications.

8 Final remarks

In this chapter we conclude this thesis with some final remarks. In Section 8.1, we discuss the conclusions of our research findings. In Section 8.2, we present our recommendations for organizations and BI vendors. In Section 8.3, we discuss the limitations of our research, and in Section 8.4, we present our recommendations for further research.

8.1 Conclusions

In this research, we have sought an answer to the following research question:

What are the opportunities and limitations of using SOA concepts and technologies for building BI applications?

To answer this question, we first performed a literature study on the concepts and technologies of SOA, and on the concepts and technologies of BI. After that, we applied a Delphi study to a panel of experts in order to generate and discuss opportunities and limitations. In this Delphi study, we used models from our literature study for guidance.

The result of this Delphi study is a ranked list of opportunities and limitations, which include the arguments and comments from the experts. We have further investigated a small subset of these opportunities and limitations, by using the comments from the expert panel, having private interviews with some of our experts, and consulting additional literature.

Our research method can be considered successful, since it produced a list of 50 opportunities and limitations of using SOA concepts and technologies for building BI applications. This list actually answers the main research question. Below, we discuss the conclusions of our literature research and our general findings from the identified opportunities and limitations.

8.1.1 Concepts and technologies of SOA and BI

We identified BI as the process of gathering and analyzing data, and using the produced information to steer the organization. The process consists of four phases, which together comprise the BI-cycle: (1) planning and direction, (2) collection of data, (3) analysis of data, and (4) distribution of data. We also identified the areas of the organization for which BI delivers information, and what this information is about. We started from the observation that the components of a traditional BI architecture center around a data warehouse, which holds all the information extracted from the operational system.

We defined SOA as an application architecture within which all applications logic is defined as services, which can be called in defined sequences to form business processes. We discussed how services interact, and to what principles they should adhere. The main benefits that would justify the use of SOA are improved possibilities for both reuse and integration, which could result in increased agility and adaptability of the IT organization.

8.1.2 Opportunities of SOA for BI

In total, 40 opportunities, 10 limitations, and one general remark were identified and rated. The majority of the opportunities that have been identified and rated correspond to the characteristics and benefits that we identified for SOA:

- Reuse is identified as an opportunity for the collection of data, as well as for the distribution of information. Once services expose certain data or information, it could be easier to use that data or information in new projects.
- Integration opportunities are identified for the collection of data. SOA is perceived as offering capabilities for better integration of (operational) systems or systems external to the organization with the BI systems.
- Integration opportunities are especially valued for the distribution of information. If information produced by BI were accessible through services, it would be easier to distribute this information to many parts of the organization. This information could be used directly in operational systems, thereby integrating with the operational processes.
- Integration opportunities are also identified for components of BI systems. First of all, BI systems could be more componentized by the use of services, which could result in a more flexible BI architecture. BI components could become more interoperable by the development of standards, which might be driven by the increasing attention to SOA.

For the tactical and operational level more opportunities are identified than for the strategic level, probably because BI is most often used at these levels, and because they have a greater need for a flexible BI organization that can adapt to the more frequent changes at these levels. No considerable differences in the value of opportunities for the different focus areas have been identified.

We have further investigated Business Activity Monitoring (BAM). BAM supports the steering of operational processes by providing real-time information on the current state of these processes. BAM systems need to collect data from various operational systems, and SOA provides the means for integrating these systems. SOA supports event-based messaging, which enables real-time data collection.

Another embodiment of how SOA can support BI is through Master Data Management (MDM) . MDM serves to maintain a consistent definition of business entities throughout operational systems, and can also store the data of those entities in a central location. For managing these entities, operational systems should be able to publish changes on their entities to the MDM system, and be able to process changes in these entities that are published by other systems. SOA is identified as a suitable delivery system to integrate these systems with the MDM system. Furthermore, transformation and cleansing services could be used to build the MDM system. MDM can provide consistent and up-to-date data on entities to the BI system, and can improve the consistency of the data analyses.

8.1.3 Limitations of SOA for BI

The most serious limitation identified, which relates to several opportunities, concerns the transportation of large sets of data over services. Large sets of data often need to be transported from the operational systems to the data warehouse. Web services, often employed for the exchange of rather small messages, do not seem suitable for transporting large data sets. Standards have been defined that could lead to future solutions that can handle large data sizes, like the SOAP with attachments standard. However, more research should point out if these solutions can be helpful to BI. Another solution that could bypass the transportation of large data sets from the operational systems to the data warehouse is event-based communication. However, we have no knowledge of viable implementations of this concept.

8.2 Recommendations

Based on our conclusions, we have the following recommendations for managers that use BI for steering their organization, and for managers of BI vendor companies.

8.2.1 Organizations

Organizations can consider how SOA can play a role in the need for integration of BI with the operational systems. The needs for this integration should be identified, after which the SOA support for this integration should be considered.

Organizations that already have an SOA in place can consider the benefits they could achieve by opening up their SOA for BI purposes, by exposing data sources for BI, or by building services for the distribution of BI information to support integration of BI systems with operational systems. Organizations that consider investing in SOA, could take into account the benefits that can be achieved from a BI perspective, and consider these benefits in their decision to adopt SOA.

Either way, we advise to consider the overlap of SOA initiatives with BI initiatives, because mutual benefits might be realized. People leading the BI department could participate in decisions on SOA initiatives. BAM and MDM are examples for which cooperation between people working on SOA and people working on BI are beneficial.

Organizations should also realize that the strategy of building an SOA and hiding internal functionality behind services definitions, may not be applicable to the data. Web services are often considered inappropriate for retrieving the large amounts of data that BI needs. This implies that the data access should be performed in the way it is being done now, for example, via a direct connection to operational databases, until more suitable web services technologies for transferring large amounts of data are available and generally accepted.

8.2.2 BI vendors

We advise BI vendors to work on further integration possibilities of their products with an SOA environment, for both the collection of data and the distribution of information. Standards could be developed, or existing standards could be applied, for communication with the products and the components they consist of.

These integration opportunities especially apply to BAM systems, which need integration with multiple operational systems. By opening up these products to communication with more systems than only those from the same vendor, BAM would be more interesting for organizations to implement inside their organization.

BI vendors could consider offering MDM functionality and integrate this functionality into their BI products. MDM can provide consistent and up-to-date data on entities to the BI system, and can improve the consistency of the data analyses.

8.3 Limitations of the study

As with any Delphi-type study, the results are based on the ideas and opinions of a limited number of individuals. While we have chosen these individuals for their experience on the subject, we can make no claim about the representativeness of our expert panel. Having said this, we do believe that our expert panel is diverse, and that the experts had sufficient experience to participate in this study.

We can also make no claim about the coverage of the opportunities and limitations of the whole BI research area. Although we have used two models that are expected both said to cover the research area, this does not guarantee that all opportunities and limitations could be identified based on these models.

The results of our study pointed out that the opinion of experts varied heavily on many topics. Furthermore, the argumentations for the ratings that the experts provided were often given from very different perspectives. For future efforts, we think the Delphi study we employed could be improved. First, the size of the expert panel could be enlarged to increase the significance of the aggregated ratings. Secondly, we encourage more discussion between the experts to come to a consensus. We certainly see the benefits of a real-time Delphi study here, which enables face-to-face discussion and explanations of ideas. One could also consider performing the last round of the study by means of face-to-face or phone interviews, so that more and more relevant information can be received from the participants than what has been possible in writing.

Although originally 26 people agreed to participate in our study, the response rates were lower than we aimed for. We therefore advise for future Delphi studies to ask more experts to agree to participate in case the Delphi study is employed the way we did.

Another significant limitation has been the scarce literature on this area to support our investigation. Although we have found several sources to support our investigation for a small subset of the opportunities and limitations, this was certainly not the case for many others. The scarce literature is easily explained from the novelty of many subjects, and indicates a need for further research on this area.

In spite of the aforementioned limitations, we believe that the results of the study are useful for both research and practice.

8.4 Directions for further research

Although we believe the results of this study will prove useful, we identified several possibilities for further research into the topics that we have touched upon in our research. Some recommendations for further research are listed below:

- *Further research on opportunities and limitations identified in this study.* Most of the opportunities and limitations that are identified in this study are not further investigated, and are therefore open to further research. If these opportunities are actually valuable for practice, how they could be implemented, or for what type of BI they are most suited, could be further researched.
- *Research on the actual realization of SOA.* Currently little is known on the success of SOA implementations and to what extent the proposed benefits like increased opportunities for reuse and integration are realized. A study on (several cases of) SOA implementations could, first of all, provide more information on the issues experienced during SOA implementations, and secondly could inform about experienced benefits.
- *Research on the need for integration of BI into operational systems.* Integration of BI systems with operational systems has been identified in this study as a high valued opportunity. Further research can be performed on the actual need of organizations for integrated BI solutions, and for which type of organizations or processes the most benefits can be gained from this integration.
- *Further research on collecting data via services.* As we identified transportation of large data sizes as currently an issue for services, further research could be performed on possible solutions to this issue, and the actual benefits of such solutions over traditional data transportation methods.
- *Research on the feasibility of vendor interoperability for BI products.* Although SOA is identified as to drive standards, also for the development of standards for interoperability of products from different BI vendors, research could be performed on the feasibility of interoperability of the various components of BI products.

References

Aalst, W. M. P., Hofstede, ter, A. H. M., Weske, M. (2003). *Business Process Management: A Survey.* International Conference on Business Process Management, Springer, 2003.

Adler & Ziglio (1996) cited in Skulmoski, G. J. & Hartman, F. T. (2007). *The Delphi Method for Graduate Research.* Journal of Information Technology Education, Volume 6 (2007).

Agnar A. & Nygard, M. (1995) *Different roles and mutual dependencies of data, information, and knowledge- An AI perspective on their integration*. Data & Knowledge Engineering 16 (1995) pp. 191-222.

Akkermans, H. A., Bogerd, P., Yucesan, E. & Wassenhove, L. N. van (2003). *The impact of ERP on supply chain management: Exploratory findings from a European Delphi Study*. European Journal of Operational Research 146 (2003) pp. 284-301.

Bernhardt, D., C. (1994). *'I want it fast, factual, actionable'-Tailoring Competitive Intelligence to Executives Needs'.* Long Range Planning, Vol. 27, No. 1, pp. 12 to 24, 1994.

Bhatt, G. D. (2000). *An empirical examination of the effects of information systems integration on business process improvement.* International Journal of Operations & Production Management, Vol. 20, No. 11, pp. 1331-1359.

Bieberstein, N., Bose, S., Walker, L., Lynch, A. (2005). *Impact of service-oriented architecture on enterprise systems, organizational structures, and individuals.* IBM systems journal, Vol. 44, No. 4, 2005.

Bosworth, A., Box, D., Gudgin, M., Nottingham, M., Orchard, D., Schlimmer, J. (2003) *XML transport of binary data.* Paper by BEA Systems Inc. and Microsoft Corporation, 2003.

Brobst, S. A. (2002). *Enterprise Application Integration and Active Data Warehousing.* Proceedings of Data Warehousing 2002: From Data Warehousing to the Corporate Knowledge Center, Physica-Verlag Heidelberg, November 12-13, 2002, pp. 15-23.

Channabasavaiah, K., Holley, K. and Tuggle, E. M. (2004) *Migrating to a service-oriented architecture,* IBM DeveloperWorks, 2004.

Cherbakov, L., Galambos, G., Harishankar, R., Kalyana, S., Rackham, G. (2005). *Impact of service orientation at the business level.* IBM systems journal, Vol. 44, No. 4, 2005.

Colan, M. (2004a). *Service-Oriented Architecture expands the vision of Web services, Part 1.* IBM DeveloperWorks. http://www.ibm.com/developerworks/library/ws-soaintro.html accessed on December 4, 2007.

Colan, M (2004b) *Service-Oriented Architecture expands the vision of Web services, Part 2.* IBM DeveloperWorks, 2004.

Den Hamer, P. (2005). De *organisatie van Business Intelligence.* The Hague: SdU Publishers.

Dyché, J. & Levy, E. (2006). *Customer Data Integration: Reaching a Single Version of the Truth*. New York: John Wiley & Sons.

Driezen, W. (2008). Service-oriented and open-source software as potential markets for application management. Master thesis, University of Twente, January 2008.

Eckerson, W. W. (2006). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business.* New York: John Wiley & Sons

Erl, T. (2005). *Service-Oriented Architecture, Concepts, Technology, and Design.* Upper Saddle River, New Jersey: Prentice Hall.

Encyclopedia of Statistical Sciences (2006). *Multiple Range Test, Duncan's,* in Encyclopedia of Statistical Sciences (2006). New York: John Wiley & Sons, Inc.

Folinas, D. (2007). *A conceptual framework for business intelligence based on activities monitoring systems.* International Journal Intelligent Enterprise, Vol. 1, No. 1, 2007.

Gile, K., Moore, C., Hefner, R., Fossner, L., Eric. K. (2005). *The BI Architectural Innovations Shell Game.* Forrester Research, September 2005.

Hsu, C. & Sandford, B. A. (2007). *The Delphi Technique: Making Sense of Consensus.* Practical Assessment, Research & Evaluation, Vol. 12, No. 10, August 2007.

Inmon, W. (2002). *Building the Data Warehouse*. New York: Wiley & Sons.

Jamieson, S. (2004). *Likert Scales, How to (Ab)use them.* Medical Education 2004, Vol. 38, pp. 1212-1218.

Jukic, N. (2006). *Modeling strategies and alternatives for data warehousing projects. Communications of the ACM, April 2006, Vol. 49, No. 4.*

Kahaner, L. (1996). *Competitive Intelligence*. New York: Touchstone.

Karel, R. (2006). *Introducing Master Data Management.* Forrester Research, November 2006.

Karel, R. & Fulton, L. (2008). *How to make MDM And SOA Better Together.* Forrester Research, April 2008.

Khoshafian, S. (2007) *Service Oriented Enterprises.* Auerbach Publications.

Leeuw, A.C.J. de (1982). *Organisaties: management, analyse, ontwerp en verandering*. Assen: Van Gorcum.

Keen, M., Acharya, A., Bishop, S., Hopkins, A., Milinski, S., Nott, C., Robinson, R., Adams, J., Verschueren, P. (2004). *Patterns: Implementing an SOA Using an Enterprise Service Bus.* IBM Redbooks, July 2004.

Kimball, R. & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.* New York: John Wiley & Sons.

Leymann, F., Roller, D., Schmidt, M. T. (2002). *Web services and business process management.* IBM Systems Journal, Vol. 41, No. 2, 2002.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications.* Reading, MA: Addison-Wesley.

Melchert, F., R. Winter & M. Klesse (2004). *Aligning Process Automation and Business Intelligence to Support Corporate Performance Management*. Tenth Americas Conference on Information Systems, New York.

Mertz, D. (2001). *XML Matters: XML and compression.* IBM Developerworks, http://www.ibm.com/developerworks/xml/library/x-matters13.html accessed on May 12, 2008.

Mittereder, D. (2005). *Pervasive Business Intelligence: Enhancing Key Performance Indicators.* DM Review Magazine, April 2005. http://www.dmreview.com/issues/20050401/1023894-1.html accessed on May 7, 2008.

Moitra, D., Ganesh, J. (2005). *Web services and flexible business processes: towards the adaptive enterprise.* Information & Management, 2005.

Mylonopoulos, N.A., & Theoharakis, V. *Global perceptions of IS journals - Where is the bestIS research published?* Communications of the ACM (44:9), September 2001, pp. 29-33.

Nguyen, T. M., Schiefer, J., Tjoa, A. M. (2005). Sense & Response Service Architecture (SARESA): An Approach towards a Real-time Business Intelligence Solution and its use for a Fraud Detection Application. DOLAP'05, November 4–5, 2005.

OASIS (2006) *Reference Model for Service Oriented Architecture 1.0*. Committee Specification 1, August 2, 2006.

Oracle (2005). *Attachments in SOAP Messages, An Oracle White Paper.* Oracle, December 2005.

Oracle (2007). *Master Data Management, An Oracle White Paper. Oracle, November 2007.*

Papazoglou, M. P. and van den Heuvel, W. (2007a) *Service oriented architectures: approaches, technologies and research issue.* The VLDB Journal (2007) 16, pp. 389–415.

Papazoglou, M. P., Traverso, P., Dustdar, S., Leymann, F. (2007b) *Service-Oriented Computing: State of the Art and Research Challenges.* Computer, IEEE 2007.

Pérez, V. L. and Schüler, R. (1982). *The Delphi Method as a Tool for Information Requirements Specification.* Information & Management 5 (1982) pp. 157-167.

Peyret, H., Mendel, T., Le Quoc, K. (2005). *Business Activity Monitoring Slowly Prepares For SOA.* Forrester Research Inc.

Philips, E. & D. Vriens. (1999). 'Business Intelligence'. Deventer: Kluwer Bedrijfsinformatie.

Pressman (2001). *Software engineering: A practitioner's approach.* New York: McGraw-Hill.

Rowe, G. & Wright, G. (1999) *The Delphi technique as forecasting tool: issue and analysis* International Journal of Forecasting 15 (1999) pp. 353-375.

Schmidt, R. C. 1997. *Managing Delphi Surveys Using Nonparametric Statistical Techniques.* Decision Science, Volume 28, Number 3, 1997.

Scheibe, M., Skutsch, M. and Schofer, J. (1975). *Experiments in Delphi methodology*, in Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications.* Reading, MA: Addison-Wesley, pp. 262-287.

Schiefer, J., List, B., Bruckner, R.M. (2003). *Process Data Store, A Real-Time Data Store for Monitoring Data Processes.* DEXA 2003.

Shields, T.J., Silcock, G.W.H., Donegan, H.A., Bell, Y.A. (1986). *Methodological problems associated with the Use of the Delphi technique.* Fire Technology, Vol. 23, No. 3, August 1987, pp. 175-185.

Schmidt, R. C., Lyytinen, K., Keil M., & Cule, P. (2001). *Identifying Software Project Risks: An International Delphi Study.* Journal of Management Information Systems, Spring 2001, Vol. 17, No. 4, pp. 5 – 36

Schmidt, M.T., Hutchision, B., Lambros, P., Phippen, R. (2005). *The Enterprise Service Bus: Making service-oriented architecture real.* IBM systems journal, vol. 44, no. 4, 2005

Scholl, W., König, C., Meyer, B., Heisig, P. (2004). *The Future of Knowledge Management: an International Delphi Study.* Journal of Knowledge Management, Vol. 8, No. 2, 2004.

Skulmoski, G. J. & Hartman, F. T. (2007). *The Delphi Method for Graduate Research.* Journal of Information Technology Education, Volume 6 (2007).

Wikipedia (2008). *Mashup (web application hybrid).* http://en.wikipedia.org/wiki/Mashup_%28web_application_hybrid%29 Accessed on June 5, 2008.

Wu, L., Barash, G., Bartolini, C. (2007) *A Service-oriented Architecture for Business Intelligence.* IEEE International Conference on Service-Oriented Computing and Applications 2007.

Yang, Y. (2007). *Faster Data Transport Means Faster Web Services with MTOM/XOP.* http://www.devx.com/xml/Article/34797 accessed on May 12, 2008.

Ying, Y., Huang, Y., Walker, D.W. (2005). *A Performance Evaluation of Using SOAP with Attachments for e-Science.* Procedures of the UK e-Science All Hands Meeting, 2005.

Zhao, J. L. and Cheng, H. K. (2005). *Web services and process management: a union of convenience or a new area of research?* Decision Support Systems 40, 2005.

Appendix A – Duncan's MRT for separate categories

BI-cycle – Phase 2 (opportunities)

Number	N	Subset for alpha = 0.10	
		1	
12	15	1,20	
11	10	1,30	
10	13	1,62	
3	15	1,87	
4	13	1,92	
2	14	2,07	
Sig.		0,17	
Means for groups in homogeneous subsets are displayed.			

BI-cycle – Phase 2 (benefits)

Number	N	Subset for alpha = 0.10	
		1	2
9	12	0,33	
8	12	1,17	1,17
7	13		1,54
5	13		1,69
6	13		1,69
Sig.		0,20	0,46
Means for groups in homogeneous subsets are displayed.			

BI-cycle – Phase 4 (opportunities)

Number	N	Subset for alpha = 0.10		
		1	2	3
22	15	0,47		
23	12	0,75		
21	13	1,38	1,38	
20	11		1,91	1,91
19	15		2,33	2,33
18	15			2,53

Sig.		0,13	0,12	0,30
Means for groups in homogeneous subsets are displayed.				

BI-cycle – general (opportunities)

Number	N	Subset for alpha = 0.10	
		1	2
27	13	0,54	
26	12	1,08	1,08
25	14	1,29	1,29
24	15		1,87
Sig.		0,20	0,18
Means for groups in homogeneous subsets are displayed.			

BI functions – tactical level (opportunities)

Number	N	Subset for alpha = 0.10
		1
32	13	1,15
31	13	1,54
30	13	1,69
Sig.		0,24
Means for groups in homogeneous subsets are displayed.		

BI functions – operational level (opportunities)

Number	N	Subset for alpha = 0.10
		1
37	13	1,46
38	13	1,54
36	13	1,62
35	12	1,75
34	14	2,14
Sig.		0,25
Means for groups in homogeneous subsets are displayed.		

BI functions – focus areas (limitations)

Number	N	Subset for alpha = 0.10
		1
45	12	-0,83
44	13	-0,54
43	13	0,23
Sig.		0,18
Means for groups in homogeneous subsets are displayed.		

Catch-all question (opportunities)

Number	N	Subset for alpha = 0.10
		1
49	15	1,80
48	13	2,08
47	15	2,40
Sig.		0,20
Means for groups in homogeneous subsets are displayed.		

Appendix B – Significance difference means $2^{nd}/3^{rd}$ round

Kruskal-Wallis one-way analysis of variance test of difference between means of 2^{nd} and 3^{rd} round.

Ranks general rating			
	Round	N	Mean Rank
2	2	8	8,4
	3	8	8,6
	Total	16	
3	2	8	8,5
	3	8	8,5
	Total	16	
12	2	8	8,2
	3	8	8,8
	Total	16	
13	2	8	9,1
	3	8	7,9
	Total	16	
18	2	9	9,1
	3	9	9,9
	Total	18	
19	2	9	9,5
	3	9	9,5
	Total	18	
22	2	9	10,8
	3	9	8,2
	Total	18	
24	2	9	9,0
	3	9	10,0
	Total	18	
25	2	8	8,3
	3	8	8,7
	Total	16	
34	2	9	10,0
	3	9	9,0
	Total	18	
47	2	9	9,5
	3	9	9,5
	Total	18	
48	2	8	8,4
	3	8	8,6
	Total	16	
49	2	9	9,9
	3	9	9,1
	Total	18	

Ranks technical feasibility			
	Round	N	Mean Rank
tf 2	2	7	7,5
	3	7	7,5
	Total	14	
tf 3	2	8	8,7
	3	8	8,3
	Total	16	
tf 12	2	8	6,9
	3	8	10,1
	Total	16	
tf 18	2	7	6,3
	3	7	8,7
	Total	14	
tf 19	2	8	7,5
	3	8	9,5
	Total	16	
tf 22	2	7	7,7
	3	7	7,3
	Total	14	
tf 24	2	9	8,2
	3	9	10,8
	Total	18	
tf 34	2	9	9,5
	3	9	9,5
	Total	18	

Kruskall-Wallis test for significant differences between Round 2 and Round 3 means (general rating)													
	2	3	12	13	18	19	22	24	25	34	47	48	49
Chi-Square	0,01	0	0,07	0,30	0,15	0,00	1,10	0,17	0,03	0,25	0,00	0,01	0,11
df	1	1	1	1	1	1	1	1	1	1	1	1	1
Asymp. Sig.	0,91	1	0,79	0,59	0,69	1,00	0,29	0,68	0,87	0,61	1,00	0,91	0,74
a. Kruskal Wallis Test													
b. Grouping Variable: Round													

Kruskall-Wallis test for significant differences between Round 2 and Round 3 means (technical feasibility)								
	tf 2	tf 3	tf 12	tf 18	tf 19	tf 22	tf 24	tf 34
Chi-Square	0,00	0,03	2,20	1,27	0,76	0,06	1,09	0,00
df	1	1	1	1	1	1	1	1
Asymp. Sig.	1,00	0,86	0,14	0,26	0,38	0,81	0,30	1,00
a. Kruskal Wallis Test								
b. Grouping Variable: Round								

Appendix C - Stability Measurement

The following table is an example of how the Stability Measurement (Scheibe et al., 1975) is performed.

Table 24: example of stability measurement between round 2 and 3 for opportunity # 18

Rating	-3	-2	-1	0	1	2	3
Round 2 ratings	0	0	0	0	3	1	5
Round 3 ratings	0	0	0	0	1	3	5
Absolute difference in number selecting rating, rounds 2-3 (a)	*0*	*0*	*0*	*0*	*2*	*2*	*0*
Total units of change (b)	4						
Net person-changes (c)	2						
Number of participants	9						
Percent change (d)	22%						

(a) These numbers are the absolute differences between the second and third round in the total number of selections each rating option has.
(b) These numbers are the sums of the absolute differences in a.
(c) These are the net changes divided by 2. Each change made by a person always changes 2 numbers.
(d) Percent change is net person-changes (c) divided by the number of participants

Appendix D - Correlation general rating and technical feasibility rating

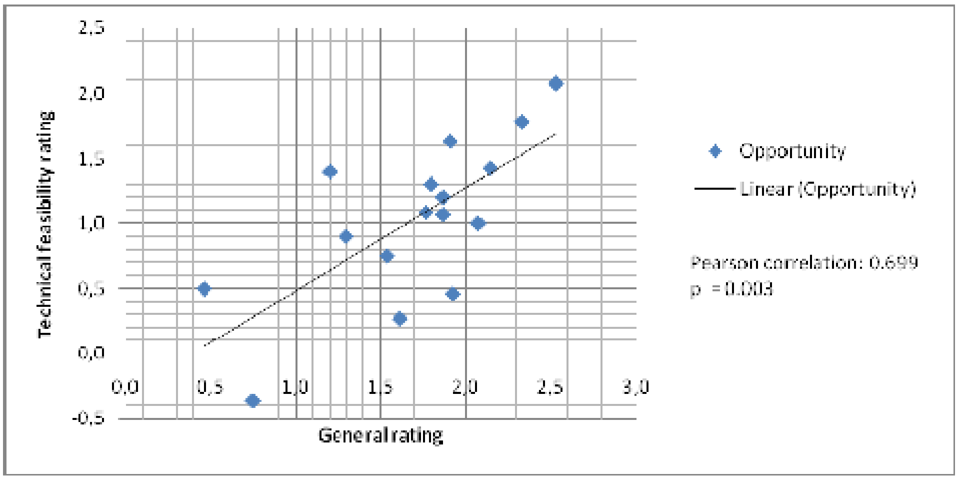


Figure 20: Correlation general rating and technical feasibility rating of an opportunity for general rating means and technical feasibility means

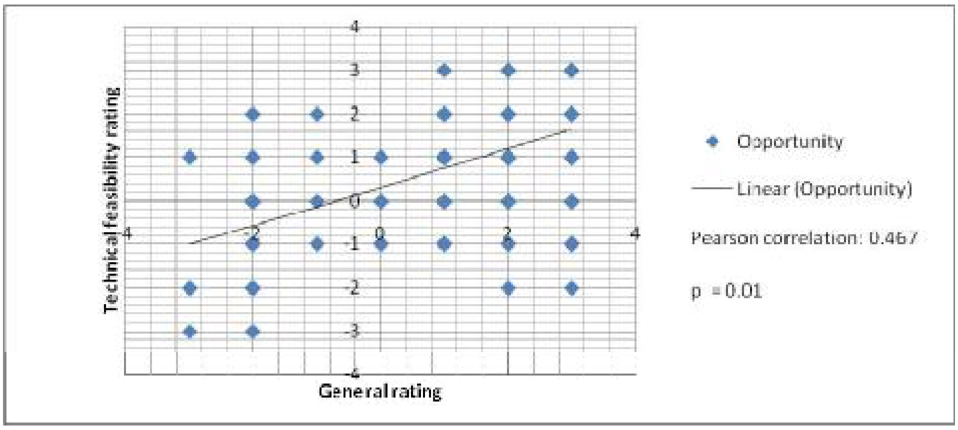


Figure 21: Correlation general rating and technical feasibility rating of an opportunity for all points of individual experts

Appendix E – Round 1 questionnaire

Opportunities and limitations of Business Intelligence in SOA-based Architectures: a Delphi Study

Introduction

Service Oriented Architecture (SOA) has gained increasing attention among both practitioners and academics. However, it is yet unclear what the increasing adoption of SOA can mean for Business Intelligence (BI), which is still mainly implemented based on non-SOA architectures. This study serves to provide insight in this matter.

This study is part of the final thesis I am writing for my masters degree on Business Information Technology at the University of Twente in the Netherlands. The thesis bears the title "Opportunities and Limitations of Business Intelligence in SOA-based architectures: a Delphi Study". I am supervised by Roland Müller, Assistant Professor of Knowledge Management and Information Services, and Luís Ferreira Pires, Associate Professor of Architecture and Service of Network Applications. At the consulting firm Capgemini, I am guided by Drs. Niels van der Zeyst.

This document is structured as follows:

- Section 2 provides an overview of this study.
- Section 3 introduces the main concepts of this study: BI and SOA.
- Section 4 contains the questions of this study.

Overview Study

Goal of this study
The purpose of this study is to produce an overview of the opportunities and limitations of BI in an SOA-based architecture. The overview will list the opportunities, and can include their technical realization and their business value. The results of this study will be further investigated, and will serve as foundation for my master thesis.

The first round of this study serves to identify initial ideas on the opportunities and limitations, which are elicited by means of questions on two BI models. The structure and contents of the second round is based on the output of the first.

Your role in this study
Since literature on this subject is rather scarce, experts with knowledge of BI, SOA, or both are approached to take place in an expert panel to discuss their ideas and opinions. As you are one of the participating experts, you are expected to provide your own ideas and discuss the ideas of others. For structuring the discussion, we have adopted the Delphi method, which enables experts to discuss a complex problem through a structured communication process (Linstone and Turoff, 1975).

Participating experts
The expert panel currently consists of 9 consultants of several consulting companies, 8 representatives of well known BI vendors, 3 clients of Capgemini using BI in their company, and 5 researchers affiliated to a university or research institute. Most of the participants have technical as well as conceptual knowledge of BI and/or SOA.

Communication method

Communication proceeds through documents sent by e-mail. In the beginning of each round, I will sent each participant the same questionnaire. After receiving all results, they are collated into one overview. This overview, including a new questionnaire, is then sent out again to each participant for the following round.

Time frame
The exact time frame of the study is –due to the nature of the study- unknown. We estimate that three rounds of questions will be necessary, with two weeks per round.

Confidentiality
Anonymity of participants is one of the features of the Delphi study, allowing the participants to freely express their opinions. Therefore, the individual responses are restricted to the expert himself, my supervisors as mentioned in the introduction, and myself. The final version of the thesis contains the name and employer of the participating experts, unless objected to, but the names will not be connected to individual statements in any way.

Concepts: SOA and BI

This section introduces the BI and SOA concepts. The descriptions of the concepts serve to create a common view and vocabulary for the expert panel.

Service Oriented Architecture

A concept that is gaining increasing attention from academics and practitioners is the Service-Oriented Architecture (SOA). The purpose of this architecture is to address the requirements of loosely coupled, standards-based, and protocol-independent distributed computing, for aligning information systems to business processes in a way that accommodates business agility.

In an SOA, software resources are packaged as “services”, which are modules that provide standard business functionality. Services can be regarded as building blocks that communicate with each other in order to collectively support a common business task or process. A service in an SOA is a piece of functionality with three main characteristics. First, the service is self-contained, which means that the service maintains its own state. Second, services are platform-independent, and therefore can communicate with services built on other platforms. Third, services can be dynamically located, invoked, and (re-)combined.

An SOA provides a flexible architecture by modularizing large applications into services. With SOA, an enterprise can create, deploy, and integrate multiple services and choreograph new business functions by combining new and existing application assets into a logical flow. In this way, SOA can deliver the flexibility and agility that business users require, defining services which can be aggregated and reused as the key building blocks of enterprises, to facilitate ongoing and changing needs of business. ***(Papazoglou, 2007)***

Business Intelligence

In steering their organization, companies experience an increasing need to obtain information about the processes and people within their company, about their environment, and about other factors influencing their business. Business Intelligence (BI) provides support for delivering this information. In this research, we define Business Intelligence as ***the process, supported by corresponding facilities, of gathering and analyzing data, and using the produced information to steer the organization.*** This section introduces the models that serve as a basis for the questions of the first round of this research.

BI Functions

To structure the broad scope of BI, we use the BI function model of Den Hamer (2005). The model, which is depicted in Figure 4, aligns the generally known functions of BI to two axes: the organizational layers, and the focus areas. The organizational layers consist of the strategic, tactical and operational level. The focus areas consist of the internal organization, customers, suppliers, market, environment, and competitors of the organization.

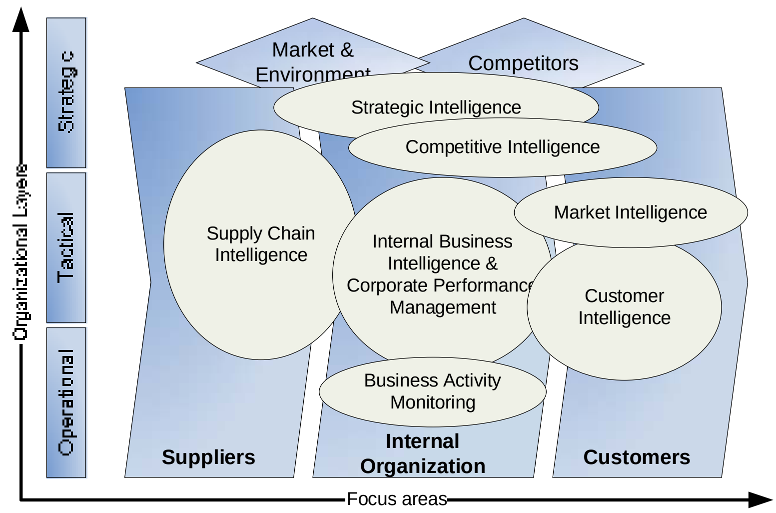


Figure 22: BI Function Model (Den Hamer, 2005)

The following BI functions are distinguished:

- *Internal BI* is the currently most used function of BI, which can serve all organizational levels, but most often serves the tactical level, by providing management information for steering operational processes.
- *Business Activity Monitoring* (BAM) mainly serves the operational processes. By continuously providing information on the state of the process, BAM enables a fast steering response by the people responsible for the process.
- *Corporate Performance Management* (CPM) focuses on providing internal information for the strategic level.
- *Market Intelligence* focuses on providing information from a general market perspective, e.g. on certain customer groups, regions or market developments.
- *Customer Intelligence* focuses on the individual customer, e.g. by providing a complete profile of the individual customer.
- *Competitive Intelligence* and *Strategic Intelligence* focus on information about the world outside the company. Competitive Intelligence informs on competitors and markets, and Strategic Intelligence extends this by providing multiple year overview on e.g. macro-economic trends, technological developments and government policies.
- *Supply Chain Intelligence* focuses on information about purchasing, logistics, and inventories of an organization. The information often comes *from* suppliers, or information is provided *to* suppliers.

BI cycle

The *BI cycle* (Den Hamer, 2005; Philips & Vriens, 1999; Kahaner, 1996) defines the basic steps of the BI process. The cycle consists of four phases:

1. **Planning and direction:** consists of structuring the BI cycle.
2. **Collection of data:** the needed data sources are identified, after which the data can be converted, edited, aggregated, and stored in a structured way.
3. **Analysis of data:** producing information, by providing a context to the data, or discovering patterns and connections in the data collection.

4. **Distribution of information:** getting the produced information to the right people in the right format.

Traditional BI Systems

BI delivers information by collecting and analyzing data, using corresponding facilities. We discuss here the traditional setup of these facilities, which uses a *Data Warehouse* (DW). Main components of the system are the operational source systems, the data staging area, the data presentation area, and the data access tools, as displayed in Figure 23.

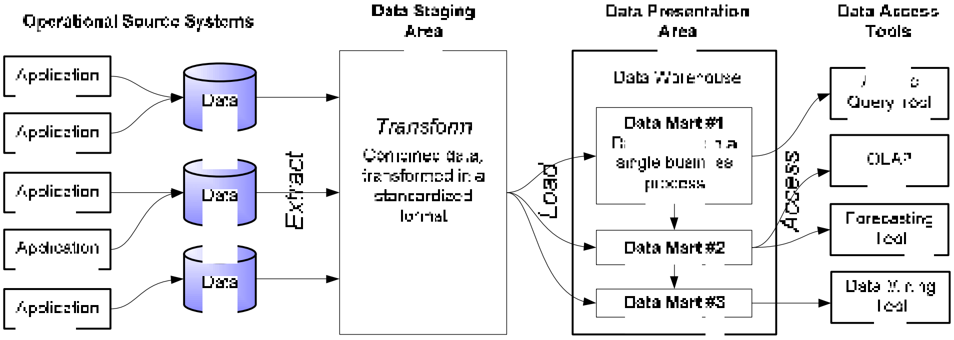


Figure 23: Basic elements of a BI system according to Kimball & Ross (2002)

Operational source systems are systems used by the business. Assumed is that all the needed data is already in these systems. Source systems maintain little historical data, opposed to the business intelligence system.

The data staging area is both a storage area and a set of processes that work on this data. These processes are called *extract-transform-load* (ETL). Extraction is the first step, in which data is copied from the operational source systems into the staging area. Transformation is the second step, in which the data is combined from the various sources and transformed in a common format. Loading the data into the presentation area is the third step.

The data presentation area is where data is organized, stored, and made available for data access from other systems. In the DW community, there are two different views on how this should be organized. Figure 23 displays the view of Kimball and Ross (2002), in which the presentation area consists of a series of integrated *data marts* that presents the data from a single business process in a multidimensional model. Figure 6 displays the view of Inmon (2002), where the presentation area consists of an integrated database containing all the data, which then serves as a source for *independent* data marts.

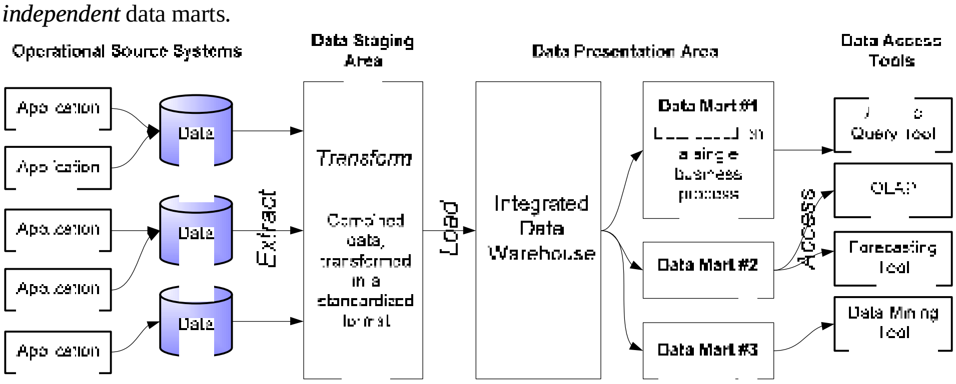


Figure 24: Basic elements of a BI system according to Inmon (2002)

Data access tools are the applications that actually use the data that is in the presentation area, to deliver the information for which the system is built. The type and complexity of data access tools can vary heavily. Examples of these tools are reporting tools, which process a standardized report, forecasting tools, which use the data to identify trends, or Online Analytical Processing (OLAP), to quickly provide answers to analytical queries.

Questions Round 1

This section contains the questions of the study. You can write the answers in this document, below the questions, preferably using another font color. In case the questions are not clear to you, please do not hesitate to contact me. My contact details are at the end of this document.

When answering the questions, please keep the following in mind:

- Any idea is open for discussion, which means that all ideas that come to your mind can be written down.
- Be concise in your writing, by putting down your ideas in one or a few sentences.
- Avoid mentioning vendor-specific products. This research focuses on general ideas that can be used by everyone.

Self-assessment

The following questions are meant to identify the collective experience of the panel of experts. Please fill in the following information:

Profession	
Number of years experience with BI (if applicable)	
Number of years experience with SOA (if applicable)	
Number of BI projects participated in (if applicable)	
Number of SOA projects participated in (if applicable)	

Publicity of identity

Would you mind being mentioned in the list of participating experts in the final version of the thesis?
Answer: Yes/No

Questions on the subject

Only a few questions are asked, but multiple answers can be given on each question. First, we focus on the BI-cycle, which presents the general BI process. Second, we focus on the BI-functions model, which presents the different business areas of BI. Both models can be viewed from a business perspective and a technical perspective. You are encouraged to take both views in your answers. You should first read through all the questions, before answering them, because there can be certain overlap in the questions.

BI cycle: opportunities of SOA for implementing BI

These questions are based on the BI cycle as presented in section 0, and the conventional BI architecture as presented in section 0. Please structure your answer per identified opportunity.

1. Which opportunities can you identify for applying SOA to the general process of BI? Use the BI cycle as a guideline and take phases together if your answer requires this.
   - Phase 1: Planning and direction.
   - Phase 2: Collection of data.
   - Phase 3: Analysis of data.
   - Phase 4: Distribution of information.
2. For each opportunity, can you shortly explain in words how this can be realized from an architecture perspective (components and their relations)?
3. How do you see these opportunities in relation to the conventional BI architecture as presented in section 0?

Please provide your answers, per identified opportunity, below (use as much space as you need):
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
………………………………………………………………

BI-functions model: answering the business value SOA/BI

These questions are based on the BI-functions model as presented in 2.4.

4. For which elements of the BI functions model do you think SOA provides opportunities? Take levels together if your answer requires this.
   a. Level of organization
      i. Operational level
      ii. Tactical level
      iii. Strategic level
   b. Area of focus
      i. Suppliers
      ii. Internal organization
      iii. Customers
      iv. Market and environment
      v. Competitors
2. For each opportunity, can you shortly explain in words how this can be realized from an architecture perspective (components and their relations)?

Please provide your answers, per identified opportunity, below (use as much space as you need):
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

……………………………………………………………………………………………………
……………………………………………………………

General
The BI-cycle model and the BI-function model are meant to structure the answers and encourage to think about several areas in which opportunities for SOA can be identified. However, because these models can also limit the type of answers, the following question is asked:

5. What other opportunities or limitations for BI in an SOA environment can you identify?

Please provide your answers, per identified opportunity or limitation, below (use as much space as you need):
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………

Sending back the document

First of all, thank you very much for filling out the document for this first round. You can send this document to my e-mail address: sefan.linders@capgemini.com. You will be notified of the results of this first round at the same time as you will receive the questionnaire for the second round. If you have any questions in the mean time, please don't hesitate to contact me.

Kind regards,

Sefan Linders

Contact details

Sefan Linders
Graduate Student at University of Twente
Enschede, The Netherlands
Tel +31 (0) 6 2336 4709
E-mail sefan.linders@capgemini.com

References

Den Hamer, P. (2005). De *organisatie van Business Intelligence.* The Hague: SdU Publishers.

Inmon, W. H. (2002). *Building the data warehouse.* Third edition. New York: John Wiley & Sons, Inc.

Kahaner, L. (1996). *Competitive Intelligence*. New York: Touchstone.

Kimball, R. & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.* New York: John Wiley & Sons.

Linstone, H. A., & Turoff, M. (Eds.). (1975). The Delphi method: Techniques and applications. Reading, MA: Addison-Wesley.

Papazoglou, M. P. and van den Heuvel, W. (2007a) *Service oriented architectures: approaches, technologies and research issue.* The VLDB Journal 16:389–415

Philips, E. & D. Vriens. (1999). *Business Intelligence.* Deventer: Kluwer Bedrijfsinformatie

Appendix F – Round 2 questionnaire

Delphi Study - Round 2

Opportunities and limitations of Business Intelligence in SOA-based applications

Introduction

This document presents the questionnaire of the second round of the Delphi study on Opportunities and Limitations of Business Intelligence in SOA-based Applications. This round is based on the BI-models as introduced in the first round, with exception of the models of traditional BI architectures. The collective output from round 1 is presented, which in this round will be rated by the same expert panel, enabling statistical analysis. As in the whole study, individual inputs of experts will not be published or connected to the name of those experts in any way.

In the first round, 25 persons (researchers, consultants, BI vendor representatives and clients from Capgemini) have been sent a questionnaire. Of those participants, 17 have responded by sending in the document with their answers. For this round, the original group of 25 persons is approached again.

The output of round 1 has been filtered on relevancy, rewritten to adhere to a common understanding of used definitions, and compacted to an acceptable size. Answers reflecting the same opinion have been combined into one answer.

The full contents of the introduction and questionnaire of round 1 is attached in the same e-mail as this document has been attached with. It may help you understand the context of this questionnaire, but is not required for answering this questionnaire.

Questions Round 2

The output from round 1 is presented as a list of hypotheses of opportunities and limitations, which in this round will be rated on agreement level. For answers involving a technical perspective, also the technical feasibility is rated. Please select your agreement level for each item. If you do not understand the question, e.g. because you do not think you have the right knowledge, select 'no opinion'. You are encouraged to provide comments, in case you think these are important arguments for your answer. There is a comment box provided below each group of hypothesis's.

Definitions in context of this study
Opportunity: concept or pattern that involves the use of SOA and can improve the organization (design, development, maintenance, etc.) and use of BI.
Limitation: restriction of using SOA for the organization and use of BI.
Technically feasible: capable of being implemented with a reasonable amount of effort, considering currently available technologies.

Assume mature SOA organization
For the mentioned concepts and patterns, you can assume that a mature SOA organization is in place, in which all data from operational systems is accessible through services.

BI Cycle

Phase 1: Planning and Direction

1 An important limitation of using SOA in the planning and direction phase is that this phase mainly consists of creative processes that are not very well structured and requires a lot of human interaction.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Do you have any comments on the planning and direction phase (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations?

Phase 2: Collection of data

2 A good opportunity for applying SOA to the collection of data is to use services as a data source.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

3 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

4 An important benefit of using services as a data source is that external data sources are easier accessible through services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

5 An important benefit of using services as a data source is that it enables reuse of these data sources. This makes the BI organization more agile in response to changing information needs.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

6 An important benefit of using services as a data source, is that by having an overview of the services for BI (e.g. though a services library) you have insight in the data available for BI, instead of the data being scattered and 'hidden' across the application landscape in non SOA environments.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

7 An important benefit of using services as a data source is that it is then easier to access data in a (near) real-time, event driven model.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

8 An important benefit of using services as a data source is that collecting data from published services ensures that there is a separation of concern and that the risk that the data is wrong is reduced. The responsibility for the quality of data is taken by a different project team who must have thought about it and dealt with it to have been able to publish a service definition.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

9 A good opportunity for applying SOA to the collection of data is by performing bulk synchronization between the operational systems and the BI systems, using conventional methods (which are able to handle the voluminous size of the data), and receiving regular updates via services. More up to date information is then available.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

10 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

11 A good opportunity for applying SOA to the collection of data is by using a distributed/virtual/federated data warehouse. Raw data can be distributed across a grid, and are accessed for analysis via grid services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

12 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

13 A good opportunity for applying SOA to the collection of data is by providing access to Master Data Management (MDM) based on SOA. Every layer of the BI cycle can provide input to and/or use MDM information.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

14 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

15 A good opportunity for applying SOA to the collection of data is by transformational services that fulfill the role of a transformation step of the conventional BI architecture. These services can increase the use of standard data formats.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

16 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

17 A good opportunity for applying SOA to the collection of data is by wrapping components like data validation and data cleansing as services, so multiple data integration and BI projects can reuse those components. The input should provide the source data needed to be processed and filter conditions that the service should apply to the source data. The output will be pre-processed data or a subset of changed data.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

18 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

19 An important limitation for applying SOA to the collection of data is the large data size involved, which is usually not suitable for transportation through services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Do you have any comments on the collection of data phase (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations for this phase?

Phase 3: Analysis of data

20 A limitation of using SOA in the analysis phase is that this phase is often very diverse and specific, limiting the role of services to pre defined queries and standard analysis (row count, null values, etc.)

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

21 A limitation of using SOA in the analysis phase is that analysis through services is only suitable for lower organizational levels, where decisions are structured more clearly.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

22 A good opportunity for applying SOA to the analysis of data is for communication between different components of the BI solution.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

23 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

24 A good opportunity for applying SOA to the analysis of data is that for providers of specific analytic functionality are provided a low friction way of deploying their specific expertise.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Do you have any comments on the analysis of data phase (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations for this phase?

Phase 4: Distribution of information

25 A good opportunity for applying SOA to the distribution of data is by exposing the produced BI information through services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

26 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

27 A good opportunity for applying SOA to the distribution of data is by exposing information for use in other applications (so-called closed-loop BI).

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

28 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

29 A good opportunity for applying SOA to the distribution of data is by wrapping complex queries and other BI functionalities, demanded by users, as services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

30 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

31 A good opportunity for applying SOA to the distribution of data is by using configurable services as data provider to the OLAP query process. The input can be a cube and filter conditions, the outputs can be a cube or a data mart.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

32 This opportunity is feasible using currently available technologies.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

33 A good opportunity for applying SOA to the distribution of data is the concept BI for All: using services to provide BI functionality and information to many different types of users throughout the whole organization, in an easy way

Strongly disagree ○ 3 ○ 2 ○ 1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

Do you have any comments on the analysis of data phase (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations for this phase?

Whole BI Cycle

34 A good opportunity for applying SOA to the BI process is by modeling and executing processes in the Business Process Management layer of SOA. This provides insight into process modeling and process execution information available in the application landscape and available for BI, instead of the process information being scattered and 'hidden' across the application landscape in non-SOA environments.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

35 The data/information flows in organizations from the operational source systems to data warehouse and from data warehouse to the end user are often very complex. Moreover, errors and mistakes in the components in the path of the information flows are easily hidden due to aggregations and statistical analysis. A good opportunity for applying SOA to the BI process is the self containedness of services, which allows better fault prevention and tolerance, and more flexibility in reconfiguring the information flows.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

36 A good opportunity for applying SOA to the BI process is by Information Integration based on SOA, which can decrease the latency of data (e.g. caused by overnight processing), by collecting, combining, analyzing and distributing information together with the information from the traditional data warehouse.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

37 A good opportunity for applying SOA to the BI process is that the use of services enables componentization of the BI architecture, to support loosely coupled BI systems.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

38 This opportunity is feasible using currently available technologies.

Strongly disagree ○ 3 ○ 2 ○ 1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

Do you have any comments on the BI Cycle as a whole (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

BI Functions - Organizational Levels

Strategic level

39 An important limitation for applying SOA to BI on the strategic level, is that BI on this level often involves large amounts of data, which is not suitable for use with services.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

40 A good opportunity for applying SOA to the strategic level is that using SOA for this level can save costs and reduce development time of BI solutions.

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ No opinion

Do you have any comments on the strategic level (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

Tactical level

41 An important limitation for applying SOA to BI on the tactical level, is that this often requires combining multiple sources. Processing these datasources would result in large amounts of programming.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

42 A good opportunity for applying SOA to BI on the tactical level exists for integrating data from multiple sources, which is easier through (platform independent) services.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

43 A good opportunity for applying SOA to BI on the tactical level is by creating a flexible architecture, as changes often happen at this level.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

44 A good opportunity for applying SOA to BI on the tactical level is that it enables shortening the time frame between demand and delivery of information.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

Do you have any comments on the tactical level (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

Operational level

45 A good opportunity for applying SOA to BI on the operational level is that it enables shortening the time frame between demand and delivery of information.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

46 A good opportunity for applying SOA to BI on the operational level is in Business Activity Monitoring, to monitor business processes in (near) real time fashion.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

47 This opportunity is feasible using currently available technologies.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

48 A good opportunity for applying SOA to BI on the operational level is by creating a flexible architecture, as changes often happen at this level.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

49 A good opportunity for applying SOA to BI on the operational level exists for integrating data from multiple sources, which is easier through (platform independent) services.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

50 A good opportunity for applying SOA to BI on the operational level is in closed-loop BI, transporting produced information directly into other applications and processes.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

51 This opportunity is feasible using currently available technologies.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

Do you have any comments on the operational level (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

All organizational levels

52 A good opportunity for applying SOA to BI in general is that on all organizational levels, business process improvement based on BI information is easier through SOA.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

53 A good opportunity for applying SOA to BI in general, is that through SOA application integration services, BI can turn into operational BI as one part of the IT systems of the organization. The input data should be source specification and refresh rate and subscribed data structure. The output can be CDC data (Change Data Capture, data changed in specific time period) for BI near real time BI processes for operational level and tactical level.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

54 This opportunity is feasible using currently available technologies

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

Do you have any comments on all organizational levels (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

BI Functions - Focus area's

55 Good opportunities for applying SOA to BI exist within the internal organization.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

56 Good opportunities for applying SOA to BI exist in the supply chain, when the processes are complicated and has a high volume of orders.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

57 An important limitation for applying SOA to BI for market & competitor analysis is that this is often an unstructured process and therefore mainly a human process, not offering opportunities for SOA.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

58 An important limitation for applying SOA to BI for market & competitor analysis is often delivered now and then, therefore processing batches would be more suitable than using services.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

59 A good opportunity for applying SOA to BI for customer intelligence only exists in an industry with a large number of customers.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

60 There is not much difference between the focus areas in opportunities for applying SOA to BI. Everywhere where data is transferred from one system to another this might be done through services. This can be within an organization but also between organizations.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

Do you have any comments on all organizational levels (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

General remarks

61 A good opportunity for applying SOA to BI is pervasive or ubiquitous BI, which means that BI will be used everywhere and in every process throughout the organization. Just like SOA can add business value by disconnecting the business process from the underlying applications, BI can be a component of this.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

62 A good opportunity for applying SOA to BI is in the creation of a more agile organization, which can react faster to changes in the business.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

63 A good opportunity for applying SOA to BI is in easier integration of other components with BI, like the data access tools.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

64 An important limitation for applying SOA to BI is that SOA is often vendor specific, which poses limitations to the use of BI and SOA.

Strongly disagree -3 -2 -1 0 +1 +2 +3 *Strongly agree* *No opinion*

64 An important limitation for applying SOA to BI is that SOA is often vendor specific, which poses limitations to the use of BI and SOA

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

65 An important limitation for applying SOA to BI is that the (Business) Process Focus is not very common in the BI community, but is important for working with SOA

Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Do you have any comments on opportunities and limitations for BI in SOA-based applications (e.g. providing arguments for your (dis)agreements), or can you identify additional opportunities or limitations that apply in general to the BI process?

In the previous questionnaire, you have been asked to provide some details on your profession and experience. Also the question has been asked if you would mind to be mentioned in the final thesis. Because this question was perceived by some as ambiguous, it is here rephrased, including your answer as interpreted. Please check your answer if provided, as well as the other items that can be used for your listing

Allows to be listed (with name, profession, and company) in thesis as participating expert	
Name	
Profession	
Company	
Number of years experience with BI (if applicable)	
Number of years experience with SOA (if applicable)	
Number of BI projects participated in (if applicable)	
Number of SOA projects participated in (if applicable)	

Sending back the document

First of all, thank you very much for filling out the document for this second round. You can save & send this document to my e-mail address: stefan.linders@capgemini.com. If you have any questions in the mean time, please don't hesitate to contact me.

Kind regards,

Stefan Linders

Contact Details

Stefan Linders
Graduate Student at University of Twente
Enschede, The Netherlands
Tel +31 (0) 6 2006 4709
E-mail stefan.linders@capgemini.com

Appendix G – Round 3 questionnaire

Delphi Study - Round 3

Opportunities and limitations of Business Intelligence in SOA-based applications

Introduction

This is the questionnaire of the third (and last) round of the Delphi study on Opportunities and Limitations of Business Intelligence in SOA-based Applications. In the second round, 10 persons have responded by sending in the document with their ratings. Some of them have provided comments to specific hypotheses or groups of hypotheses. This last questionnaire is sent out to the original group; we also encourage you to participate if you have not participated in previous rounds.

The items of the previous round that received a high rating will be further researched in this round, so we can elaborate on those in the thesis where this study is part of. Also included are some items that received a high spread of ratings, representing items on which opinions vary heavily, making them interesting for identifying the reasoning behind these variations.

Thirteen items from the previous round are repeated here: 9 identified opportunities with a high rating, 3 opportunities with a high spread of ratings, and 1 limitation with a high spread of ratings. We added 1 not yet rated opportunity. The items include:
- the hypothesis
- the general rating provided (the central tendency as the Mean (M) and spread of ratings as Standard Deviation (SD))
- if mentioned in the previous round, the rating on technical feasibility (in Mean (M-TF) and Standard Deviation (SD-TF))
- the comments from round 2
- your rating, if provided

For each item, please provide the arguments for your rating. You are also allowed to adjust your rating. Explain why you think it is a good opportunity (or not) or limitation, and the role of SOA, using specific characteristics of SOA. We acknowledge the overlap of some hypotheses, and you are allowed to mention other hypotheses in your answer.

Questions Round 3

Definitions in context of this study
Opportunity: concept or pattern that involves the use of SOA and can improve the organization (design, development, maintenance, etc.) and use of BI.
Limitation: restriction of using SOA for the organization and use of BI.
Technically feasible: capable of being implemented with a reasonable amount of effort, considering currently available technologies.

Assume a mature SOA organization
For the mentioned concepts and patterns, you can assume that a mature SOA organization is in place, in which all data from operational systems is accessible through services.

	Hypothesis	Mean	Standard Deviation (SD)	Mean (Techn. Feas.)	SD (Techn. Feas.)	
	Bulk synchronization + SOA updates	M	SD	M-TF	SD-TF	Comments previous round
1	A good opportunity for applying SOA to the collection of data is by performing bulk synchronization between the operational systems and the BI systems, using conventional methods (which are able to handle the voluminous size of the data), and receiving regular updates via services. More up-to-date information is then available.	0,9	2,0	0,8	1,1	*It could be done, but why is this a good idea?*

General: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Technical feasibility: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Master Data Management — M SD M-TF SD-TF

2 A good opportunity for applying SOA to the collection of data is by providing access to Master Data Management (MDM) based on SOA. Every layer of the BI cycle can provide input to and/or use MDM information.

M	SD	M-TF	SD-TF
2,2	0,8	1,1	1,0

On technical feasibility: Many BI tools do not have proper MDM capabilities to support this.

General: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Technical feasibility: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Transformational services

3 A good opportunity for applying SOA to the collection of data is by transformational services that fulfil the role of a transformation step of the conventional BI architecture. These services can increase the use of standard data formats.

M	SD	M-TF	SD-TF
1,9	0,9	1,0	1,4

General: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Technical feasibility: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Large data size

4 An important limitation for applying SOA to the collection of data is the large data size involved, which is usually not suitable for transportation through services.

M	SD
1,2	2,2

Comments previous round

A conventional ETL program will serve that purpose much better.

High transaction volumes can be handled through calling a service that would only pick up one transaction, or adjust the services to just add another datastore to fill through the ESB. Then large datasets are handled on a transaction by transaction base but in a continuous way so that at the end of the day very large datasets are handled.

Most important limitation is system performance (CPU) and network performance. SOA seems useful for bulk synchronisation, but servers are not always up to the job.

Some (e.g. Amazon.com) are already using services for this purpose, so I don't see why this should be a limitation.

General: *Strongly disagree* ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 *Strongly agree* ○ *No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

	Distributing by services	M	SD	M-TF	SD-TF	Comments previous round
5	A good opportunity for applying SOA to the distribution of data is by exposing the produced BI information through services.	2,2	0,7	1,3	1,4	*Simple low level (operational) decision support may be supported in this way. High level decisions (at CEO level) are impossible to support in this way. They need too much human interaction between DW, data mining, and* [illegible] *An important aspect of SOA is the availability of both push and pull mechanisms for distributing information linked with Web 2.0 mechanisms such as syndication (RSS/Atom) and mashups.* *Services will be predominantly used to integrate BI-related information between applications. You should be careful when exposing services to end users as they are not used to deal with this, so a user interface, wrapping the service, needs to be in place to view the results.*

General:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*
Technical feasibility:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

	Closed loop BI	M	SD	M-TF	SD-TF
6	A good opportunity for applying SOA to the distribution of data is by exposing information for use in other applications (so called closed loop BI).	2,3	0,9	1,2	1,4

General:	*Strongly disagree*	3	2	1	0	+1	+2	+3	*Strongly agree*	*No opinion*
Technical feasibility:	*Strongly disagree*	3	2	1	0	+1	+2	+3	*Strongly agree*	*No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

	Mashups	
7	A good opportunity for distributing information by services is by creating mashups: combining information sources from (often internet based) 3rd party providers with internal information (e.g. from the data warehouse). An example is visualizing sales or stock levels using Google Maps.	*This hypothesis was not part of round 2 and therefore has no rating yet. However, it has been mentioned in round 1, as well as in round 2, by several participants, which is why we have added the subject.*

General:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*
Technical feasibility:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

	Complex queries as services	M	SD	M-TF	SD-TF
8	A good opportunity for applying SOA to the distribution of data is by wrapping complex queries and other BI functions, demanded by users, as services.	1,2	1,9	0,2	1,2

General:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*
Technical feasibility:	*Strongly disagree*	-3	-2	-1	0	+1	+2	+3	*Strongly agree*	*No opinion*

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Decrease latency of data — M SD

9 A good opportunity for applying SOA to the BI process is by information integration based on SOA, which can decrease the latency of data, by collecting, combining, analyzing and distributing information together with the information from the traditional data warehouse. — 1,2 2,0

General: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Componentization BI architecture — M SD M-TF SD-TF

10 A good opportunity for applying SOA to the BI process is that the use of services enables componentization of the BI architecture, to support loosely coupled BI systems. — 1,8 0,9 0,9 1,3

General: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Technical feasibility: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Business Activity Monitoring — M SD M TF SD TF

11 A good opportunity for applying SOA to BI on the operational level is in Business Activity Monitoring, to monitor business processes in (near) real time fashion. — 2,2 0,8 1,5 1,0

General: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Technical feasibility: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Pervasive/Ubiquitous BI — M SD

12 A good opportunity for applying SOA to BI is pervasive or ubiquitous BI, which means that BI is used everywhere and in every process throughout the organization. Just like SOA can add business value by decoupling the business process from the underlying applications, this can be applied to BI as well. — 2,2 0,8

General: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Agile organization — M SD

13 A good opportunity for applying SOA to BI is in the creation of a more agile organization, which can react faster to changes in the business. — 2,1 0,9

General: *Strongly disagree* -3 -2 -1 0 +1 +2 +3 *Strongly agree* — No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Integration other components with BI	M	SD
14 A good opportunity for applying SOA to BI is that SOA facilitates the integration of other components with BI, like data access tools.	2,4	0,7

General: Strongly disagree ○ -3 ○ -2 ○ -1 ○ 0 ○ +1 ○ +2 ○ +3 Strongly agree ○ No opinion

Can you provide arguments for your rating? Which characteristics of SOA are important here and why?

Personal Info

In the previous questionnaire, you have been asked to provide some details on your profession and experience. Please check this information here, or complete them in case we have not received this information from you yet.

Allows to be listed (with name, profession, and company) in thesis as participating expert

Name

Profession

Company / Institute

Number of years experience with BI (if applicable)

Number of years experience with SOA (if applicable)

Number of BI projects participated in (if applicable)

Number of SOA projects participated in (if applicable)

Sending back the document

First of all, thank you very much for filling out the document for this third and last round. You can save & send this document to my e-mail address: stefan.linders@capgemini.com. If you have any questions in the mean time, please do not hesitate to contact me.

Kind regards,

Stefan Linders

Contact Details

Stefan Linders
Graduate Student at University of Twente
Enschede, The Netherlands
Tel +31 (0) 6 2336 4709
E-mail stefan.linders@capgemini.com