Designing a big data software-as-a-service platform adapted for small and medium-sized enterprises

by

Lyubomir Nedyalkov

June, 2013

Delft, The Netherlands

MASTER THESIS

Designing a big data software-as-a-service platform adapted for small and medium-sized enterprises

by

Lyubomir Nedyalkov (4185110)
L.A.Nedyalkov@student.tudelft.nl
Faculty of Technology, Policy and Management, Delft University of Technology
Jaffalaan 5, 2628BX Delft, Netherlands

Graduation Committee:

Chair: Prof. Dr. Y. H. Tan

(Prof. Dr. W.A.G.A. Bouwman)

First supervisor: Dr. Ir. G.A. (Mark) de Reuver

Second Supervisor: Dr. M.E. (Martijn) Warnier

External Supervisor: Bastiaan Walenkamp

TU Delft Delft University of Technology

DIALOGUES TECHNOLOGY

June, 2013

Delft, The Netherlands

Abstract

Identified as one of the biggest technological trends for 2013, "big data" will be a major driver for organizations' technology spending for the years to come. Usually, the term is associated with three main characteristics: volume, variety, and velocity, also known as the 3 V's. In this thesis *big data technologies and architectures have the purpose to economically draw insights from large volumes of a wide variety data by enabling high-velocity capture, discover, and analysis.* Big data analytics add to traditional business analytics by dealing not only with structured data kept in relational databases but also with unstructured data from a variety of sources (e.g. social media, e-mails, sensors, etc.). Thus, in essence, big data is not a new phenomenon.

The basis of this thesis is an idea for making big data analytics more accessible to small and medium-sized enterprises (SMEs). Technological advancements in cloud computing, mainly virtualization and reduced cost of storage, remove some of the barriers that stand between SMEs and big data analytics. However, big data has inherent issues (e.g. hidden biases) that must be addressed appropriately if reliable insights are to be generated from it. As SMEs are assumed to lack the needed competencies for dealing with these issues, we argue that it will be beneficial to connect them to data scientists. Thus, in this thesis *we designed a big data software-as-a-service platform (that connects SMEs to data scientists) and the underlying business model from the platform provider's perspective*.

One design cycle has been executed during the project. We started with identifying whether SMEs perceive particular business needs for big data analytics. Companies were found to experience such needs from different calibers. Thus, it was decided to build the platform in a modular way so that it allows SME customers to flexibly outsource only these parts of a big data analytics process that they are willing to. Then, we set two goals that the big data SaaS platform must fulfill once realized. First, it needs to *provide SMEs with access to big data services within different application service layers in order to meet different calibers of business needs.* And, second, the platform must *facilitate the collaboration between SMEs and data scientists for the development of reliable big data analytics services.* Also, two ideas for pilot services were investigated. After setting the goals, three sets of requirements were defined: functional, user, and contextual requirements. The final activity, part of the information system (IS) design, was setting the platform's technical architecture that can achieve the specified goals and meet the set requirements. In addition, detailed specifications for the two pilot services were also described.

After all planned activities for the IS design were concluded, we moved to business modeling by applying the STOF method. The basic STOF method was extended with eight additional critical design issues (CDIs) identified from business model, platform and business ecosystem theories. We first outlined the business model behind the big data SaaS platform as part of the STOF method's quick scan stage. The outline was then evaluated on different criteria: completeness, consistency, viability, scalability and sustainability. Both internal evaluation within the project team and external evaluation with experts and practitioners were performed. The main conclusions were that at this point the business model is not yet complete and viable thus refinement with CDIs in the future is required.

Acknowledgements

In the five months between February and June 2013 I got the opportunity to work on a thesis project within a field that is of great interest to me. I would like to express my gratitude to the people mentioned below for helping me through this period in particular and my Master studies in general.

First, I want to express my gratitude to my TU Delft supervisor Mark de Reuver. Your guidance and valuable feedback provided direction in my research from its beginning until its very end. I sincerely thank you for your mentorship and for always being professional (even when faced with my strange printing requests).

I would also like to thank my external supervisors Bastiaan Walenkamp and Marc van Gent for giving me the opportunity to conduct my thesis project within Dialogues Technology. You provided me with valuable connections and at the same time gave me all the freedom I needed. I deeply appreciate that.

A would like to stress my gratitude to my aunt Katya and cousin Pavel for supporting me through this final period of my studies. Dear aunt, thank you for opening your home and for being so patient with me. Pavel, thank you for introducing me to the guys from Dialogues Technology and for all the laughs that we have had. You both made me feel like I was home.

I want to thank my family for providing me with all the support that I needed during my studies. If it was not for you I would have never studied at TU Delft in the first place. Also, I want to thank my girlfriend for her understanding, patience and support. And, finally, I thank all my old and new friends from Delft who made this period of my life so exciting and fun.

I thank you all!

Lyubomir Nedyalkov,

Amsterdam, June 17th, 2013

Table of Contents

List of Figures

List of Tables

1

Introduction

Nowadays, the term "big data" has become hype and many IT professionals and business people rank it, together with cloud computing, as the top information technology trends for the years to come. Usually, big data is associated not only with its volume but also with its variety and velocity (the 3 V's of big data). Many organizations started to regard data as a strategic asset. However, it is important to mention that data on its own is not information. First, data need to be translated to information, while information in turn needs to be translated into insights (or knowledge) in order to bring value to companies. Traditionally, this is the domain of business intelligence (BI).

Traditional BI solutions were built for on-premise use and consequently required relatively large investments from companies. First, the supporting IT infrastructure (e.g. data centers) had to be built which, in turn, brought high maintenance and staff costs. Second, organizations had to purchase database management systems and specialized analytics software provided by large vendors (e.g. Oracle, SAS). Usually, the vendors' pricing schemes were based on system usage per user, per annum hence license costs amounted thousands of euros. In general, these pricing schemes ask for careful early planning by management but even then they are quite inflexible. Not surprisingly, mainly large companies could afford traditional BI solutions due to the high initial cost and large total cost of ownership associated with implementing such on-premise information systems.

With the emergence of cloud computing, however, this situation has changed. According to the NIST definition of cloud computing [6], the cloud model is composed of five essential characteristics (resource pooling, rapid elasticity, on-demand self-service, measured service, broad network access) and three service models (Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS)). Storage, processing and software can now be moved from the organization's premises to the "cloud" by subscribing to services based on the three service models. In this way, organizations do not have to create and maintain IT infrastructure, thus the initial costs and total cost of ownership can be significantly reduced. Additionally, the pricing schemes of service providers are usually based on monthly subscriptions. Consequently, companies have more flexibility in suspending subscriptions when they need to cut costs. As a result, many cloud-enabled services emerged (including in the BI domain) that are not only accessible to large companies but also to SMEs[1]. Another advantage of cloud services that is especially important for small companies and start-ups is the ease of scaling of operations. Sometimes these enterprises can have difficulties in predicting customer demand for their products/services thus scalability at low cost could be very attractive to them.

[1] According to the European Commission, a company can be categorized as an SME if it has less than 250 employees or less than €50 million/€43 million in turnover/balance sheet total.

Cloud computing and the inherent service delivery models are removing some of the perceived barriers that stand between SMEs and big data analytics adoption as noted above. However, there are still other issues that need to be addressed in order to enable big data analytics adoption within SMEs. It can be assumed that some SMEs cannot transform data into useful insights without external help since their organizational competencies have not been developed sufficiently. Product/service development, increasing customer base and securing cash flows are usually the activities that SMEs are mostly focused on. Thus, a small/medium business would perceive business analytics as non-vital. However, analytics insights can help in improving the product/service offering of a company, as well as its customer and market selection. Consequently, finding a way to connect SMEs to data scientists[2] could be beneficial to both groups.

Until recently business intelligence has been focused mainly on structured data. This type of data is stored in relational database systems to which enterprises (usually) have direct access. Unstructured data, on the other hand, can be regarded as a combination of data from different formats like e-mails, images, audio/video files, social media content, etc. According to IDC, 90% of the generated digital content worldwide is in the form of unstructured data [7]. It can be assumed that when a business is still small, external (mainly unstructured) data analyses could provide more useful insights because structured data is not yet abundant enough. Therefore, if SMEs can access the right data (structured/unstructured and internal/external) and are able to analyze it, they can make more sound decisions and eventually become more profitable.

In today's service-oriented IT landscape, more and more data will become available to companies in the web space. The service orientation trend that we are witnessing is shifting from building infrastructure to the development of platforms and eco-systems [7]. Social networks, e-commerce and many other websites are providing third-party developers with means to connect to their systems and extract data (i.e. through web services). Thus, many types of external and unstructured data become accessible to organizations. Recognizing how these data can be leveraged, i.e. how they can provide valuable insights, and subsequently performing the right type of analysis might earn an enterprise a competitive position in their industry.

Before the beginning of this thesis, an idea for making big data analytics more accessible to SMEs was introduced to the researcher by the CTO (Chief Technology Officer) of Dialogues Technology (a Dutch software development company). After getting more acquainted with the research domain, the researcher further refined the idea by proposing a particular implementation of an information system. Thus, the idea of developing a big data SaaS platform that connects SMEs to data scientists came to be. Figure 1 presents an overview of the main stakeholder groups and value flows as envisioned at the start of the thesis.

The researcher also searched if there are any existing cloud-enabled platforms that connect companies to data scientists in the domain of big data analytics. The goal of this search was to assess the innovativeness and competitive landscape for the envisioned platform. The researcher perceives that academic literature does not contain such information. Thus, searches were performed in two

[2] The group of data scientists includes data analysts, BI consultants and other professionals who have experience in drawing information from raw data.

web publication portals that focus on technology news and analysis - TechCrunch and TechRepublic. The search terms that were used are “big data”, “analytics”, and “SME”.

Several analytics platforms similar to the envisioned platform were identified. The services are described shortly in Appendix A. However, none of the identified existing platforms create the business network described in Figure 1.

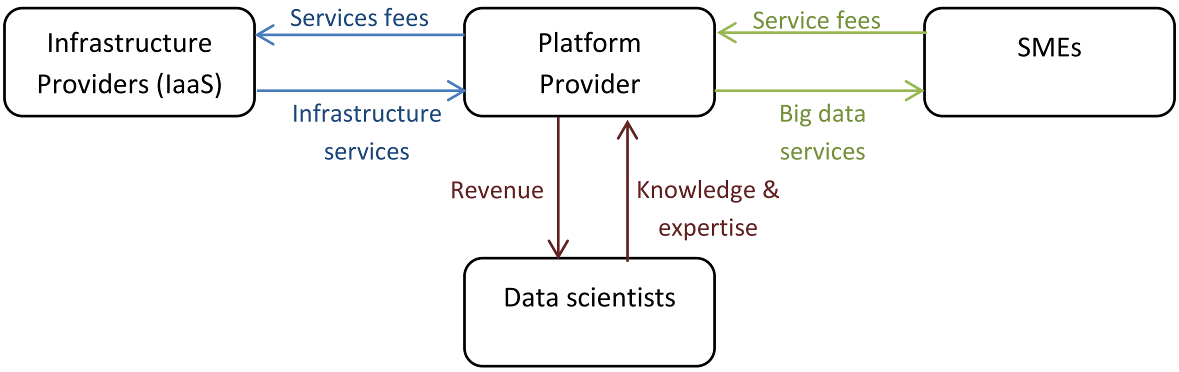

Figure 1: Platform's main stakeholder groups and value flows

Based on that, we formulate the problem statement in this research project:

> ***Currently, there is no existing SaaS platform that enables SMEs to draw insights from big data by collaborating with data scientists***

Finally, several assumptions underlying the described idea must be made explicit:

Assumption 1: SMEs lack the resources necessary for drawing insights from big data.
Resources include IT infrastructure, human, or financial resources.

Assumption 2: SMEs can recognize cloud computing as effective means to access IT infrastructure at low cost without the need for maintenance.
The assumption is that cloud computing is perceived by SMEs as means to access IT infrastructure for using specialized hardware and software (for big data analytics) which they cannot afford in an on-premise fashion.

Assumption 3: SMEs lack the competencies necessary for drawing insights from big data.
Competencies relate to the abilities of SMEs to recognize which data should be analyzed and what type of analysis can be done in order to meet a particular business need.

Assumption 4: Hiring big data analytics consultants is not feasible for SMEs due to high costs associated with consultancy services.
It is assumed that consultancy services can be perceived by SMEs as expensive due to high consultancy fees.

Assumption 5: Value can be created by connecting SMEs to data scientists in a business ecosystem around a SaaS platform.

It is assumed that a platform can bring value to both SMEs and data scientists by reducing the costs (for IT infrastructure, consultancy fees and transaction costs) for the former and being a source of revenue for the latter.

Now that the research problem and a potential solution were defined we can set the research goal of this thesis project.

1.1. Research Goal

This thesis aims at designing a big data SaaS platform which will bring together SMEs and data scientists in a common business ecosystem. In such an ecosystem, service offering aspects like value creation and value appropriation are central and must be addressed explicitly. Therefore, a business model must be designed together with the platform. Thus, we set the research goal of this thesis project as:

> ***To design a big data analytics software-as-a-service platform and the underlying business model from the platform provider's perspective***

In order to clarify more the research objective, the embedded concepts are described further below:

- ***Big data analytics*****.** Traditional business intelligence includes gathering, storing, analyzing and accessing structured data in an enterprise. In this project big data analytics can be defined in the same way with the difference that raw data used as input can come from different sources (e.g. information systems, web pages, social media, sensors) and in different types (e.g. text, image, audio, video) thus is unstructured in nature.

- ***Software-as-a-service platform*****.** A SaaS platform is a combination of hardware, software and networking technologies on top of which a number of services run. These services are distributed in a multi-tenant fashion to subscribers of the platform. In the SaaS model, subscribers do not manage or control the underlying infrastructure in any way [6].

- ***Big data Software-as-a-Service platform*****.** The platform will enable SMEs to draw insights from big data by collaborating with data scientists by using existing or developing new analytics services.

- ***Business model.*** According to Bouwman, Faber [8], a business model is a blueprint that describes how a network of organizations jointly creates and captures value from technological innovation.

Now that we have defined the goal of this research, we shall formulate the main research questions. The research approach for answering these questions is discussed after that.

Q1) What are business models, platforms, and business ecosystems and what critical design issues related to these concepts can be drawn from the existing knowledge base

This question has the purpose of identifying critical design issues (CDIs) that must be addressed during business models design. The list of CDIs provided in the STOF method is used as a generic list which is supplemented by artifact-specific CDIs identified in the literature. That is, platform and business ecosystem theories are reviewed for identifying platform-specific CDIs which supplement the ones from the STOF method.

Q2) Are big data analytics and the envisioned SaaS platform relevant for small and medium-sized enterprises?

This question aims at understanding if SMEs perceive any particular business needs for adopting big data analytics and the envisioned IS artifact. In this way relevance of the research project and the proposed solution is sought.

Q3) What are the goals, requirements and structural specifications for the big data SaaS platform?

This question embodies the activities to be performed in the first three stages of design cycle described by Verschuren and Hartog [9]. The design activities in the next stages (prototyping and evaluation) fall out of the scope of this research due to time limitations.

The third research question is further broken down to the following sub-questions:

Q3.1) What are the goals of the platform?

This question relates to the first stage of the design cycle, i.e. the specification of goals that the platform aims at achieving when implemented.

Q3.2) What are the functional, user and contextual requirements for meeting the goals?

This question relates to the second stage of the design cycle where the requirements for achieving the goals are specified. Functional, user and contextual requirements are drawn before proceeding to the next stage. Within the latter two groups non-functional requirements are also specified.

Q3.3) What are the structural specifications of the platform?

This question relates to the third stage of the design cycle where the structural specifications are derived from the requirements.

Q4) How does the business model of the designed big data SaaS platform look like?

An outline of the business model from the platform provider's perspective is created.

Q5) To what degree the designed business model can be considered as complete, consistent, viable, scalable, and sustainable?

Finally, the completeness, consistency, viability, scalability and sustainability of the business model are assessed.

1.2. Research Approach

The research approach adopted in this study is the design science approach. First, the design science research framework that is applied in this research is presented. And, second, the research methodology and the corresponding data collection methods are described.

1.2.1. Research Framework

The framework selected for conducting the research is the one proposed by Hevner, March [5]. The framework is widely used and referenced by many scientific authors[3]. We perceive this as a positive indication for its high quality thus we have decided to adopt it in this thesis also. Another reason is that the framework provides of a set of guidelines for "conducting and evaluating good design-science research" [5] (see *Appendix C*).

Figure 2 shows a graphical representation of the research framework adapted for the purposes of this project.

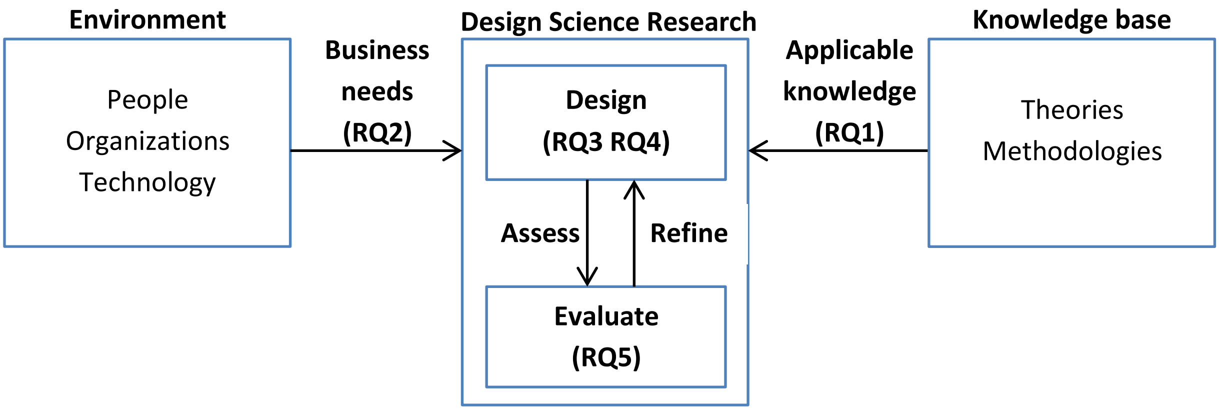

Figure 2: Design science research framework and positioning of research questions, adapted from [5]

On the right side is the existing knowledge base. When designing IS artifacts, an understanding of the nature of an artifact should be achieved. For that purpose, the first research question addresses the definitions of business models, platforms, and business ecosystems by looking at the existing scientific literature. Concepts closely related to these theories and applicable for the design of a SaaS platform are also addressed. Furthermore, critical design issues for designing feasible and viable business models are identified from these theories.

On the left side is the environment which is composed of people, organizations, and technology. People within organizations perceive different business needs which are defined in terms of goals, problems, and/or opportunities [5]. The second research question of this study aims at identifying whether SMEs perceive particular business needs for big data analytics and the envisioned SaaS platform. An artifact that does not address any problems or opportunities will most probably fail to succeed. Also, input from the environment is crucial for defining the goals of the artifact which is done during the first step of the artifact design.

In the middle is the domain of the IS research where two core processes run in turn. The first one is design. In this thesis both an IS artifact and a business model are designed. The IS design consists of three stages that make up an adaptation of the Verschuren and Hartog [9] design cycle. The third research question directly addresses these stages by asking for the goals (1), requirements (2) and structural specifications (3) of the envisioned platform. The fourth research question addresses the business model design. The second process is evaluation. The fifth and final research question addresses the evaluation of the designed business model only.

[3] There were 1,768 references of the framework of Hevner et al. within Scopus.com by June, 2013

1.2.2. Research Methodology

The research project is conducted in the following phases which address each research question in turn:

1. ***Theoretical Background (RQ1).*** First we review business model, platform and business ecosystem theories with the purpose of identifying critical design issues (CDIs) for creating viable business models.
   Methodology: Literature review

2. ***Business Needs Identification (RQ2).*** In this phase we interview representatives from several SMEs. The results of the interviews are presented and examined in order to understand the particular business needs of SMEs for adopting big data analytics. Additionally, all assumptions, underlying the research problem, are tested in order to see whether the proposed implementation of the big data SaaS service can indeed realize the identified business needs.
   Methodology: Interviews

3. ***Information System Design (RQ3).*** In this phase, the big data SaaS analytics platform is designed by taking the following subsequent steps.
   3.1 ***Goals definition.*** The goals of the designed IS artifact are defined by taking into account the identified business needs from the previous research stage.
   Methodology: Desk research

   3.2 ***Requirements specification.*** Functional, user and contextual requirements are defined as part of the requirements specification step.
   Methodology: Desk research (functional and contextual requirements) and Interviews (user requirements)

   3.3 ***Structural specifications.*** In this step the technical architecture of the envisioned information system that fulfills the identified requirements is defined.
   Methodology: Desk research

4. ***Business Model Outline (RQ4).*** In this phase, we outline the business model underlying the designed big data SaaS platform. From the different business model methodologies compared in Appendix B, we have chosen the STOF method. The reason is that this method gives the most detailed breakdown of business model components for ICT innovations. Additionally, only the STOF method provides explicitly an extensive list of CDIs related to creating both customer and network value.
   Methodology: Application of the STOF method – Step 1: Quick scan.

5. ***Business Model Evaluation (RQ5).*** In this final phase, the business model is evaluated by assessing its completeness, consistency, viability, scalability, and sustainability. For the purposes of viability assessment, the list of critical success factor (CSF) questions within the STOF method is modified on the basis of the identified CDIs in the first project phase. The modified list of CSF questions is applied in this phase.
   Methodology: Project team discussion session and Interviews with experts and practitioners

1.3. Thesis Structure

The thesis structure is aligned with the main research phases and is presented in Figure 3 below.

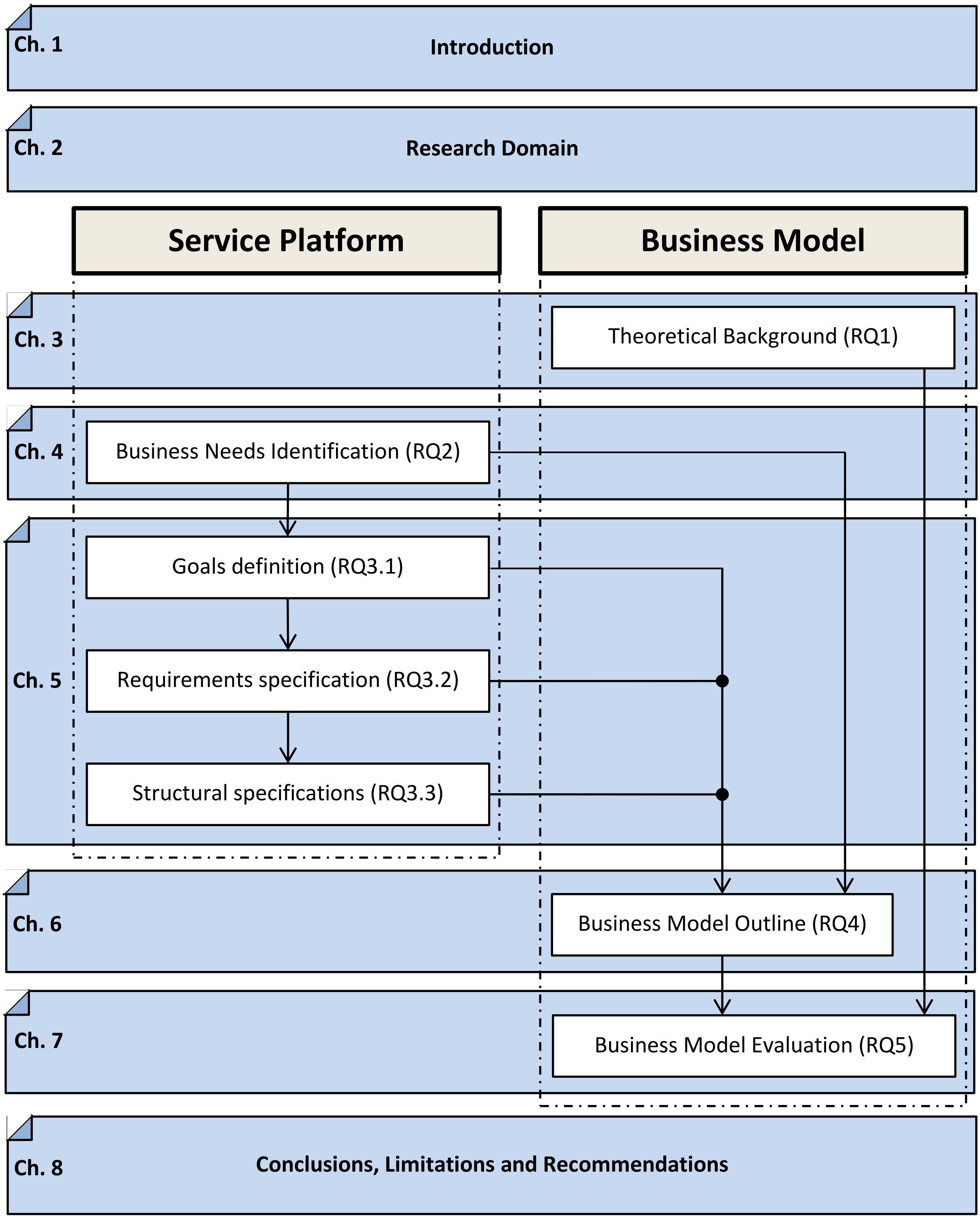

Figure 3: Thesis Outline

Before addressing the first research question, we need to position the envisioned artifact within the research domain. This is done in the following chapter.

2

Research Domain

The purpose of this chapter is to describe the research domain and then to position the envisioned platform within that domain. While doing so, the researcher has been able to obtain relevant domain knowledge which was very useful during the actual design process.

This chapter is structured as follows. In the first section cloud computing is defined, its building blocks are described in detail, and its relevance for SMEs is discussed. The second section describes business intelligence, big data analytics and their relation to each other. Finally, in the chapter conclusion, the designed artifact is positioned within the described research domain.

2.1. Cloud Computing

When looking at the scientific literature, various definitions of cloud computing can be found [6, 10-14]. In a Master thesis project Dihal [15] compared several definitions which were built by combining common characteristics from other cloud computing definitions. His results are presented in Table 1 below.

	[10]	[11]	[12]	[13]	[6]
Virtualization of resources	Yes	-	Yes	-	Yes
Variety of resources	Yes	-	-	-	Yes
Shared resource pool	-	-	-	-	Yes
Scalability	Yes	Yes	Yes	-	Yes
Convenience in use	Yes	Yes	-	-	Yes
Network enabled	-	Yes	-	Yes	Yes
Service level agreements	Yes	Yes	Yes	-	Yes
Pay-per-use	Yes	-	-	Yes	Yes

Table 1: Common elements and their presence in several cloud computing definitions [15]

The definition of Mell and Grance [6], or also known as the NIST definition of cloud computing, is adopted in this research project:

> *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (pp. 2)*

The NIST definition of cloud computing has been chosen for this project because of two reasons. First, it has been found to be most comprehensive by Dihal [15] since it covers all common elements in the cloud computing definitions which he had reviewed. Second, it provides a further breakdown of the cloud computing concept in three components: characteristics, service models, and deployment models. This breakdown helps for gaining deeper insights on both technology and service aspects of cloud computing. Next, these three components of cloud computing are described in more detail.

2.1.1. Essential Characteristics

There are five essential characteristics of cloud computing [6]. They describe the way the cloud resources are provisioned to the consumers of the cloud service. The characteristics are the following:

1. *On-demand self-service.* Consumers can get access to computing resources (e.g. storage, processing, memory, etc.) on their own without any human interaction with the cloud service provider [6]. For instance, consumers can conveniently order data storage over the Internet by only submitting a request form through a web browser.

2. *Broad network access.* Cloud resources are available over the network (or the Internet) and can be accessed through standard mechanisms from any client device [6]. This allows consumers to access cloud resources and capabilities from various geographical locations where there is network connectivity. Moreover, they can use different devices (e.g. smart phones, tablets, PCs, etc.) to access the cloud services without having to worry about data/software availability across these devices.

3. *Resource pooling.* This characteristic refers to the capability of cloud service providers to pool computing resources for serving customers in a multi-tenant fashion. Physical and virtual resources are dynamically allocated to consumers according to their current demand. Consequently, consumers might have no control or knowledge over the location of the cloud resources they are consuming [6].

4. *Rapid elasticity.* Cloud resources can be scaled appropriately, i.e. they can be flexibly allocated or released, depending on current consumer demand. This could be done automatically at any time and in any quantity, thus providing the customer with a sense of resource absoluteness [6]. This characteristic could be very advantageous to consumers who are having difficulty with projecting their demand (e.g. start-ups).

5. *Measured service.* This characteristic refers to the monitoring, control and reporting of resource usage to both the cloud service provider and consumer. The metering capability of the cloud allows service providers to adopt a pay-per-use revenue model where the service consumers pay for the portion of cloud resources that they have actually used [6].

2.1.2. Service Models

According to the NIST definition, cloud computing is made up of three service models: Infrastructure-as-a-Service, Platform-as-a-Service, and Software-as-a-Service (see Figure 4).

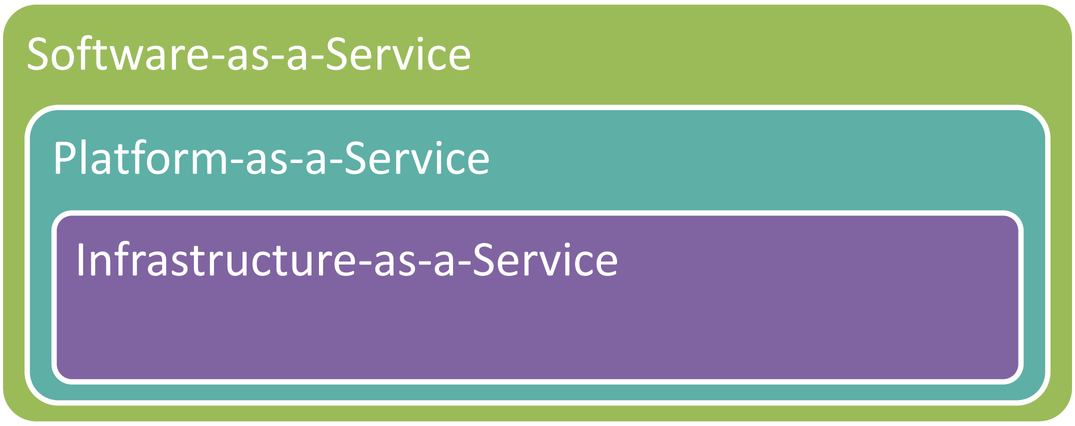

Figure 4: Cloud computing service models [6]

- *Infrastructure-as-a-Service (IaaS)*

The IaaS service model includes the provision of fundamental computing resources like processing, storage, and networks to consumers [6]. In the past, remote resources were leveraged for large scale computing work for the first time with grid computing. However, control of these resources stayed on site and remote users did not have the freedom to use different software applications or operating systems than the ones supported by the provider [16]. The IaaS model changed this situation with the help of virtualization. The IaaS provider turns over the control of part of its resources to the consumer who in turn can deploy and run any software (e.g. operating systems or applications) on the cloud. Consumers, however, cannot manage or control the cloud's underlying hardware and networking infrastructure [6].

Sotomayor, Montero, Llorente, & Foster [17] argue that an IaaS cloud allows for the provisioning of computational resources to consumers in the form of Virtual Machines (VMs) that are physically deployed in a IaaS provider's data center (i.e. resource pooling). Virtualization allows the leased resource to be isolated from the rest of the infrastructure in a secure way [16] thus protecting other service consumers. Amazon is an example of one such provider with their Elastic Compute Cloud (EC2) service. EC2 consumers can allocate or release VMs conveniently through a web interface or a programmers' Application Programming Interface (API) (i.e. on-demand self-service).

- *Platform-as-a-Service (PaaS)*

According to NIST, the PaaS model provides consumers with the capability to deploy consumer-created or acquired applications on the cloud. These applications can be created with the help of programming languages, libraries, services and tools supported by the PaaS provider. The consumers have control over the deployed applications and possibly for the configuration of the application environment. However, they have no control over the underlying infrastructure including network, servers, operating systems, and storage [6].

The PaaS model allows consumers (i.e. software developers) to focus mainly on the development and monetization of their software applications without having to invest in and maintain the underlying infrastructure [18]. Instantiations of PaaS can support the entire software life-cycle by allowing developers to design, develop, test, deploy and host their applications in a single environment [19].

Usually, PaaS instantiations are typical representatives of two-sided platforms which have a mediating function between the demand and supply sides of the market. Platform theory on two-sided markets is reviewed in the next section thus discussion on the mediating role of PaaS instantiations is limited at this point.

- *Software-as-a-Service (SaaS).*

In the SaaS model, consumers use the service providers' applications which are running on the cloud infrastructure. The applications can be accessed from any user device through a web page or a native smartphone application, for instance [6]. Thus, high flexibility is achieved in terms of time and location of access [18]. The consumer, however, is not allowed to manage or control the underlying cloud infrastructure with the possible exception of limited user-specific application configuration settings [6].

In the SaaS model software was moved from the organizations' premises to the cloud. This shift has changed the way software applications are being sold. In the traditional on-premise model, software licenses were bought and used on hardware owned by the licensee for an unlimited time or on per annum basis. In the SaaS model, the consumer pays a fee for accessing software that is ran on a third-party server. As soon as the subscription is terminated, the consumer loses access to the application. Usually, SaaS revenue models are either pre-paid subscription or on pay-as-you-go basis. The first model entails a combination of allocated resources (e.g. storage, data transfer, number of users, etc.) for a certain fee per month [18]. Consequently, companies have a higher degree of flexibility in comparison with the on-premise model when choosing the exact quantity of resources they will require. On the other hand, the pay-as-you-go model provides the highest degree of flexibility since consumers are charged only for what they have actually used [18].

2.1.3. Deployment Models

There are four cloud computing deployment models according to the NIST definition (see Figure 5). They specify the number and type of subscribers that can have access to the cloud services.

1. *Private cloud.* The cloud services are only accessible to the business units within a single organization. The infrastructure can be owned, managed and operated by the consumer organization, a third party or a combination of the two [6].
2. *Community cloud.* The cloud services are exclusively provisioned to a community of consumers which share similar concerns (e.g. mission, security requirements, etc.) [6]
3. *Public cloud.* The infrastructure is open and can be used by the general public. It may be owned, managed and operated by a business, government or academic organization. [6]

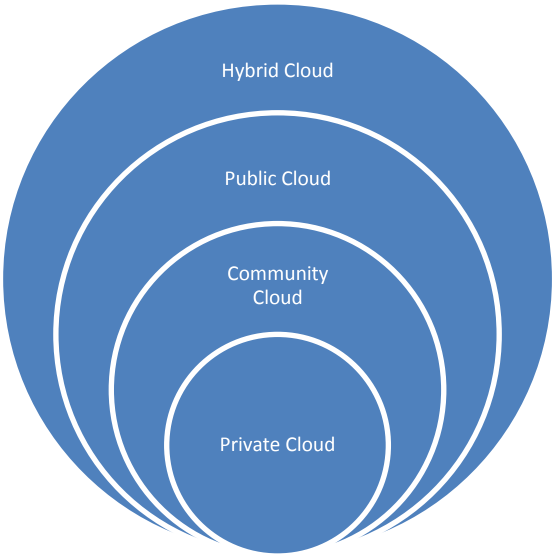

Figure 5: Cloud computing deployment models [15]

4. *Hybrid cloud.* The cloud infrastructure is a combination of two or more distinct infrastructures realized through different deployment models (private, community, or public). These infrastructures remain separate entities, but they are connected by means of standardized or proprietary technology which allows for data and application portability [6].

2.1.4. Relevance of Cloud Computing for SMEs

Cloud computing and its inherent characteristics, service models and deployment models can be regarded as economically appealing to small and medium-sized enterprises for two main reasons. First, SMEs can significantly reduce the total cost of ownership of IT when choosing cloud services since they no longer have to create and maintain the inherent IT infrastructure for internal service delivery. They can gain immediate access to computing resources without any upfront capital expenditure [14] which also reduces their time to market [13]. Second, the fees and quality of service brought by cloud service providers are far better than what most SMEs can realize on their own due to the economies of scale enabled by virtualization [14]. For instance, the pay-as-you-go revenue model allows SMEs to pay for the resources they have actually used [18]. In contrast, the traditional on-premise license payment schemes require customers to pay the total license fee even though they might not use the software application for weeks or even months.

Another advantage of cloud services that is particularly important for SMEs is the ease with which they can scale their operations depending on client demand. Again, no large capital investments are required, while the scaling up or down is immediate and seamless [14]. High-technology start-ups are representatives of the SME group for which rapid elasticity seems very advantageous due to limited financial capital and difficulty when projecting customer demand.

When compared to large organizations, SMEs lack an extensive legacy IT infrastructure that would impede the migration to the cloud. In other words, it will be much easier for SMEs to start using cloud services than for large organizations [14].

Finally, we should also mention that there are several downsides of cloud computing when compared to on-premise solutions. First, a constant Internet connection is required in order to access cloud services. Thus, in situations when network connection problems occur, end-users will not be able to access the services running on the cloud. For real-time analytics services this could be a major issue. In addition, keeping sensitive data secure might be perceived as problematic by companies since they have less control over cloud platforms. However, we argue that the advantages of cloud computing outweigh its disadvantages for SMEs (which in not necessarily the case for large companies) since it enables the adoption of different services (e.g. big data analytics) in the first place.

2.2. Business Intelligence and Big Data

This section presents the domain of business intelligence and big data. First, business intelligence is described. A definition of BI is provided for the purposes of this thesis. After that the BI systems' IT infrastructure and BI in the cloud are discussed shortly. Additionally, BI trends from the fourth quarter of 2012 Gartner BI survey are presented. Second, big data and big data analytics are described. At the same time, business intelligence and big data analytics are compared and contrasted.

2.2.1. Business Intelligence

Just like cloud computing, business intelligence (BI) has been defined many times by numerous scientific authors and industry specialists. Azvine, Cui [20] loosely define BI as "how to capture, access, understand, analyse and turn one of the most valuable assets of an enterprise — raw data — into actionable information in order to improve business performance". According to Jourdan, Rainer [21], BI is both a process and a product. The process includes methods that enterprises leverage for developing intelligence that is used for thriving in the marketplace. The product, on the other hand, is the information that helps organizations in predicting the behavior of different players in their environment, for instance competitors, customers, partners, etc. [21]. Negash and Gray [22] describe business intelligence in terms of "BI systems [that] combine data gathering, data storage, and knowledge management with analytical tools to present complex and competitive information to planners and decision makers". According to Elbashir, Collier [23], "BI systems are defined as specialized tools for data analysis, query, and reporting, (such as OLAP and dashboards) that support organizational decision-making that potentially enhances the performance of a range of business processes". BI is also viewed as "an umbrella term that is commonly used to describe the technologies, applications, and processes for gathering, storing, accessing, and analyzing data to help users make better decisions" [24]. Even though many definitions exist, it is widely accepted that:

> *Business intelligence has the purpose of translating data into meaningful insights (i.e. knowledge) for improving decision-making within organizations.*

The definition of business intelligence above is kept broad for a particular reason. That is, the artifact that this research study aims at designing will enable the creation of business analytics solutions that provide insights for different organizational domains. These range from strategic planning to tactical and operational process improvement, supply chain production and customer services [23] In order to support decision-making in these domains, business analytics systems include features like trend analysis, predictive modeling, customer behavior analysis, and reporting and visualization [20]. These features and the translation of raw data into meaningful insights are supported by a specialized IT infrastructure.

IT Infrastructure

Specialized IT infrastructure is necessary for the deployment and effective use of BI systems [23]. According to Azvine, Cui [20], BI systems require three main technology categories: data warehouses, analytical tools, and reporting tools. Data warehouses are used for gathering raw data from different sources and integrating these data for subsequent analysis. Analytical tools have the purpose to analyze the already structured data in the data warehouse and derive insights from it. Finally, visualization and reporting tools present the data to the decision-maker (who in most cases is not an analyst) in a user-readable manner [20].

Business Intelligence in the Cloud

Cloud computing allows organizations to outsource any non-strategic activities that do not lead to better competitive position in the market. The cloud service providers' core competencies and economies of scale are transferred to their subscribers usually resulting in a cheaper, better and more timely service delivery [25]. Naturally, this is also valid for business intelligence activities.

According to Gartner [26], adoption of cloud/SaaS business intelligence services is driven by three main factors. First, SaaS BI could lead to shorter time to value in terms of faster deployment and insights generation. This is particularly valid for SMEs where IT is constrained by existing work or limited financial budgets. Second, the cost dynamic is different between the SaaS and on-premises models. Services can be expensed on the balance sheet, while IT infrastructure needs to be capitalized. Other benefits of the SaaS model indicated by Gartner are the availability of cash flow and reduced IT maintenance costs. Third, SaaS applications can overcome the lack of the business analytics expertise needed by companies to build their own analytic applications by having prebuilt intellectual property. Moreover, these applications can provide easy and natural language interfaces that reduce the competence gap for decision makers [26]. Furthermore, data analysis can consume big quantities of computing resources and can take long time to complete. With cloud computing, the processing time can be reduced significantly by scaling up the used computing resources conveniently in a matter of seconds.

Business Intelligence Trends

In the end of 2012, Garner distributed a survey on the digital technology trends in the coming years to more than 2,000 CIOs from 36 industries and 41 countries [27]. According to the results, "Analytics and business intelligence" is seen as the highest priority digital technology for 2013. Additionally, "big data/analytics" is seen by 55% of the CIOs as the second most disruptive technology in the coming decade. The first place was given to mobile technologies while social media and the public cloud were ranked third and fourth respectively. Another quite important insight from the survey is that CIOs see the real disruptive power of these technologies in combination, rather than in isolation. Consequently, the idea of a developing a cloud-based BI service platform seems relevant for businesses on the surface of it.

2.2.2. Big Data

Essentially, big data is not a new phenomenon. However, technological advancements in virtualization, reduced cost of storage, growth in installed sensors, and innovative software and analysis tools have contributed to the term becoming hype [28]. Usually, big data is associated with three main characteristics: volume, variety, and velocity, also known as the 3 V's [29]. The definition of big data that in this thesis is adapted from [28] and is based on the 3 V's:

Big data technologies and architectures have the purpose to economically draw insights from large volumes of a wide variety data by enabling high-velocity capture, discover, and analysis

Next, big data's volume, variety and velocity of are discussed more thoroughly.

Volume

According to IDC, the total of all data created and replicated in the world at the end of 2013 will reach 4ZB. This number represents a nearly 50% increase in 2012 volumes and a quadruple increase in 2010 numbers [30]. These figures illustrate the rapid growth of data that we are generating. One of the main reasons for this growth in volume is the increasing number of individuals using the Internet from both wired and mobile devices. Figure 6 below shows more than a three-fold increase in the percentage of people using the Internet from 2001 to 2011. The number of mobile broadband subscriptions increased with similar rate in only five years – from 2007 to 2011.

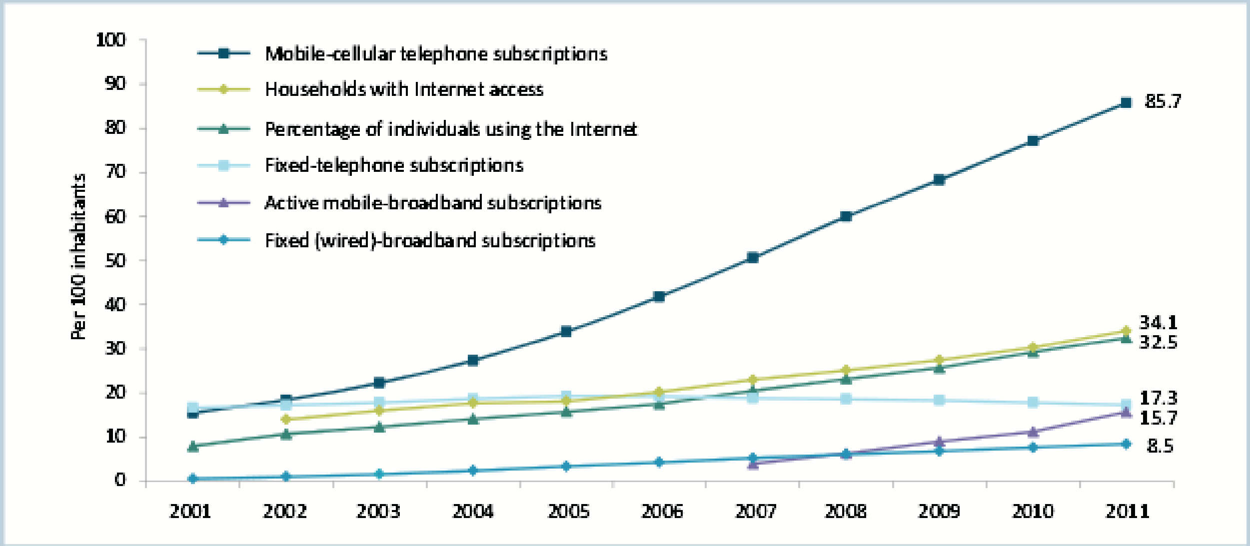

Figure 6: Global ICT Developments 2001-2011 [31]

With the increasing number of Internet users, the content generated by them also increased. This notion is related to the concept of Web 2.0 introduced by O'Reilly in 2005 for the first time. In Web 2.0 Internet users are stimulated to collaborate for creating and maintaining web content thus building collective intelligence [32]. Blogs, wikis and RSS feeds are all examples of how people are contributing to the massive growth of data on the web. Additionally, Web 2.0 supports the establishment of social networks of people with similar interest [32]. Example services are YouTube, Twitter and Facebook.

In recent years we have also witnessed a rapid growth in the number of APIs that are available to software developers (see Figure 7). Data generated within social networks (e.g. Twitter, Facebook), news portals (e.g. The Guardian), e-retailers (e.g. Amazon), etc. can be accessed either openly or with paid subscription. As more data from different sources becomes available through web services and APIs, businesses get more opportunities for different data analyses that can support decision-making.

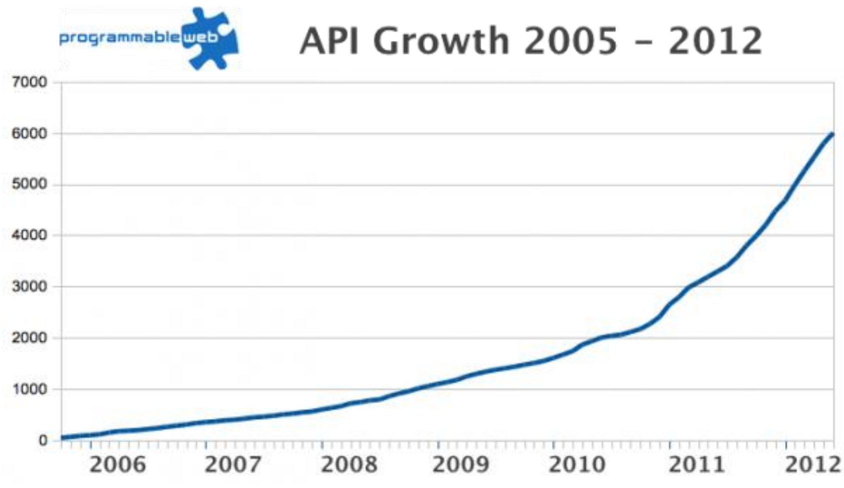

Figure 7: API Growth 2005-2012 [33]

Human activity however is not the only source of data. Different types of physical objects are being turned into 'smart' objects by equipping them with Radio Frequency Identification (RFID) tags and different sensors. In this way the objects can communicate with each other, but they can also send raw data to centralized locations [34]. Technologies that are used for making objects 'smart' are referred to as "Internet of Things" (IoT) technologies. For instance, utility companies use smart meters to collect and transmit consumption data remotely.

Variety

Data can be divided in two groups – structured and unstructured. Data that is being stored in a particular format in relational databases is referred to as structured data. According to Gartner [35], traditional data analyses within the business intelligence domain have been focused mainly on structured data. However, new and diverse types of data (text, video, audio) had become available to organizations. All these distinct types of data are broadly referred to as unstructured data and according to IBM they form around 80% of all of world's data. Sources of unstructured data include different files (e.g. word processing documents, PDFs), emails, social media, web content, video feeds, etc. [36]. That is, unstructured data lacks a common format and therefore cannot be stored in a relation database.

Velocity

Next to volume and variety, big data is also characterized by high velocity. More precisely, velocity refers to the speed with which data is created as well as to how fast the business needs it processed [37]. In some cases, data needs to be analyzed in real time. For example, Twitter introduced a set of data stream APIs which push tweets to a subscriber's endpoint at the moment of their creation. If tweets related to changes in the financial markets are analyzed in real-time (e.g. through sentiment analysis) then investment opportunities can be recognized and acted upon in a short time frame, possibly resulting in high financial returns. So, in some cases the longer it takes to retrieve insights from big data, the later a decision can be taken which eventually reduces the value of the insights and data analysis.

White [38] distinguished two techniques for analyzing big data: *the store and analyze* approach and the *analyze and store* approach. With the former approach, data is integrated in a traditional data warehouse (e.g. in a relational database management system) prior to running any analyses on it. As advantages of this approach, White [38] lists improved data integration and data quality management, as well as the ability to maintain historical information. The latter approach analyzes data on the fly as it flows across networks, business processes and between systems. The results from data analyses can then be pushed into a data stores or can be directly displayed in dashboards. Of course, if historical data needs to be accumulated, storing the analyses results is required. An important advantage of the '*analyze and store'* approach is that big data can be filtered and cleaned prior to pushing it into the data warehouse. Naturally, one of the main factors for selecting one of the two approaches is the length of the time span between the availability of the data and its analysis. If data gets outdated fast and/or the business needs to act upon it quickly after their generation then the *'analyze and store'* approach is more suitable.

2.2.3. Big Data Analytics

Traditional business intelligence has been more focused on the analysis of structured data which is usually being generated by organizations internally [35]. However, as already described above, external sources of unstructured data are becoming widely available to enterprises. Big data analytics can be regarded as an extension to traditional business intelligence in a sense that it includes analyses of unstructured data from various external sources. Rita Sallam, research vice president at Garner stated that "correlating, analyzing, presenting and embedding insights from structured and unstructured information together enables organizations to better personalize the customer experience and exploit new opportunities for growth, efficiencies, differentiation, innovation and even new business models" [35]. Additionally, White [38] indicates that big data analytics has many uses among which are advertisement placement, fraud detection, call center optimization, social media and sentiment analysis, and smart power grids.

Big data analytics however can be problematic: correlations do not necessarily indicate causality [39]. This is not always clear as people can be tempted to infer a cause-and-effect relationship between two events that are observed to happen together. Ultimately, the problems that we aim to solve and the questions that we ask in order to solve them determine whether big data analysis should focus on finding correlations *or* causalities [40].

Another major issue associated with big data is that it might not reflect the truth in an objective way. Datasets are the results of human design and as such they can be biased. Biases can be introduced during different stages of a data analytics process, for instance data collection or data analysis [39]. If explicit attention is not paid on such hidden biases then considerable risks can be introduced for decision makers. That is, basing decisions on biased (or even faulty) data that does not reflect the truth objectively is potentially devastating for businesses.

2.3. Conclusion

This chapter discussed the concepts of cloud computing, business intelligence and big data analytics as they form the research domain of the thesis project.

Cloud computing and big data analytics are both considered as top priorities for the years to come by the IT industry. For SMEs in particular, the essential characteristics and service models of cloud computing can enable the adoption of different technologies in an economically viable way. They can now take advantage of business intelligence and big data analytics solutions: something that was hardly possible for SMEs before the emergence of cloud computing. Moreover, with the increased availability of external data sources, SMEs can now leverage external data and analyze it for solving particular problems or realizing new opportunities.

The platform that this research project aims at designing is positioned within the cloud computing domain as an instance of the SaaS delivery model. The five essential characteristics of cloud computing and the SaaS delivery model can enable the high-velocity capture, discovery, and analyses of large volumes of data from different data sources (i.e. big data analytics). Thus, the platform is positioned within the big data analytics domain as well.

Furthermore, we argued that the advantages of cloud computing outweigh its disadvantages since the technology enables big data analytics adoption for SMEs in the first place. We also discussed two major issues associated with big data: problems with inferring causalities and hidden biases. Eventually, the envisioned platform has to provide reliable insights to its SME end users thus these issues must be addressed accordingly.

After the designed artifact has been positioned within the research domain, we now turn to the existing scientific literature on business models, platforms, and business ecosystems. In doing so, critical design issues that should be explicitly addressed during business model design are identified. The next chapter presents the results of the literature review.

3
Theoretical Background

This chapter's position within the research framework can be seen in Figure 8. Its purpose is to answer the first research question, that is:

> ***What are business models, platforms, and business ecosystems and what critical design issues related to these concepts can be drawn from the existing knowledge base?***

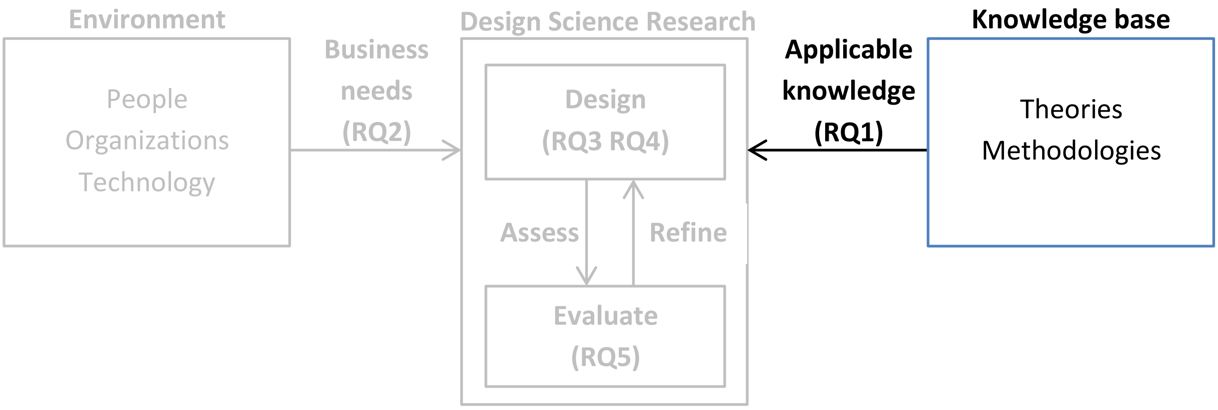

Figure 8: Positioning of the "Theoretical Background" chapter within the research framework

The first research question is answered by conducting a review of the existing scientific literature. The concepts of business model, platform, and business ecosystem are defined in the context of this project. In addition, other concepts that are closely related to platforms and business ecosystems are also reviewed and described. The objective of the literature review is to identify different business model design issues related to the nature of the envisioned IT artifact prior to the start of the design activities. Such design issues are critical for the viability of the business model [41] and from now on will be referred to as critical design issues or CDIs.

The first section of this chapter presents the research methodology selected for answering the first research question: literature review. The types of publications, selected databases and used search terms are presented.

The concept of business model is defined in the second section. Ultimately, the aim of this thesis is to design a platform that facilitates successfully the creation, transfer and capture of value between the ecosystem's members. For this reason, a business model should be carefully drawn. The STOF method has been chosen for business model design in this project as it includes an explicit list of critical success factors (CSFs) and CDIs that are used for evaluation and business model refinement purposes. In addition, the VISOR method is reviewed for relevant CDIs which can supplement the ones within the STOF method. At the end of the section, a list of generic CDIs relevant for innovations in the ICT industry is devised.

Furthermore, a clear understanding of the platform concept is required since this thesis aims at designing one such artifact. The third section of the chapter has the purpose of defining what a platform is in the context of this project. In addition, different concepts related to platforms (e.g. network effects, multi-sided markets, pricing issues, etc.) are identified in the existing literature and are described as well. CDIs related to platforms are explicitly mentioned in the text of the section.

It is common for platforms (especially digital ones) to bring together different groups of stakeholders in business ecosystems. This is also the case in this project where SMEs and data scientists are identified as key stakeholders. For this reason, the fourth section of the chapter defines the business ecosystem concept. Literature relevant for creating and governing such ecosystems is discussed as well, as it gives important input on CDIs for the business modeling activity.

Finally, a conclusion summarizes the chapter and provides the answer to the first research question, containing the definitions of business models, platforms, and business ecosystems. More importantly, the list of generic CDIs in the STOF method is extended in order to take into account artifact-specific CDIs identified in platform and business ecosystem theories.

3.1. Research Methodology

The research methodology selected for answering the first research question is literature review. The review is limited to primary publications including books, scientific journal articles, conference proceedings, and Master thesis reports from several databases: Scopus, Web of Science, Google Scholar, and TU Delft Discover. The following search terms per concept are used:

- Business Models: "business models AND definition", "business models AND literature analysis
- Platforms: "platform thinking", "digital platforms", "multi-sided platforms", "platform openness", "platform AND pricing"
- Business Ecosystems: "business ecosystems", "governance AND business ecosystems"

Next, business models are defined and generic CDIs for innovations in the ICT industry are identified from the STOF and VISOR methods.

3.2. Business Models

There is a general lack of consensus among scientific authors regarding the definition of business models. The reason can be attributed to interest in business models from a variety of perspectives all of which relate to the term [42]. Examples of such perspectives were given by Zott, Amit [43]: e-business and use of IT within organizations [44-48]; strategic issues such as value creation and competitive advantage [49-51]; and innovation and technology management [51, 52]. Furthermore, these three perspectives are not mutually exclusive.

The business model definition that this thesis project adopts must be in line with the research objective. In that sense, several aspects are deemed important. First, the definition should revolve around service delivery as the envisioned platform is not a product that can be bought and owned by its customers. Second, information technology should be part of the definition as it broadly represents the main enabler for the artifact. Last but not least, the selected business model definition needs to address value creation for all business ecosystem members, i.e. not only for the platform provider. Bouwman, Faber [8] provide a definition that conforms to the requirements described above. According to them, a business model is:

> *"A blueprint for how a network of organizations co-operates in creating and capturing value from technological innovation".*

As this definition is broad, the authors provide a more detailed one. In it, several business model components, which were also found relevant by other authors, are made explicit:

> *"A business model is a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational an financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors."*

In business networks around digital platforms, platform providers need to find a way to balance value creation, conversion and capture between the different network members. That is, balanced business models have to be designed in order to secure the network's survival and thriving [53].

In the design process, CDIs are of crucial importance for the viability and sustainability of a business model [4]. Next, CDIs provided explicitly by the STOF method are presented. Additionally, implicit CDIs are derived from the VISOR method. Finally, the two lists are compared and merged.

3.2.1. CDIs in the Business Model Methodologies

This section reviews the STOF and VISOR methodologies for identifying relevant CDIs for the business modeling activity.

The STOF Method

The STOF method is a step-by-step approach for business model design that uses pre-defined design variables (or Critical Design Issues) which are considered of crucial importance for the viability and sustainability of the future business model. Table 2 provides an overview of the CDIs within two categories – CDIs related to creating customer value and those related to creating network value [4].

Customer value	Network value
1. Accessibility for Customers	1. Value Contributions and Benefits
2. Targeting	2. Division of Investments
3. Value Elements	3. Division of Costs and Revenues
4. Pricing	4. Pricing
5. Branding	5. User Profile Management
6. Security	6. Customer Retention
7. Quality of Service	7. Accessibility for Customers
8. System Integration	8. Acceptable Customer Base
9. User Profile Management	9. Network Openness
10. Customer Retention	10. Network Governance
	11. Network Complexity
	12. Partner Selection

Table 2: Critical Design Issues within the STOF method [4]

Within the STOF method, an evaluation of a quick scan stage has to be performed, based on eight Critical Success Factors (CSFs). If one of these factors is assessed negatively, then CDIs related to that factor need to be further refined until the corresponding CSF is evaluated positively [4].

The VISOR Method

El Sawy and Pereira [53] provide a small set of questions that should be answered by business model designers when dealing with each VISOR component. Even though the authors do not frame them as design variables or critical design issues, they could be interpreted as such. Table 3 provides the list of CDIs per component that can be derived from the questions given by El Sawy and Pereira [53].

Value Proposition	Interfaces	Service Platform	Organizing Model	Revenue Model
Target customers; Value elements;	Selection of interfaces	Selection of service platforms	Selection of partners	Pricing structure; Revenue sharing among partners;

Table 3: CDIs derived from the VISOR model

These CDIs are defined as perceived by the researcher from the articulation of the VISOR components given in [53]. The following section confronts the two methods and provides an merged list of CDIs that can be considered as generic business model CDIs for innovations in the ICT industry.

Confrontation

When it comes to CDIs, only the STOF method puts explicit attention on design variables that need to be addressed during the business model design. The VISOR method draws the designer’s attention to questions that must be addressed and answered. Several CDIs were derived from these questions. The confrontation between CDIs from the STOF method and the VISOR method is presented in Figure 9.

It can be seen from the figure that the STOF method provides more comprehensive list of CDIs. Most of the VISOR-specific CDIs are covered within the STOF method. However, there are two CDIs identified in the VISOR method that are missing in STOF – selection of interfaces and selection of service platforms. The former requires choices of interfaces that would enable the delivery of a specific type of value thus leading to a “wow” experience [53]. The latter relates to choices about specific service platforms that can enable the value delivery to the intended customers (e.g. IaaS platforms in this thesis project).

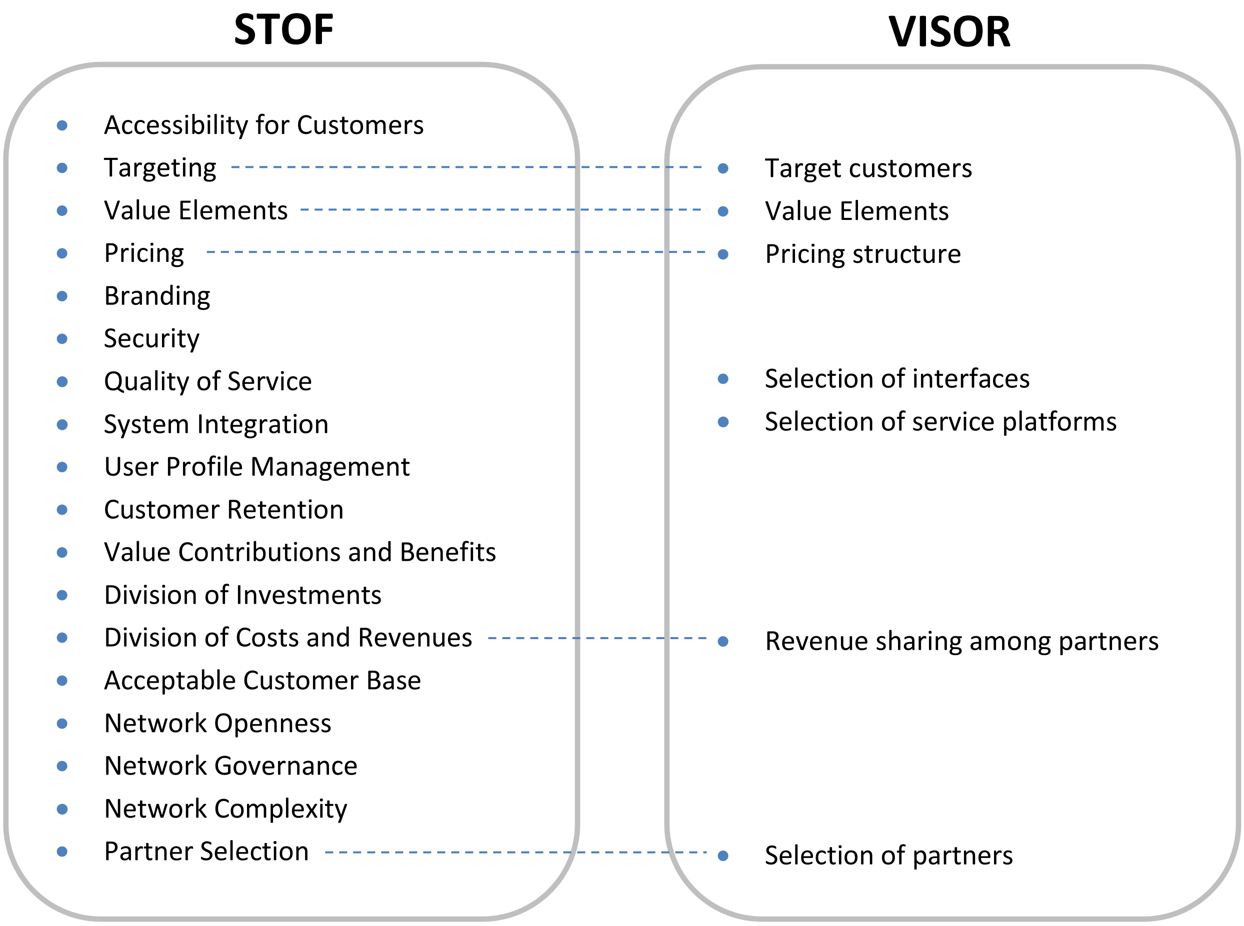

Figure 9: Confrontation of CDIs identified in the business model theory

Figure 10 provides the merged list of business model CDIs identified while reviewing the STOF and VISOR methods.

STOF

- Accessibility for Customers
- Targeting
- Value Elements
- Pricing
- Branding
- Security
- Quality of Service
- System Integration
- User Profile Management
- Customer Retention
- Value Contributions and Benefits
- Division of Investments
- Division of Costs and Revenues
- Acceptable Customer Base
- Network Openness
- Network Governance
- Network Complexity
- Partner Selection

VISOR

- Selection of interfaces
- Selection of service platforms

Figure 10: CDIs from the STOF and VISOR methods

We regard the list of CDIs given in Figure 10 as generic for product/service innovations within the ICT industry. Next, literature review identifies CDIs that are related more closely to the nature of the envisioned IS artifact – a multi-sided platform. For that reason, platform theory is reviewed first.

3.3. Platforms

The concept of 'platform' has been defined in many different ways in the existing scientific literature. Several 'families' of definitions can be recognized – product platforms, industry platforms, service platforms, and ICT platforms.

Robertson and Ulrich [54] define a *product platform* as a collection of assets from four categories (i.e. components, processes, knowledge, and people and relationships) that are shared by a set of products. Similar to this definitions is the definition of Meyer and DeTore [55]. They define product platforms as "common architectures spanning multiple products that are implemented with common subsystems and subsystem interface". Or put more simply, a product platform is a set of elements and interfaces between them that can be used across different products. An example of a product platform is a universal electric motor for all consumer power tools up to 650 watts in power [55]. Van Vuuren and Halman [56] describe a *platform* (i.e. a product platform) as neither a single product nor a product family[4], but instead as "the common basis of all individual products within a product family". They argue that decoupling of product elements is crucial for achieving a separation of common (i.e. platform) elements from differentiating (i.e. non-platform) ones. Furthermore, two important requirements exist if a product family is to be built on top of a product platform. First, a certain degree of modularity for decoupling of elements is needed. And, second, a possibility must exist for standardizing product subsystems and/or interfaces [56].

Moving on to the second family of platform definitions, Cusumano [58] argues that *industry platforms* have the same functions as product platforms (i.e. to provide core technology that can be reused in different products) but at the same time they also differ in two essential aspects. First, an industry platform is a part of a bigger technology "system" which includes components from different companies, called complements. And, second, the industry platform has little value to its consumers in the absence of these complementary products or services [58]. Examples of industry standards are operating systems, game consoles, and smartphones. Gawer [59] defines an industry platform as a building block which provides an essential function to technological systems and thus acts as a foundation upon which other organizations can develop complementary technologies, products, or services. She gives Microsoft Window as an example of an industry platform, for which many complementary software applications were developed by software developers.

The third family of platform definitions is around *service platforms.* According to Evans, Hagiu [60] platforms comprise of various modules of code that "provide services to be used by other parts of the operating system or by software applications". An Application Programming Interface (API) is used for providing access to these services. Bouwman, Faber [8] describe service platforms as middleware platforms which provide generic business functions like authentication, billing and customer care, or more specific functions like the provision of location or context information.

[4] A product family is a series of closely related products that share identical components 57. Meyer, M.H. and J.M. Utterback, *The Product Family and the Dynamics of Core Capability*, in *MIT Sloan Management Review*. 1993.

Service platforms can be incorporated in the technical architecture of a new IT service thus their functions do not have to be developed as they can obtained from a service provider.

The final family of definitions is formed around *ICT platforms.* According to Bresnahan [61], "A platform is a shared, stable set of hardware, software, and networking technologies on which users build and run computer applications". Liu [62] defines SaaS platforms as the underlying hardware and software used for distributing hosted applications in a multi-tenant, on-demand fashion.

Table 4 below gives a concise summary of the definitions presented above.

Definition Family	Description	Authors
Product platform	Set of elements and interfaces between them that can be used across different products.	Robertson and Ulrich [54]; Meyer and DeTore [55]; Van Vuuren and Halman [56];
Industry platform	A building block that acts as a foundation upon which other organizations can develop complementary technologies, products, or services.	Cusumano [58]; Gawer [59];
Service platform	Middleware platforms which provide generic or more specific business functions	Evans, Hagiu [60]; Bouwman, Faber [8];
ICT platform	A shared, stable set of hardware, software, and networking technologies on which users build and run computer applications	Bresnahan [61]; Liu [62];

Table 4: Families of platform definitions in the literature

It can be observed that all definitions have one thing in common. That is, platforms provide a basis of functionality upon which different (innovative) products or services can be developed.

When we consider the nature of the artifact that this study aims at designing (i.e. SaaS platform) then we can argue that the most comprehensive definition of a platform is the one given by Bresnahan [61]. The cloud infrastructure, software development tools, and the global Internet represent the set of hardware, software and networking technologies on top of which the service provider and platform users can build and run different applications created for the subscribers of the service.

After we have defined the platform concept for the purposes of this research project we turn again to the existing literature for describing different phenomena and concepts that are closely related to platform theory and need to be considered during the artifact design and business modeling activities. First, we discuss network externalities and their relation to platforms.

3.3.1. Network Externalities

Network externalities can be exhibited in two occasions. First, direct networks externalities arise when the benefit for a consumer from using a service increases when the number of other consumers of the same service increases [1, 63]. A classic example of one such service is telephony. Second, indirect network externalities can also arise when complementary goods are important.

More precisely, consumers would use a particular service only when a set of complementary products or services is available to them [1].

In the case of an ICT platform, the different applications that run on top of the platform can be considered as the complementary products. That is, the more applications a platform offers, the more consumers would be willing to subscribe for the platform. In turn, the more consumers subscribe, the more developers of complementary applications would be willing to join. If complementors are allowed to join the platform, then the number of complementary products will rise even more. Thus, the self-reinforcing cycle from see Figure 11 can be created [1].

Whether a platform provider decides to let developers of complementary products join the platform is then a strategic decision. This relates to the concept of platform openness which is discussed later in this chapter.

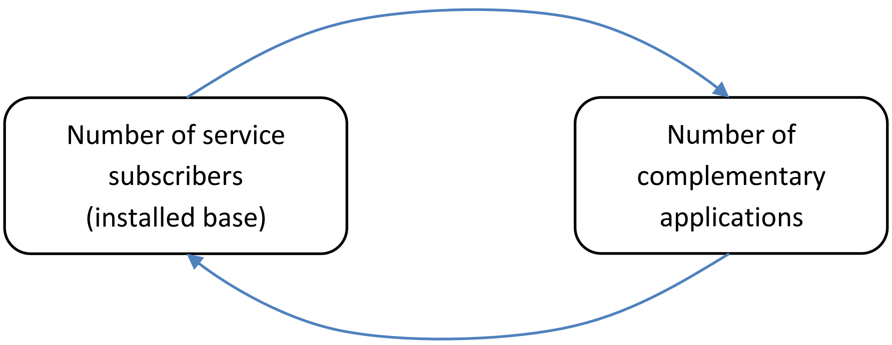

Figure 11: Self-reinforcing cycle of number of service subscribers and number of complementary applications, adapted from [1]

Indirect network externalities usually lead to the so called, “chicken-and-egg” problem [64]. In essence, the problem lies within the fact that a new ICT platform would not be appealing to complementors since there are not that many service subscribers yet, while at the same time it would not be appealing to subscribers too since there are not that many complementary applications. Finding a way to solve this problem is of critical importance to platform providers.

Platforms that connect distinct groups of users create multi-sided markets (also two-sided markets in case there are only 2 groups). The next section describes the main characteristics of one-sided and multi-sided markets, as well as their differences.

3.3.2. One-Sided and Multi-Sided Markets

One-sided markets are characterized with a traditional value chain, where value creation is sequential. That is, value is added by having the output of one’s activities as the input of another’s [3]. Figure 12 presents the value and revenue flows in one-sided markets - value is flowing from suppliers, through enterprises, to customers and end-users, while revenue is moving in the opposite direction. Everything left of the company is cost, while everything on the right is revenue [65].

Figure 12: Value and revenue flow in one-sided markets, adapted from [3]

Car manufacturing can be considered as a good example of one-sided markets. Car manufacturers buy materials or pre-made car parts (e.g. engine) from their suppliers then they assemble the whole product (i.e. the car) and sell it to their customers. Customers pay the price for the car to the manufacturer who, in turn, pays to his suppliers for the materials or parts.

On the other hand, in multi-sided markets, two or more parties interact on a platform where indirect network externalities exist [63]. Rochet and Tirole [66] argue that most of the markets exhibiting network externalities are characterized by the presence of two (or more) distinct sides whose benefit comes from their interaction on a common platform. That is, a platform provider acts as an intermediary that matches the needs of some groups with the activities of others. Thus, the join participation of the each group makes the platform more valuable to them [60]. Figure 13 below represents the flows of value and revenue in two-sided markets.

Figure 13: Value and revenue flow in two-sided markets, adapted from [3]

Application stores illustrate this type of markets. The application store (i.e. the platform) is the intermediary between application developers (i.e. complementors) and the end-users. Application developers get visibility through the store, while end-users get access to a large pool of applications with advanced search capabilities [3].

Multi-sided markets are different from one-sided markets since multi-sided platforms must serve two or more distinct groups of consumers in order to generate demand from them [67]. In such markets participants' value rises as demand from both sides is being matched [65]. Furthermore, each of the involved groups represents a source of both cost and revenue [3], [65]. However, usually one of the sides is treated as a profit source, while the other is treated as financially neutral or as a loss [66].

Interestingly though, markets that are usually organized around multi-sided platforms could prove to be viable as one-sided markets as well. If a company chooses to integrate vertically into the supply of a product/service component then it can effectively exclude third-party suppliers and thus operate in a one-sided market [60]. Evans illustrates this scenario with the case of Apple's iPod, where the company produces its own hardware, software and operates its own content-provision software.

According to Evans [67] three conditions must be met for the emergence of a multi-platform business. First, there must be two or more distinct groups of potential customers for the platform. Second, externalities must emerge when connecting or coordinating these groups. And, third, an intermediary is necessary to internalize these externalities. Thus, the platform provider must make sure that these conditions are met in the market they want to position the platform in.

After reviewing the theory on network externalities and multi-sided markets, the following CDI can be identified:

CDI: Two-sided Market Dynamics

How to start the market dynamics between the different sides of a new platform in order to attract consumers to each side, or in other words - how to solve the 'chicken-and-egg' problem?

Next, the subject of platform openness is discussed as it deals with the degree to which distinct groups are allowed to participate in the development of the platform and its complements.

3.3.3. Platform Openness

Platforms are regarded as 'open' or 'closed' to the extent that they restrict their users in participating in the development, commercialization, or use of the platform. In 'open' platforms restrictions may still be placed on reasonable terms and in a non-discriminatory fashion, i.e. is to all participants. Conforming to technical standards and license fees are examples of such restrictions [2]. Furthermore, platforms should not be characterized as open or close on a generic level, but instead on a participant role level. Four roles were described by Eisenmann, Parker [2] - demand side users, supply side users, platform providers, or platform sponsors (see Figure 14).

Platform providers mediate the transactions between users from both sides and serve as their main point of contact with the platform. On the other hand, platform sponsors do not deal with users, but instead they can modify the underlying technology and have the rights to determine who may participate in the platform as providers and users [2].

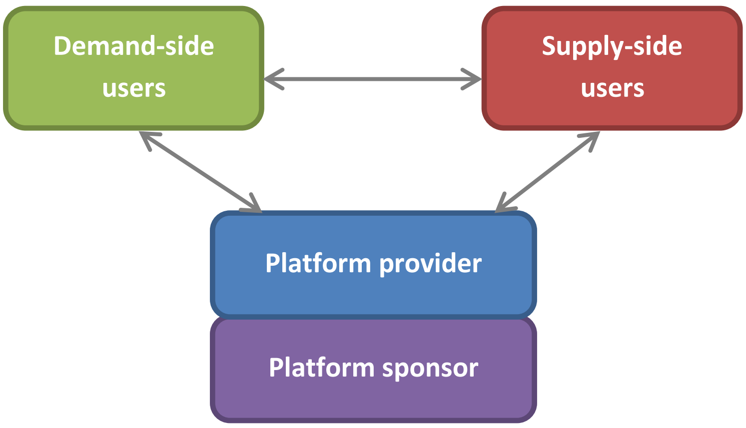

Figure 14: Platform participant roles [2]

Platforms can be open to some of the participant groups (e.g. demand side and supply side users) while staying closed to others (e.g. platform providers and platform sponsors) [2]. Cusumano [68] describes such platforms as "open, but not open" or "closed, but not closed".

Furthermore, Cusumano [58] argues that companies usually lack the necessary resources and capabilities to create all the complementary applications and services that would attract end-users. Thus, platform providers (or platform sponsors) should build a strategy for opening up their technology. Other firms should be economically incentivized to join the business network around the platform and start developing complementary applications. By opening up the platform for particular groups of users, the self-reinforcing cycle between the number of service subscribers and the number of complementary applications (see Figure 11) can be initiated.

CDI: Platform openness

- Which stakeholders are considered as demand side users, supply side users, platform providers, and platform sponsors?
- How open should the platform be for each of these user roles?

Platform openness represents the next CDI identified in platform theory. Important questions related to this CDI are given in the box above.

3.3.4. Platform Leadership

Cusumano and Gawer [69] describe platform leadership as "the ability of a company to drive innovation around a particular platform technology at the broad industry level". They argue [70] that if a company wants to become a platform leader then they need to address both the technology and business aspects of platform leadership. The former includes designing the right technical architecture, application interfaces, and disclosing intellectual property selectively to key partners. While the latter deals with deciding whether to develop complements internally or to open up the platform for external complementors. Furthermore, Gawer and Cusumano [71] identify four levers that can assist platform management in both platform strategy formulation and implementation:

1. *Scope.* Scope refers to the amount of complements that a platform leader does in-house versus the amount it leaves for external parties (i.e. platform openness). That is, platform leaders must decide whether it is better to develop solely all complementary products, to leave this to the market, or to have a combination of both. In doing so, it is necessary to assess the dependence of the platform on complements and to consider how to increase demand for the platform.

2. *Product technology.* Platform leaders must decide on the architecture of the platform. This includes both the high-level platform design and the design of interfaces which determines how sub-systems should work together. For instance, modular architecture is deemed useful when a platform is open, i.e. it is public knowledge how components can be connected to the platform. However, in this way the inner architecture of the platform can be exposed to competitors. Thus, platform leaders must make a conscious decision on sharing technical specifications with outsiders. A critical design issue has already been identified as "Level of openness".

3. *Relationship with external complementors.* Platform leaders must balance between collaborating with external complementors (to jointly increase the potential size of the platform's market) and competing with them (to stimulate new complements market) at the same time. Thus, handling tensions of conflicts of interest is a necessary ability of platform leaders.

4. *Internal organization.* This lever is relevant mainly for large companies which take the role of platform leaders. Their internal organization must allow for an effective relationship management with outsiders. A scenario that needs to be managed is when some parts of a company are collaborating with complementors, while at the same time other parts are competing with the same outsiders.

Based on the four levers above, we can identify the following CDI:

CDI: Platform Complementary Services

- What complements to develop in-house and what complements to leave to external firms?
- What will be the balance between cooperation and competition with external complementors?

Finally, as part of the literature review on platforms, pricing strategies for platform providers are discussed.

3.3.5. Pricing Strategies

In multi-sided markets, platform providers need to determine a price for each side by taking into consideration what the consequences will be for the other sides' growth and willingness to pay [65]. Usually, platform pricing structures are heavily skewed towards one side of the market [67]. Caillaud and Jullien [64] describe such pricing strategies in terms of "divide-and-conquer", i.e. the participation of one side is subsidized (divide) while the loss is recovered from the other side (conquer).

One option for platform providers in the entry phase is to set low or zero prices for one side of the market in order to achieve critical mass of participants there [67]. By critical it is meant the mass that creates strong network effects which draw users to the other sides of the market [65]. Thus, this strategy could prove to be useful for overcoming the "chicken-and-egg" problem (see the Network Externalities sub-section above). A typical example of a zero pricing towards one end of the market is Adobe Reader where the reader software is given for free in order to increase demand for the paid production software.

If enough subsidy-side users are attracted then users from other sides would be willing to pay more to reach them thus effectively recovering the subsidy costs incurred by the platform provider [65]. Different pricing models for multi-sided platforms can be found in the literature. Dou, Wu [72] discuss three of them – subscription fee model, where users are charged with a fixed fee for a set period (e.g. a month); license fee model, where users are charged only one time; and time-limited freemium, where users are not charged initially but at a later point. Additionally, cloud-enabled platforms may charge on a pay-per-use model, where users pay for the resources that they have consumed during a fixed period. Hybrid pricing models are also possible.

Another pricing strategy involves attracting particular participants on one side of the platform which are highly attractive to users from other sides. The former are referred to as "marquee buyers" in the literature [66]. By adopting this strategy, the platform provider can charge little or not at all the marquee buyers and increase the prices to the participants from the other side as described above. The latter would still be willing to pay the high prices since they have a propensity to sell products/services to their target groups. For example, American Express was able to charge merchants with higher prices since the latter viewed the American Express cardholders (usually corporate spenders) as extremely attractive [67].

Alternatively, platform providers can achieve critical mass by assisting consumers from one side of the platform in order to lower their participation costs. For instance, providing consulting services for free could be attractive to consumers on one side of the market. It is also important that such benefits discourage consumers to switch to rival platforms [67].

Pricing represents the last critical design issue identified while reviewing the scientific literature on platforms. Questions related to pricing which should be addressed explicitly during the business model design are given in the box below.

CDI: Pricing

- Which sides of the market should be subsidized and for how long?
- Are there marquee users that should be subsidized?
- Which sides should pay a premium?
- What pricing models should be provided by the platform?

Platform providers have to be able to create healthy business ecosystems based on business models that are viable not only for them but also for all network actors and end-users [70]. Thus, business ecosystems and any associated CDIs are discussed next.

3.4. Business Ecosystems

The concept of a 'business ecosystem' was first introduced by Moore [73] in 1993. He made a strong analogy between biological and business ecosystems, which are both "formed by large, loosely connected networks of entities" [74]. Other authors also recognized this analogy in their work [75-77]. According to Iansiti and Levien [75], members of a business ecosystem ultimately share the ecosystem's faith, regardless of their strength, just like individual species share the faith of a biological ecosystem. As a consequence, organizations in business ecosystems are influenced both by their internal capabilities and by the external interactions within the ecosystem [74].

According to the Moore's definition [73], in business ecosystems different companies "co-evolve capabilities around a new innovation: they work cooperatively and competitively to support new products, satisfy customer needs, and eventually incorporate the next round of innovations". That is, companies within an ecosystem work collectively (i.e. cooperate) in order to get an advantage over other competitive networks [75]. As a result, competition is moving from the product to the network level. That is, product-related issues like quality, price, and ownership become just prerequisites for competitive success. On the other hand, network dimensions like the availability of complements, size of the installed base, and customer expectations on the ecosystem's current and future size become more important for market dominance [76].

Moore [73] also argued that companies should not be viewed as part of a single industry, but instead as part of a business ecosystem which crosses multiple industries (e.g. Apple in the music, hardware, software industries [53]). Drawing the exact boundaries of business ecosystems is not an easy task, though. The reason is that hundreds of companies can exist within a single business ecosystem. In order to cope with that, organizations should try to systematically identify ecosystem members with which their future is tightly intertwined and to determine their most critical dependencies on them [75].

Business ecosystems can emerge around an anchor point, i.e. a core technology [76] which brings different member groups together. The next sub-section presents platforms as such technology and gives more insights on the relation between platform and business ecosystems.

3.4.1. Platforms and Business Ecosystems

According to Den Hartigh and Van Asseldonk [76], a business ecosystem is "a network of suppliers and customers around a core technology, who depend on each other for their success and survival". This definition has been selected as most appropriate in the context of this project since a digital platform can be considered as a core technology around which a business ecosystem can emerge. That is, by developing a multi-sided platform and attracting participants to the different sides, a company can effectively create a business ecosystem.

Iansiti and Levien [78] also stress the role of platforms as the basis of business ecosystem emergence. They regard platforms as sets of functionality that are packaged and shared with an ecosystem's members through public interfaces (e.g. APIs). Members on their end can leverage these functionalities and create value on top of them for other customers in the business ecosystem. This view on platforms is in accordance with the platform definition of Bresnahan [61] that was selected in the previous section since the public interfaces of a platform represent the software (and underlying hardware and networking technologies) that platform users can leverage for developing complementary products or services.

Fragidis, Tarabanis [79] argue that a platform should be viewed as a central entity in customer-centric business ecosystems as it sets the technological and business standards for member interactions; it coordinates these interactions; and controls their technical aspects. In business ecosystems, platforms create value (directly and indirectly) and support its dissemination by both offering mechanisms for customer participation and leveraging resources and capabilities from other ecosystem members [79].

With respect to the above, platforms are considered as the main building block for the emergence of business ecosystems in this project. They represent the means for an ecosystem's members to create and share value and are fundamental to its functioning and competitive success [78]. Next, we provide a categorization of these members depending on their behavior which we then use within the business ecosystems governance sub-section.

3.4.2. Member Roles

Different types of members can be distinguished in business ecosystems just like different species exist in biological ecosystems. Iansiti and Levien [74] provided a categorization of these members depending on the way they influence an ecosystem's health and evolution. They distinguished keystones, dominators and niche players. Each of these roles is described next.

1. *Keystones. Keystones* are ecosystem members which serve as hubs in the network of interactions between other members. They provide the foundation upon which niches can be created; regulate connections between other participants; and strive to increase diversity and overall productivity. This foundation takes the form of a stable platform that other members depend on for creating value and disseminating it across the ecosystem. The removal of a keystone and its platform would lead to the collapse of the whole business ecosystem [74].

2. *Dominators.* In biological ecosystems, dominators are species which overtake a large part of their ecosystem by eliminating other members and/or their functions. In this way they effectively limit diversity and the value an ecosystem can provide to other species. In business ecosystems, dominators play the same role – they progressively try to take-over the ecosystem by eliminating or not allowing other firms in their market. They try to acquire most of the value generated in the ecosystem and leave as little as possible to other members. Consequently, dominators damage an ecosystem's health since they reduce diversity, eliminate competition and stifle innovation. Eventually, competitive ecosystems which have a healthier structure with bigger diversity of keystones and niches can replace a dominated ecosystem [74].

3. *Niche players.* Niche players are the most numerous members of a business ecosystem. Companies are regarded as niche players when they exhibit typical levels of connectivity to other members of the ecosystem [74].
   Oftentimes, niche players locate themselves in the innovation frontier of a business ecosystem where they develop new products or services and try to explore new markets. In that sense, they are crucial for the growth and overall health of an ecosystem. However, in some cases, a conflict between keystones and niche players can emerge. Niche players that lag behind the innovation frontier of a business ecosystem may find their products incorporated in the keystone's platform core [74].

According to Iansiti and Levien [78] keystone members should be aware of the opportunity to use platforms as means to shape their business ecosystem. When customers adopt a platform they invest in systems and capabilities that are required to integrate and use the platform's products or services. Thus, customers will incur switching costs which will prevent them from leaving the keystone's ecosystem. However, relying solely on switching costs is not a sustainable strategy for keystones since niche players can be part of many business ecosystems at the same time. Sooner or later the niche players' switching costs may diminish and when that happens they will leave the platform. Iansiti and Levien [78] argue that a more sustainable strategy for keystones is to constantly seek ways to innovate and increase the value created within their platform.

Basole [80] argues that adopting a dominator strategy could be beneficial in mature industries where the pace of innovation is slow. However, dominators can be perilous to emerging industries since such market behavior stifles innovation thus can significantly reduce the competitiveness of a business ecosystem [80].

Iansiti and Levien [74] favor keystone strategies over dominator strategies since they lead to an effective and sustainable way of dealing with innovation and niche creation in the long term. Nevertheless, some keystones may be tempted to become dominators since dominator strategies can be very beneficial in the short-term. However, if platform providers turn to dominators the whole business ecosystem may collapse in the long term as already mentioned. A platform provider's choice of an ecosystem role (i.e. keystone or a dominator) is therefore a strategic decision that has implications early in the design of a platform. Thus, the following CDI is identified:

CDI: Business Ecosystem Strategy

Which role should a platform provider take in its business ecosystem – keystone or dominator?

3.4.3. Governance

Den Hartigh and Van Asseldonk [76] consider the term governance as more appropriate than management when a company tries to influence other members of a business ecosystem. The reason is that ecosystems have networked rather than hierarchical structures and therefore cannot be managed, but instead should be governed [76]. Von Tunzelmann [81] broadly defines governance as "organizing collective action" that consists of the structure, control, and process of decision-making. Structure refers to the forms through which decisions are made, control is about the power to make these decisions, and process refers to the particular implementation of structure and control.

Vos [82] described business ecosystem governance as providing network members with the incentive and vision to strive for a common goal and giving them the freedom to reach that goal on their own without obstructing their motivation. Business ecosystem governance also includes the use of steering mechanisms to ensure that member activities will reach the common goal in an effort of improving the business ecosystem's capability of coping both with its internal pace of innovation and with external change [76].

We can see that platform providers, keystones to be more precise, are the network members who can effectively govern business ecosystems. First, they decide on the levels of platform openness for each side of the market, as already discussed in the previous section. Second, they can organize collective action towards the realization of a common goal, as discussed by Vos [82]. Keystones can facilitate the pursuit of the common goal by taking specific decisions on different aspects of the platform's core technology (e.g. components, interfaces, rules, etc.) [2]. They should also decide on the structure and power of decision-making for the platform's technology. That is, do they solely decide on the platform's components and interfaces or do they allow other members to provide input as well. A critical design issue on business ecosystem governance is identified.

CDI: Business Ecosystem Governance

- Should other members take part in decisions related to components, interfaces and other technology aspects of the platform?
- To what common ecosystem goal can all members adhere and strive to?
- What steering mechanisms can be used to help the ecosystem members in achieving the common goal?

The topic of business ecosystem governance marks the end of our literature review. Next, the chapter conclusion summarizes the findings.

3.5. Conclusion

This chapter had the purpose of answering the first research question: *What are business models, platforms, and business ecosystems and what critical design issues related to these concepts can be drawn from the existing knowledge base?* The first section of the chapter described literature review as the research methodology selected for answering the first research question.

The second section of this chapter defines the business model concept first. For the purposes of this research project, a business model was defined as: *"a blueprint for a service to be delivered,* *describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational an financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors"* [83]. The STOF and VISOR methods were reviewed with the purpose of deriving business model CDIs. We have chosen to use the STOF method for designing the business model in the first place as it provides the most comprehensive list of CDIs. Two more CDIs from the VISOR method were added to this list.

In the context of this research project, platforms were defined as "*a shared, stable set of hardware, software, and networking technologies on which users build and run computer applications*" [61]. Different concepts, closely related to platforms that were identified and discussed are network externalities, multi-sided markets, platform openness, platform leadership, and platform pricing strategies. Four platform-specific CDIs were identified during the literature review.

The third section of this chapter introduced business ecosystems. Different scientific authors [73, 75-77] gave parallels between biological and business ecosystems. The latter are formed by loosely connected networks of organizations that emerge around a core technology (e.g. a digital platform). In business ecosystems different stakeholders, or members, are engaged in cooperative activities for developing complementary products for the core technology. The types of members include keystones, dominators and niche players. Keystones are usually the members that provide a business ecosystem's core platform thus they were identified as the members which can effectively govern business ecosystems. Two CDIs related to business ecosystems were identified in the scientific literature review.

In order to address the specific nature of the designed artifact during business modeling, we provide an adapted list of CDIs (Figure 15). Our starting point is the list of generic STOF CDIs. We supplement this list with 8 additional CDIs from the VISOR method, platform theory and business ecosystem theory.

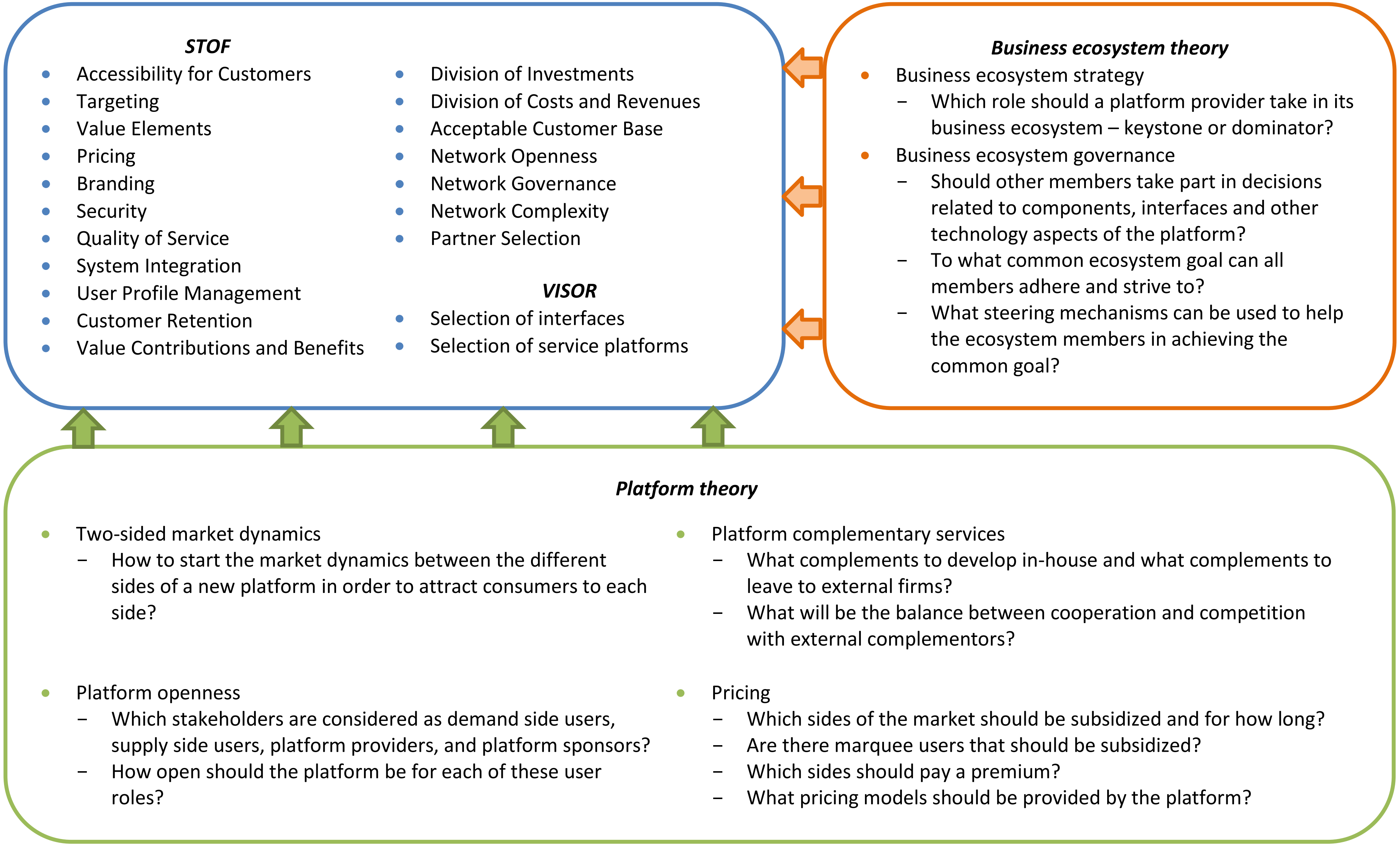

Figure 15: Generic ICT CDIs and platform-specific CDIs identified in the existing knowledge base

Next, we match each additional CDI to a particular domain and CSFs from the STOF method. Table 5 shows the results of the matching.

Critical Design Issue	STOF domain(s)	STOF Critical Success Factors
Selection of interfaces (VISOR model)	Technology domain	Compelling value proposition
Selection of service platforms (VISOR model)	Technology domain	Compelling value proposition; Acceptable quality of service; Acceptable profitability; Sustainable network strategy
Two-sided market dynamics (Platform theory)	Service domain	Compelling value proposition; Clearly defined target group
Platform openness (Platform theory)	Service domain	Compelling value proposition; Acceptable quality of service
Platform complementary services (Platform theory)	Organization domain	Acceptable quality of service; Sustainable network strategy; Acceptable division of roles
Pricing (Platform theory; STOF)	Service domain; Finance domain	Compelling value proposition; Clearly defined target group; Unobtrusive customer retention; Acceptable quality of service; Acceptable profitability; Acceptable division of roles
Business ecosystem strategy (Business ecosystem theory)	Organization domain	Acceptable profitability; Sustainable network strategy; Acceptable division of roles
Business ecosystem governance (Business ecosystem theory)	Organization domain	Sustainable network strategy; Acceptable division of roles

Table 5: Positioning of artifact-specific CDIs within the STOF method

It is necessary to draw parallels between some of the additional CDIs and the core ones within the STOF method.

First, STOF already includes a “service platforms” design element in its technology domain. However, we have decided to include an explicit CDI for this element as we are designing a SaaS information system. Thus, the selection of infrastructure service platforms is regarded as crucial for the viability of the business model.

Second, “network openness” within STOF and “platform openness” are very similar. However, we prefer using the latter CDI as it makes it very explicit that separate levels of openness can be selected for each distinct customer group (i.e. participant role).

Third, the finance domain of the STOF method includes a “pricing” CDI. However, this generic CDI does not take into account pricing in multi-sided markets. Pricing in such markets is very specific as different pricing models can be set for different customer groups. Therefore, we argue that the two CDIs should be combined.

Finally, “network governance” and “business ecosystem governance” should be compared. The former CDI addresses selection of collaboration partners, setting the collaboration rules, and rule compliance monitoring [83]. On the other hand, “business ecosystem governance” addresses how the keystone member (or platform provider) governs the evolution of the entire ecosystem. Therefore, their focus is different, so we can distinguish two separate CDIs.

The next chapter discusses the relevance of big data analytics and the envisioned artifact for SMEs. Results of interviews with SME representatives are discussed in order to identify concrete business needs for big data analytics adoption. In addition, the five assumptions that led to the idea of developing a big data SaaS platform are tested.

4
Business Needs Identification

This chapter's position within the research framework can be seen in Figure 16. It has the purpose of answering the second research question:

> ***Are big data analytics and the envisioned SaaS platform relevant for small and medium-sized enterprises?***

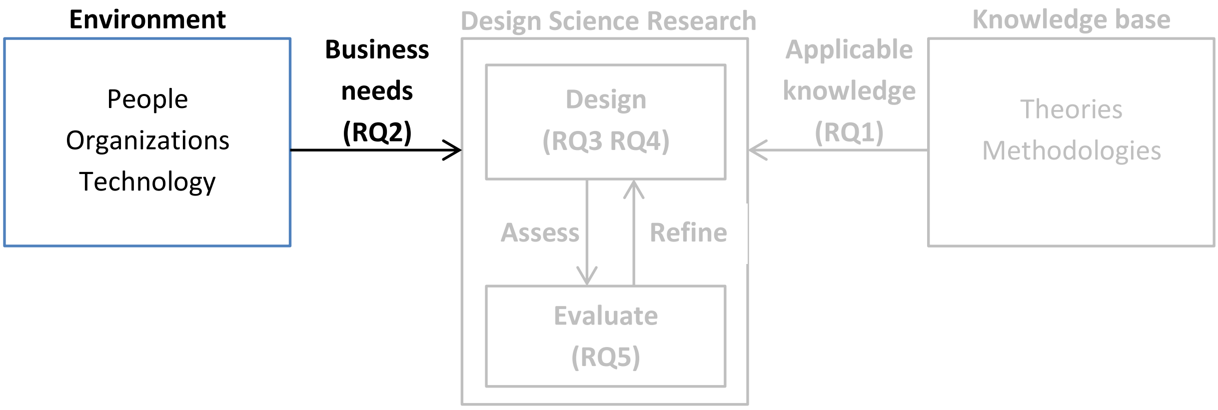

Figure 16: Positioning of the "Business Needs Identification" chapter within the research framework

This chapter discusses the relevance of both big data analytics and the designed SaaS platform for SMEs. In this way compliance of the design research with Hevner's second design guideline is ensured (see Appendix C).

The following section describes the research methodology and performed activities for answering the second research question. The second section of the chapter presents the research findings. These include the concrete business needs that were identified and the validation results of all assumptions underlying the idea. Finally, the chapter conclusion summarizes the findings and provides the main conclusions from this research phase.

4.1. Research Methodology

Qualitative data for answering the second research question is collected by conducting semi-structured interviews with SME representatives. The interviews have two distinct goals. First, identification of concrete business needs for big data analytics is sought. A business need can be stemming either from a problem that an enterprise is currently facing or from a new business opportunity that they have identified [5]. Second, validation of the assumptions underlying the idea for the designed artifact is pursued. That is, we want to assess whether the big data SaaS platform, as currently envisioned, can indeed create value for the group of SMEs. The researcher selected the semi-structured interview approach since the two goals are explorative in nature and might ask for deviations from a fully structured interview protocol.

Four main activities were performed by the researcher as part of this phase: development of interview plan and interview protocol; selection of respondents; conducting the interviews; and analysis of collected data. Next, each activity is described in more detail.

4.1.1. Interview Plan and Interview Protocol

The first activity performed by the researcher in this phase was to create an interview plan and interview protocol. The developed interview plan has the following structure and sequence:

Figure 17: Interview Structure

All conducted interviews followed the structure from Figure 17. First, SME representatives were prompted if they perceive any business needs for big data analytics within their organization. Then, the assumptions which lead to the idea for a big data SaaS platform that connects SMEs to data scientists were tested. After that point, the envisioned IS artifact was introduced to each respondent. The researcher considers that describing the designed artifact early might frame the respondents' answers and introduce bias in the collected data. Finally, data on the respondents' perception of the described solution was collected. The reader is referred to Appendix D for the detailed interview protocol, containing the complete list of interview questions and the reasoning behind each of them.

4.1.2. Selection of Respondents

We have chosen to use a convenience (non-probability) sampling technique for selecting interview respondents because of two main reasons. First, generalizability of results to the entire SME population in The Netherlands is *not* sought at this point. We argue that the platform would still be successful if it meets a single business need for a single company in an efficient and effective manner. Second, the researcher has limited time to conduct this project phase. Consequently, potential respondents that are easily accessible were contacted. Dialogues Technology provided the researcher with the necessary contacts. In total, 5 employees of SMEs within different industries were contacted by email and/or phone (see Table 6; information denoted with * is obtained from LinkedIn company profiles).

#	Person position	Company	Company Size (in employees)	Industry*
1	Interim Manager	Company A	51-200*	Financial services
2	Chief Operations Officer	Company B	220	Mortgage services
3	Manager Marketing Intelligence	Company C	250	Publishing
4	Database Marketing Manager	Company D	25	Loyalty management
5	Director Operations Media	Company E	51-200*	Media production

Table 6: Overview of contacted respondents and companies

Four out of the five contacted candidate companies replied and were interviewed (Company A, B, C, and D). The respondent from Company A is currently taking the position of Interim Manager. He has practical experience in the following areas: outsourcing; transition management; IT strategy; IT innovation; and Internet of Things. The second respondent is the COO of Company B. He has extensive experience in both operations management and financial management. The third respondent is the manager of Marketing Intelligence department within Company C. His department is responsible for, among other things, maintaining the customer databases; analyzing whether customers are satisfied with the newspaper and their future needs; and reporting. Finally, the fourth interviewee is responsible for the database marketing team and customer insights within Company D. He has extensive marketing (management) experience in different industries (airline, telecommunications and sports).

4.1.3. Conducting the Interviews

All interviews were recorded with audio recording device after asking the respondent for their explicit permission. During the interviews, the researcher took notes for asking follow-up questions. In this way, the respondent was not interrupted unnecessarily. After the end of the interview, usually on the same day or the following day, the researcher wrote down the interview report. The interview recordings were used as an aid during this process.

All of the respondents were asked if they are willing to read the interview report for validation purposes. All of them indicated their willingness to do so however only two (Companies A and B) actually provided back their comments. In one of the cases, a few minor adjustments were reflected back in the final interview report.

4.2. Interview Findings

For the full transcripts of all four conducted interview, the reader is referred to Appendix E. The following sub-section presents all identified business needs for big data analytics adoption as perceived by the respondents.

4.2.1. Identified Business Needs

All four participants indicated that they perceive particular business needs for incorporating big data analytics in their current business processes. We categorize these needs based on the source of needed data.

Social media

Social media content was identified by all participants as a potential source of unstructured external data that could add value to their companies. As social media incorporates many types of data, we will further specify the specific types that were mentioned during the interviews. Table 7 gives an overview of these data types and some examples for each type.

Data type	Examples
Personal data	Relationship status, employment status, health status
Likes, interests	Companies, public figures
Post content	Personal opinion on particular products/services

Table 7: Types of social media data that were identified as relevant by the respondents

Respondents from both companies within the financial industry indicated that it would be of value for them to incorporate personal data in their current processes. Company A could assess the creditability of a person more accurately if they knew their relationship status. Company B could use changes in the relationship, employment, or health status of people to identify individuals who are likely to get into financial trouble before they actually do.

The respondent from company C indicated that they could provide more tailored news content to their subscribers based on their needs and wants as specified in social media (e.g. likes, interests).

The respondent from company D indicated a business need for matching social media posts to people profiles from their database and subsequently analyzing the post content (e.g. sentiment analysis) for marketing purposes.

Website contents

In the particular case of company C, the respondent recognized a business need for analyzing corporate website content data for enriching their current company profiles database.

Audio

The respondent from company C also recognized a need for a service that could analyze audio data and generate meta-data for subsequent discovery purposes.

Real-time Positioning

The interviewee from company D communicated a business need for having access to real-time positioning data of individuals from their customer database that meet certain criteria related to shopping patterns.

4.2.2. Assumptions Validation

The validation results for each assumption that led to the idea for the designed artifact (i.e. big data SaaS platform which connects SMEs to data scientists) are discussed in turn next.

Assumption 1: SMEs lack the resources necessary for drawing insights from big data.
When we look more closely at the selected companies we can see that for three of them business intelligence can be regarded as one of their core competencies. Consequently, it is of strategic importance for them to have BI-related resources in-house.

In contrast, the respondent from company C indicated that they do lack the necessary resources for drawing insights from big data. Even though, company C also has a business intelligence department, BI is not one of their core competencies.

Consequently, in three of the four interviews we reject this assumption.

Assumption 2: SMEs can recognize cloud computing as effective means to access IT infrastructure at low cost without the need for maintenance.
The respondent from company B indicated that it is not feasible for them to put personal financial data on the public domain (i.e. public cloud infrastructure) due to strict privacy policies imposed by their data providers (i.e. banks and insurers).

In all other cases (except for the case of company B) we validated this assumption.

Assumption 3: SMEs lack the competencies necessary for drawing insights from big data.

The respondents from companies B, C and D indicated that they have the in-house capabilities needed to draw insights from big data. The reason can be attributed to the fact that they currently employ business intelligence professionals.

In the case of company A, the respondent differentiated between the competencies necessary for drawing insights from structured data and big data respectively. In his opinion, company A has limited competencies when the latter is concerned. Thus, we validated assumption 3 for this case.

Consequently, this assumption is rejected in three of the four cases. The reason for this result could be attributed to the sampling of the SMEs. Each of the companies has undergoing business intelligence initiatives for which they employ professionals who are likely to know which data should be analyzed and what type of analysis can be done in order to meet a particular business need.

Assumption 4: Hiring big data analytics consultants is not feasible for SMEs due to high costs associated with consultancy services.

Another interesting and unexpected result was that all companies indicated that if it is necessary, they would hire business analytics consultants to help them with drawing insights from big data. Companies B and C are in the top segment of the SME category which could be the reason why they can afford hiring business analytics consultants. In the case of Company C, the respondent explicitly indicated that he found such consultants very expensive, but that they are also “worth it”.

Companies A and D are smaller than the other two. The former is part of a large holding while the latter is owned by several of its large clients. Consequently, both companies can access financial resources from larger players in their network.

In conclusion, we reject this assumption in all four cases.

Assumption 5: Value can be created by connecting SMEs to data scientists in a business ecosystem around a SaaS platform.

The results of the validation for this assumption mirror the results for assumption 3. The reason is attributed to the fact that respondents who perceive that they have the internal competencies for drawing insights from big data do not see value in getting connected to external data scientists.

We reject this assumption in three of the four cases. Again, we believe that the sampling of SMEs is the reason for most rejected assumptions. Nevertheless, the interviews led to one important design choice. They are presented next in the chapter conclusion.

4.3. Conclusion

This chapter had the purpose of answering the second research question: *Are big data analytics and the envisioned SaaS platform relevant for small and medium-sized enterprises?*

Four interviews with representatives from SMEs were conducted in order to answer this research question. Table 8 provides a summary of the business needs for big data analytics that were identified during the interviews. Additionally, the columns labeled as A1 to A5 describe whether the initial assumptions of the researcher are verified or rejected for each separate interview.

#	Company	Business Needs	A1	A2	A3	A4	A5
1	Company A	Build a better picture of the creditability of debtors, thus lower risk and define new propositions	✘	✔	✔	✘	✔
2	Company B	Identify clients that might get into financial problems before they are in serious trouble	✘	✘	✘	✘	✘
3	Company C	Match social media data to existing customer database; Supplement in-house data on Dutch companies with data from the web; Analyze sound to derive meta-data	✔	✔	✘	✘	✘
4	Company D	Match product-related messages from social media (Twitter) to existing purchaser profiles database; Determine in real-time if a big customer is about to enter a particular store to place a purchase	✘	✔	✘	✘	✘

✔ **Assumption Verified** ✘ **Assumption Rejected**

Table 8: Identified business needs and assumptions validation interview results

From the collected data we can see that all four companies perceive particular business needs for incorporating big data in their business processes. Thus, we conclude that *big data analytics is indeed relevant for SMEs*.

However, only one of the respondents indicated that connecting his company to a pool of data scientists could add value to them since they have limited competencies for big data analytics. The rest of the respondents indicated that they already have the necessary competences in-house. Thus, we can conclude that organizations employing professionals with competencies in data analytics would not find value in having access to a pool of data scientists on the SaaS platform. Consequently, we have decided to interview respondents from SMEs that lack such competences in-house for gathering their system requirements (see next chapter).

Furthermore, some of the respondents indicated that they want to analyze big data only in-house. The reason is that they want to have more control over data privacy and security issues. Consequently, we conclude that the platform, as envisioned at this moment, *is irrelevant for SMEs with similar business needs.*

Thus, the main finding from the interviews is that *SMEs have business needs for big data analytics from different calibers*. In order to adapt the platform to different calibers of business needs, we make one particular design decision. That is, *the platform should be built in a modular way*. On one hand, basic services (e.g. data collection, pre-processing, and data storage) should be available to SME end-users. On the other hand, the same basic services should be re-used as "building blocks" for more complex (end-to-end) services. We try to illustrate this choice with Figure 18.

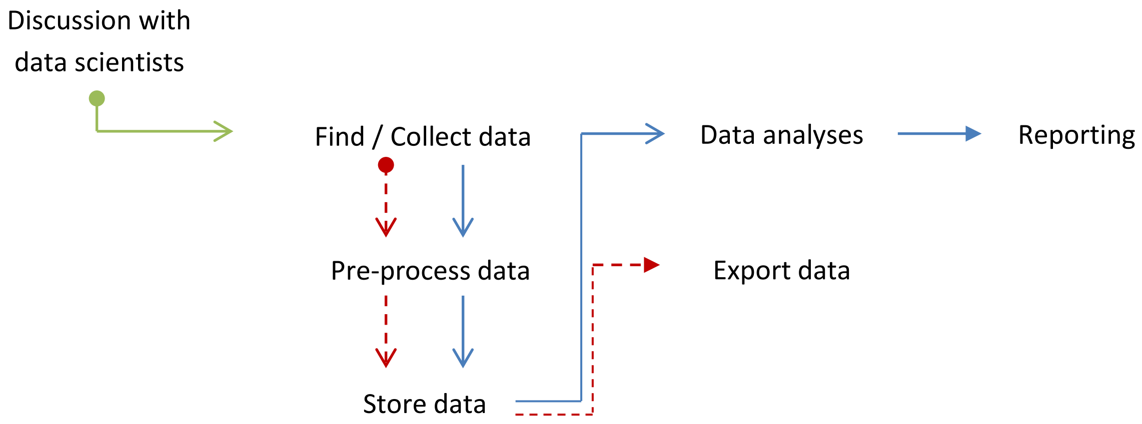

Figure 18: Example combinations of big data services on the platform

One organization (e.g. Company C) might need an end-to-end service (solid arrows), while another one might need a more limited service (dashed arrows) that exports the required data to a particular format for subsequent on-premises analyses (e.g. Company A). We argue that by implementing a modular architecture, the platform will be capable of fulfilling different calibers of big data business needs like the ones of the four SMEs in our sample. That is, SMEs shall be able to use/subscribe exactly to these features/services that they need.

By taking into account the conclusions drawn from these business needs interviews and the corresponding design choices we can now move on to the next project phase: information system design.

5
Information System Design

This chapter's position within the research framework can be seen in Figure 19. Its purpose is to answer the third research question:

> ***What are the goals, requirements and structural specifications for the big data SaaS platform?***

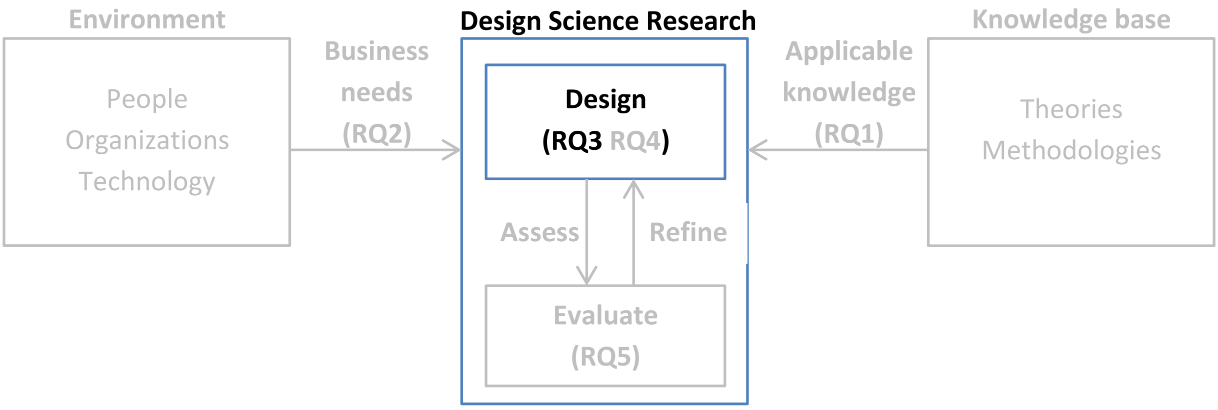

Figure 19: Positioning of the "Information System Design" chapter within the research framework

As already discussed in the introduction chapter, the research methodology selected for designing the IS artifact is the design cycle of Verschuren and Hartog [9]. Only the first three stages of this design cycle are conducted as part of this thesis due to time limitations. These stages are: (1) first hunch, (2) requirements specification, and (3) structural specifications. Figure 20 shows the outcomes of the three design stages.

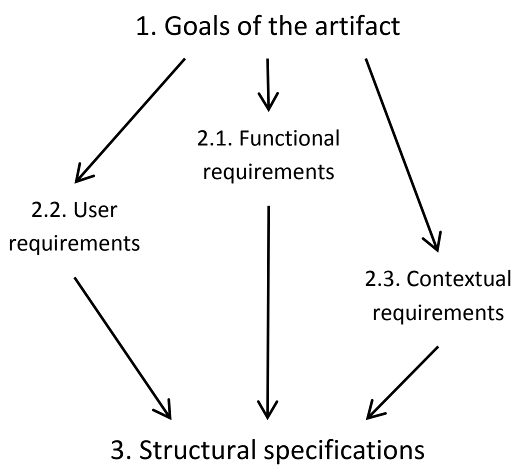

Figure 20: Outcomes of the three IS design stages

The starting point of the IS design is the definition of the goals that the platform must achieve once realized. The goals are set in the following section by taking means of desk research. In the second section we set the requirements for the system. Verschuren and Hartog [9] divide system requirements in 3 groups: functional, user and contextual requirements. *Functional requirements* indicate the functions that the designed artifact should fulfill once it is realized given its goals. *User requirements* reflect the demands of the future designed artifact end-users. *Contextual requirements* are prerequisites set by the political, economic, juridical, or social environments. Finally, in the third section, we set the structural specifications for the IS. These specifications are the characteristics, modules and parts of the artifact that satisfy the whole set of requirements defined in the previous IS design stage [9].

5.1. First Hunch Stage

In the first IS design stage, we formulate a small set of goals that the artifact has to achieve once realized. The goals are defined in the next sub-section by means of desk research.

5.1.1. From Business Needs to Goals of the Designed Artifact

The goals of the big data SaaS platform are set by taking into account both the findings from the business needs interviews and the research domain.

The main finding of the conducted business needs interviews is that *SMEs have business needs for big data analytics from different calibers*. On one hand, for some of the companies it is only feasible to use services for collecting, pre-processing, storing and exporting big data (see Figure 18). The reason is that data security and privacy restrictions might prevent them from analyzing data in an off-premise fashion. On the other hand, other companies which do not use sensitive data for analytics purposes could also use data analysis and reporting services on the SaaS platform. Thus, we decided that *the platform should be built in a modular way.*

We argue that the platform shall have several application service layers since some services will be used as "building blocks" for other (more complex) services. Within such modular architecture, SMEs shall be able to subscribe exactly to these services that they need. Thus, the first goal of the platform is:

To provide SMEs with access to big data services within different application service layers in order to meet different calibers of business needs

Furthermore, while describing the research domain, we saw that some major issues are associated with big data analytics (see *Research Domain*). That is, people sometimes mistakenly confuse correlation with causality and big data sometimes contain hidden biases that untrained people might not be aware of. Thus, we argue that SMEs and data scientists should collaborate if reliable big data services are to be developed for the former group. Consequently, the second goal of the big data SaaS platform is:

To facilitate the collaboration between SMEs and data scientists for the development of reliable big data analytics services

Once realized, the platform must meet the two goals defined above. In addition, two ideas for pilot big data services were provided by Dialogues Technology. They are described shortly in the following sub-sections.

5.1.2. Twitter Sentiment Analytics Service

The first service is sentiment analytics based on content from Twitter messages. SMEs end-users will be able to specify search terms (e.g. product/brand name) for finding relevant tweets first. After all tweets are found, content analysis will be used to determine the sentiment behind each search result.

Deriving accurate sentiments from Twitter messages (a.k.a. tweets) is a challenging task because of several reasons. First, tweets are very short (up to 140 characters) thus may lack context. This lack of context makes the extraction of emotions unreliable in some cases: for instance, sarcastic and ironic comments are hard to classify accurately. Second, tweets are relatively unstructured when compared to longer texts. Third, abbreviations and slang are regularly used by Twitter users. All of these issues reduce the overall reliability of tweet sentiment analysis.

Coming up with an accurate and reliable algorithm for sentiment analysis is considered as a major challenge for Dialogues Technology. The project team has therefore decided to use an external service provider for determining the sentiments behind tweets. SNTMNT (www.sntmnt.com) was selected for that purpose. Their algorithm is "trained to deal with unstructured context, abbreviations & slang" by using a combination of semantic models, support vector machines (machine learning) and natural language processing classifiers. Thus, several of the challenges related to tweet content analysis are addressed by SNTMNT. The company claims an accuracy of 84.7% on a binary scale and 86.9% on a three point scale which is an outperformance of approximately 10% over the best generic algorithms that they are aware of [84].

Finally, sentiment results will be aggregated and displayed in the form of easily readable graphs on the platform. By using this service SMEs end-users can choose between competing suppliers, for instance. That is, they can compare sentiment analysis results of competing brands and based on that to determine with which suppliers to work or which products to buy/sell.

5.1.3. Weather Analytics Service

The second service is predictive analytics based on weather data (e.g. air temperature, cloudy/sunny/raining, precipitation, etc.) and internal data supplied by SMEs (e.g. sales figures, number of lead calls, etc.). This service will allow SMEs first to see how weather influences different business aspects and second to predict similar weather influences in the future.

For this service to work, data scientists are required that must analyze the provided datasets first. They will determine whether correlation or causality exists between the selected (by the SME) dependent and independent variables. A data model might eventually be created which will predict future values of the dependent variable based on predicted values of the independent (weather) variables. A tailored big data service can then be built which will make use of the created predictive data model.

More information on how the two pilot services will function is given later in this chapter (see Structural Specifications Stage). The definition of goals of the designed SaaS platform marks the end of the first hunch design stage. The next stage is the specification of requirements for the designed artifact. Functional, user and contextual requirements are described in turn.

5.2. Requirements Specification Stage

This design cycle stage deals with the requirements that the artifact should fulfill given its goals set in the previous stage. Verschuren and Hartog [9] classify these requirements in three categories: functional, user, and contextual. Functional requirements describe the functions that should be fulfilled by the designed artifact once it is realized. User and contextual requirements regard the interface between the platform and its environment [9]. We have structured this section according to the above categorization.

During the requirements elicitation process several guidelines that were described by Sommerville and Sawyer [85] were followed. According to one of them, system stakeholders should be identified early in the process as they are likely sources of system requirements. Thus, before proceeding to the collection of system requirements we will first identify the different stakeholder groups of the big data SaaS platform.

5.2.1. Stakeholders Identification

Several stakeholder groups can be identified for the designed platform. Table 9 provides an overview of these groups, description of their stakes, and example representatives of each group.

#	Stakeholder group	Stake	Examples
1.	SMEs	Representatives of this stakeholder group form the demand side of the platform. They are the users of big data analytics services on the platform.	Physical / online retailers
2.	Data scientists	Representatives of this stakeholder group form the supply side of the platform. They provide knowledge for building new data analytics services.	Data analysts; Business intelligence experts
3.	Data providers	Representatives of this stakeholder group can provide data that is used as input in the big data analytics services. Data can be provided by means of Application Program Interfaces (APIs) for instance	Twitter; Weather Underground
4.	Data analytics service providers	Representatives of this stakeholder group can provide APIs to existing data analytics services that can be incorporated in the designed platform	SNTMNT
5.	Infrastructure providers	Representatives of this stakeholder group can provide IT infrastructure as a service based on the IaaS cloud delivery model.	Amazon, Microsoft
6.	Big data technology developers	Representatives of this stakeholder group provide technologies which enable the development of big data analytics services	Apache, Elastic Search, 10gen Inc.
7.	Regulators	Representatives of this stakeholder group can impose specific regulations on the system	European Commission; Dutch Government

Table 9: Big data SaaS platform stakeholders groups and stakes

The first two stakeholder groups are the future users of the designed big data SaaS platform. Therefore, their requirements must be collected during the design process and fulfilled during the development process. User requirements are collected from representatives of both SMEs and data scientists stakeholder groups (see the *User Requirements* sub-section in this chapter).

The third and fourth stakeholder groups provide either data or data analytics services which can be used as part of the platform's services. Therefore, contextual requirements from these stakeholders groups are addressed as well (see the *Contextual Requirements* sub-section in this chapter).

The fifth and sixth groups are infrastructure providers and big data technology developers. Possible requirements coming from these two stakeholders should also be considered during the elicitation of contextual requirements as they are the technology enablers of the designed artifact.

The last stakeholder group could impose requirements on the system by passing laws and monitoring for their compliance.

After the identification of system stakeholders we can continue with the elicitation of system requirements. We start with functional requirements.

5.2.2. Functional Requirements

Functional requirements indicate the functions that the designed artifact should fulfill once it is realized given its goals [9]. Consequently, we have decided to derive the platform's functional requirements from its goals. Next, we list and describe the requirements for each of the two goals identified in the first design stage. In this way, we use business needs (as the goals were derived from the business needs of several SMEs during the first round interviews) to drive the functional requirements elicitation process. Thus, another guidelines of the ones given by Sommerville and Sawyer [85] is followed.

Additionally, we were able to identify several functional requirements which can be regarded as generic since they cannot be linked directly to one of the two goals. These requirements are described last.

Big Data Services Requirements

This section describes functional requirements derived from the first goal of the designed artifact:

> *To provide SMEs with access to big data services within different application service layers in order to meet different calibers of business needs*

Basic big data services from low layers shall be used by system users or re-used as "building blocks" for services at higher layers. In this way, the different business needs of different SMEs can be addressed more accurately. Thus:

FR1: Platform modularity

The platform shall be built in a modular way.

The lowest service layer shall include big data services for finding, collecting, pre-processing and storing data from various external sources. Twitter is one such external source of data that will be used for the development of the sentiment analytics pilot service.

The second service layer shall include different export and data analytics services. For illustration purposes we will use the Twitter pilot service again. An export service will generate a file in a .csv (comma-separated values) format which includes the tweet contents, its sentiment score, and the confidence coefficient. On the other hand, a data analytics service will calculate the percentage of positive, negative or neutral tweets for a given search query.

Finally, the third layer shall consist of reporting services. Services at this layer will display analyses results from data analytics services from the second layer. In the case of the Twitter sentiment analytics service, for each search query a pie chart with percentages of positive, negative and neutral results should be displayed. Reporting services at this layer should make use of different data visualization tools, thus:

FR2: Data visualization tools

Reporting services on the platform shall display analyses results by making use of tables and graph charts (pie, line, and bar charts)

The next requirements relate to the platform's user interface.

First, SMEs should be able to browse through all big data services that are available on the platform. Second, by specifying keywords and industry type they should be able to search for big data services that are relevant for them. Finally, when a particular big data service is selected, SMEs should have the possibility to "install" the service in their account with a single action.

FR3: Big data services discovery

System users from the demand side (SMEs) should be able to browse for, search and install big data services that are available on the platform.

After a service is installed, it will appear in a list of installed services for the user account. That brings us to the next functional requirement:

FR4: Overview of installed services

System users from the demand side (SMEs) shall see a list with all installed services from where they can start or manage them.

SMEs users shall start installed big data services on demand. On one hand, some services can include graphical user interface. If we take the example of the Twitter pilot service, system users first have to enter search terms for filtering purposes. After the search is complete, results have to be displayed in graphs for easier readability. On the other hand, other services could lack graphical user interface. For instance, Company A and B (interviewed during the business needs identification interviews) would need second layer services that export data to an end-point on their own IT infrastructure. SMEs must also be able to manage and configure their installed services. This includes termination of service subscriptions and service configuration.

Moving on, each service should have a list of parameters which can be used for service usage measurement. These parameters should be tracked for each system user account. If we take the Twitter pilot service as an example again, such parameters are the number of search queries or the number of returned results.

FR5: Service usage tracking

The platform shall track service usage parameters for each service.

This was the final requirement derived from the first goal of the designed platform. Next, we list the functional requirements that can be derived from the platform's second goal.

Collaboration Requirements

The second set of functional requirements is derived from the second goal of the big data SaaS platform:

> *To facilitate the collaboration between SMEs and data scientists for the development of reliable big data analytics services*

The next functional requirement entails for a catalogue of data scientists on the platform:

FR6: Data scientists yellow pages

System users from the demand side (SMEs) should be able to browse a catalogue consisting of all registered data scientists, their expertise and competencies.

These 'yellow pages' should allow SMEs to search for data scientists with particular expertise and competencies. When a particular data scientist is chosen, a discussion with them can be started:

FR7: Discussion board

System users from both platform sides should be able to start or join discussion board threads.

SMEs should be able to communicate their business needs for new or customized big data analytics services in a discussion board. On the other side, data scientists should also be able to participate in such discussions and thus demonstrate their willingness to participate in new service development. SMEs should be able to select particular data scientists for further development activities:

FR8: Group collaboration space

System users from the demand side (SMEs) should be able to create a private group collaboration space and invite selected data scientists.

The requirements for the features of the group collaboration space were collected from potential users from both platform sides during the second round interviews. Their requirements are described in the following sub-section.

Software developers from the platform provider's side should be able to join group collaboration spaces as well. The reason is that the platform provider is the one responsible for implementing new or customized services.

The next two functional requirements concern revenue for data scientists as supply side users:

FR9: Data scientists' revenue

System users from the supply side (data scientists) should be able to receive revenue from services that were developed with their participation.

and

FR10: Data scientist income overview

System users from the supply side (data scientists) shall see an overview of the income generated from subscribers to their services

The next set of functional requirements contains generic requirements for all digital platforms.

Generic Functional Requirements

The generic functional requirements are derived implicitly from the goals of the designed IS artifact. That is, the list of goals does not include generic functions however they are still vital for the proper functioning of the designed artifact. The designer's previous experience in the area of software development allowed him to identify a small list of such generic requirements for the system.

Unauthenticated users should have access only to general information about the big data analytics platform (e.g. available services, pricing). Both demand side and supply side shall authenticate themselves in order to use the services running on the platform. The reason is that other system modules (e.g. billing) require specific user information in order to function properly. For meeting this requirement, role-specific user registration and login shall be implemented. Thus:

FR11: User authentication

System users shall authenticate themselves based on their role in the market, i.e. demand side (SME) or supply side (data scientist).

The next functional requirement relates to user account management:

FR12: User account management

System users shall manage the information associated with their accounts.

Account-specific information includes general information like e-mails and passwords as well as role-specific information. For the case of SMEs, role-specific information can be company details and invoicing information. While for data scientists, it can be experience and competencies. However, these examples are only assumptions at this stage. Potential users will be prompted explicitly during the second round interviews for what data is relevant and should be entered in their user accounts.

The third generic functional requirement deals with invoicing and billing:

> ***FR13: Invoicing and billing***
>
> The platform provider shall charge system users from the demand side (SMEs) for the provision of services.

The exact way of price calculation for the different types of users is not of our concern here (the pricing schemes of the platform are discussed in the next chapter). What is important from a functional point of view is to implement means for customer invoicing and customer billing. The former shall be developed by the platform provider as it requires access to specific customer account information. The latter will be realized by integrating the system with third-party payment processors (i.e. external service providers). When choosing which particular payment providers to integrate with, user requirements should be taken into account. For this reason, a separate question was asked to potential users during the second interviews round (see Appendix F).

Overall, 13 functional requirements were derived from the goals of the designed artifact. They represent the functionalities of the platform as envisioned by the researcher and project team (within Dialogues Technology) at this stage. During their elicitation several items for which input from potential users is needed were identified. Next, we describe the approach of gathering user requirements for the IS artifact.

5.2.3. User Requirements

The next part of this stage is the elicitation of user requirements for the designed big data SaaS platform. The interview data collection method was used for the purposes of user requirements collection. Initially, a focus group study was planned as SME end users and data scientists are expected to collaborate during the development of new big data analytics services. However, due to issues with participants' availability from the SMEs side, the data collection method was changed to interviews.

Selection of Respondents

Two respondents from each end-user group (i.e. SMEs and data scientists) were interviewed in the second round of interviews. From the SMEs group, we contacted potential candidates who met two criteria. First, they must own a small company in the retail industry. The retail industry was chosen as we assume that owners in this sector would find added value in the two pilot services. Second, they must not employ professionals with competencies in drawing insights from data (see the second learning point from the previous chapter). On the other platform side, data scientists with different backgrounds were contacted in order to get a more diverse picture of the requirements they might have. Table 10 below shows the respondents that were selected for the second round of interviews.

#	Respondent Background	User group
1	Owner of a boat accessories shop	SMEs
2	Owner of a baby products shop	SMEs
3	Conducts research in the ICT services market: • user satisfaction with ICT service providers in The Netherlands; 1. user satisfaction with ICT services provided within a particular company;	Data scientists
4	Ph.D. candidate with knowledge in quantitative research methods	Data scientists

Table 10: Second round interviews respondents

Project time limitations prevented us from conducting more interviews with potential end-users. The gathered requirements will be used for the development of a system prototype. After a tangible prototype is developed (based on these requirements) Dialogues Technology will evaluate it together with a broader sample of potential users. However, the development of a prototype and its evaluation are outside the scope of this thesis. These activities are indicated as future research recommendations in the final chapter of the report.

Interview Plan and Interview Protocol

The developed plan for the interviews with potential users from both sides of the platform followed the structure and sequence from Figure 21:

Figure 21: Second round interview structure

The main functionalities of the platform (identified in the *Functional Requirements* section above) were described at the start of each interview. After that, the two pilot services (sentiment analytics on Twitter content and analytics based on weather-related data) were discussed with the respondents. By doing so we aimed to reduce the abstractness of the designed SaaS platform by providing context to the conversation. Finally, after we made sure that the respondent understands the idea for the platform and the two pilot services we proceeded by asking for their requirements.

Each group of potential users was asked two sets of questions. The first one includes general questions for both groups. In this way we followed another guideline of Sommerville and Sawyer [85] by collecting requirements from different viewpoints. The list of general questions includes items on non-functional system properties such as device accessibility, availability, customer support, etc. The second set of questions contains group-specific questions. They are divided in two categories – user interface and costs. The questions were derived from the main functions of the artifact. Also, input from the pricing CDI was used for some questions in the latter category.

The reader is referred to Appendix F for detailed information on the interview protocol and the two sets of questions.

Conducting the Interviews

As in the first round of interviews, all second round interviews were recorded with audio recording device (with the explicit permission of the respondents). During the interviews, the researcher took notes for asking follow-up questions. In this way, the respondent was not interrupted unnecessarily.

After the end of the interview, usually on the same day or the following day, the researcher transcribed the respondent's requirements. The interview recordings were used as an aid in this process. Next, we discuss how we analyzed the collected data on user requirements.

Data Analysis

For the purposes of data analysis we plotted all identified requirements on a matrix as rows. The columns of the matrix corresponded to the four conducted interviews. In this way conflicts and overlaps could be noticed more easily. The reader is referred to Appendix G for details on the results of data analysis and the set of collected user requirements.

5.2.4. Contextual Requirements

Next to the functional and user requirements we define contextual requirements in this sub-section. Verschuren and Hartog [9] describe these contextual requirements as prerequisites that are set by the designed artifact's political, economic, juridical or social environment. After listing the system's stakeholders we were able to define several contextual requirements for the big data SaaS platform.

Currently the most used big data technologies are open source, i.e. Hadoop and mongoDB. That is, their code can be modified in case very specific performance optimizations are required. Thus, we argue that open source technologies and standards should be implemented in the designed artifact. We recognize that open source can have its negative sides like lack of detailed documentation or unavailability of customer support. Consequently, before choosing which technologies to use, we need to review their documentation and see whether a sufficient developer community exists. Thus:

CR1: Open source technologies

The platform shall be built by using open source technologies and standards with detailed documentation and large developer community.

Furthermore, JSON and XML are the most used formats for data interchange[5] between different systems. Therefore, the platform shall be able to read both of them when receiving data from data providers' APIs.

CR2: Data interchange formats

The platform shall support the JSON and XML open data formats when receiving data from external data provider APIs.

The next requirement is related to browser compatibility:

[5] By May 1st, 2013 out of 9070 APIs registered in www.programmableweb.com 7970 used the JSON and/or XML data formats. This is approximately 88% of all registered APIs.

CR3: Browser compatibility

The platform should be compatible with the browsers of at least 90% of the Internet users.

According to the most recent statistics published by w3schools, the five most used browsers for March 2013 are Google Chrome, Mozilla Firefox, Internet Explorer, Safari, and Opera [86]. Together they cover 99.1% of the used browsers. From the same source we also derive the browser versions that must be supported in order to meet the above requirement. Table 11 presents the selected versions and the cumulative percentage of browser users for these versions.

Browser	Versions to be supported	Cumulative percentage of users
Google Chrome	C 24 – C 27	48.8%
Mozilla Firefox	FF 16 – FF 21	23.2%
Internet Explorer	IE 8 – IE 10	11.9%
Safari	S5, S6	4.0%
Opera	O12	1.1%
	Total:	**89.0%**

Table 11: Browsers and versions to be supported by the big data SaaS platform

By supporting the browsers and version from the table above a total coverage of 89% of the Internet users will be achieved. Since 89% is less than the required 90% we will also add FF 14 and FF 15 to the list of supported browsers. W3scools does not provide exact statistics for these two Firefox versions for 2013, but in December 2012 they accounted for approximately 2% of the used browsers. Thus, we believe that by adding support for these two browsers we will pass the threshold of 90%.

The final contextual requirement shall become applicable only in case at least one big data service on the platform processes personal data.

CR4: Personal data protection

The platform should comply with EU's Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

The obligations of the platform provider as a "data controller" are set out in EU's Data Protection Directive [87].

5.2.5. Requirements Summary

In this design stage, we formed a set of functional, user and contextual requirements that the designed big data SaaS platform needs to fulfill once realized. These requirements will be used for the subsequent development of a system prototype. The aim of Dialogue Technology is to follow an agile approach and develop and evaluate the prototype as quickly as possible with potential users. Consequently, at this stage we do not pursue high completeness of requirements since this is a very time consuming effort (due to large number of potential requirements and conflicts). In addition, creating extensive requirements documentation is in conflict with the principles of agile software development (which are *always* applied within Dialogues Technology).

5.3. Structural Specifications Stage

In this stage, we define the structure of the designed artifact that satisfies the requirements identified in the previous design stage [9]. We start by describing the service oriented architecture (SOA) concept. Second, the technical architecture of the platform is presented on a high level. The architecture consists of three layers: infrastructure layer, SOA-based application services layer and web interface layer. Finally, detailed technical architecture for the two pilot services is presented.

5.3.1. Service Oriented Architecture

In the context of this project, Service Oriented Architecture, or SOA, can be defined as a set of best practices for the organization and use of information technology [88]. SOA removes the need for building monolithic and tightly coupled applications. Instead, complex applications can be developed in a loosely coupled way by assembling atomic services [88, 89]. These atomic services must be autonomous in nature and with a well-defined functionality. In this way, a greater workflow flexibility is achieved: IT services can be tailored to a subscriber's particular business needs by adding, removing or changing the order of the atomic services [89].

In addition, SOA allows for better B2B collaboration. The services provided by partners and suppliers can also be integrated in a service oriented architecture [89]. This service integration is usually achieved by means of web services. That is, a supplier provides an API for accessing their services through the World Wide Web.

We have chosen to design the platform's application layer based on SOA. There are two reasons for our choice. First, based on input from the first round interviews, we learned that different SMEs have different calibers of business needs for big data analytics. Thus, a design decision was taken to build the system in a modular way. By implementing SOA, we can have atomic services as modules that can be reused for the development of new services from higher service layers. Second, for the creation of both pilot services, the platform needs to connect to either a partner's or a data supplier's web API. For instance, for the Twitter sentiment analytics service, tweets are downloaded from Twitter's API, while their sentiment analysis is done by SNTMNT. If it happens that SNTMNT has to be changed with another service provider then SOA allows for the greatest flexibility since changes need to be implemented in a single atomic service on the platform that communicates with the external API.

Next, we describe the technical architecture on a high level.

5.3.2. Technical Architecture

The high-level technical architecture for the platform is presented in Figure 22. The architecture consists of three layers – web interface, SOA-based application services layer, and infrastructure layer.

Web Interface

The highest platform layer is the web interface layer. Desktops, laptops, tablets and smartphone devices can be used for accessing the system. User accounts are stored in a relational database for the purposes of user authentication and user account management. Existing service discovery, cross-group collaboration and billing (for SMEs) / payment (to data scientists) are also part of this layer.

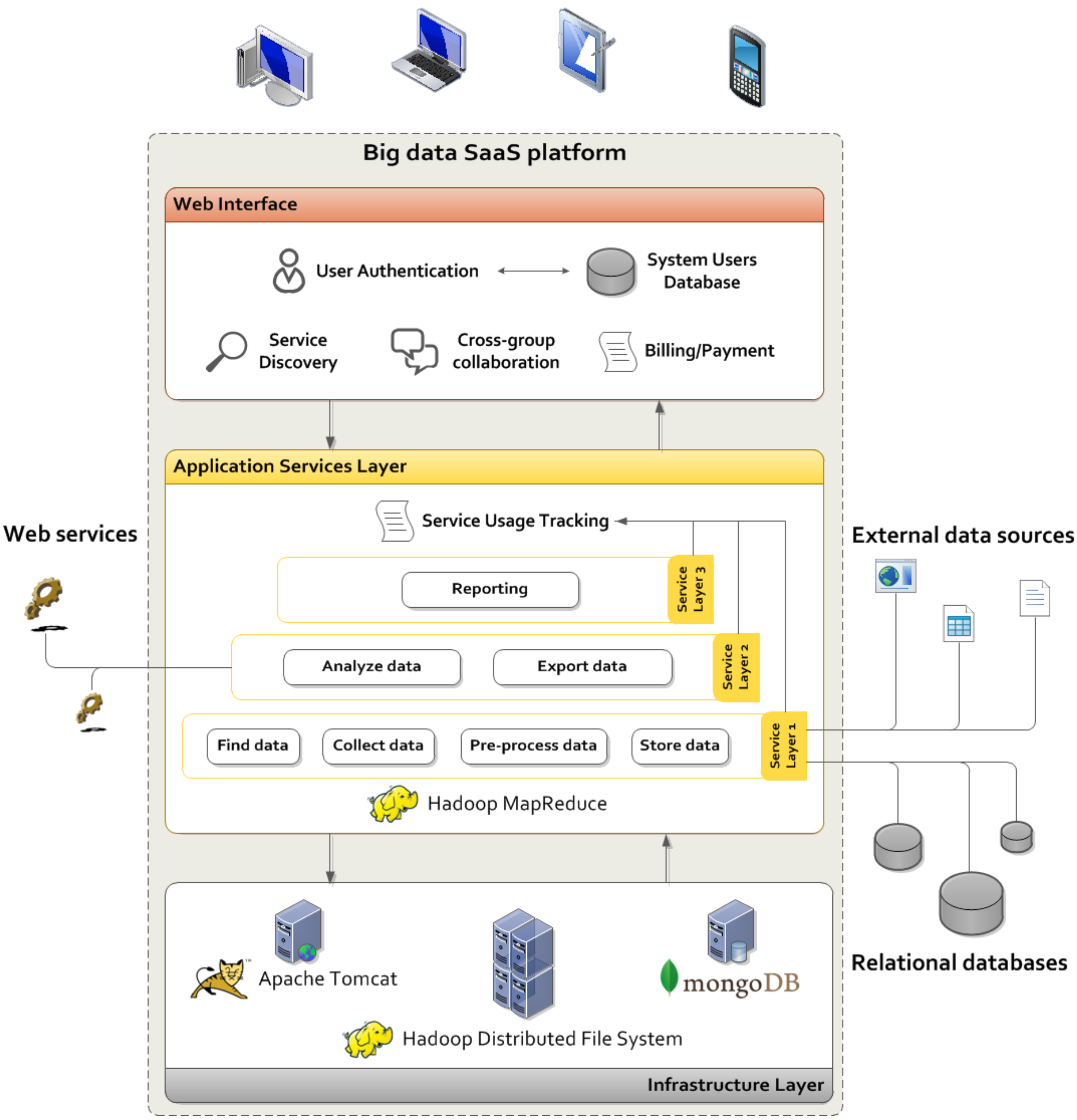

Figure 22: High-level technical architecture

Application Services Layer

The second platform layer is the application services layer. Three service layers make up the application layer of the platform's architecture. Hadoop's MapReduce framework will be implemented in order to enable parallel processing of data within services from all three layers.

The lowest service layer (i.e. service layer 1) includes services for finding, collecting, pre-processing and storing data. Two general sources of data are distinguished here. First, platform services will find and collect data from JSON or XML web APIs of external data providers (e.g. Twitter, KNMI). Second, SMEs will upload company data on the platform (e.g. sales by product/product category). In both cases, data can be pre-processed before it is stored in the mongoDB database.

The middle service layer includes data analytics and data export applications. The former will read data from mongoDB, analyze them, and write the results back in the database. Additionally, SOA allows for including third party analytics services within this layer. In the case of the Twitter sentiment analytics service, SNTMNT's web service will be leveraged for determining the sentiment score of each tweet. The latter, (i.e. data export applications) will only read data from mongoDB and either generate a dataset file to be downloaded through the platform or push the data to a customer's endpoint on their own IT infrastructure.

The highest service layer includes reporting services which display the results from the data analyses that were run on the middle level. These results will be displayed in a human-readable format. Different graphs (e.g. pie, line, and bar chart) and tables shall be used for that purpose.

By implementing SOA, services from all three layers can be flexibly assembled for the provision of applications to SMEs that are tailored to their particular needs and business processes.

Infrastructure Layer

The final and lowest layer is the infrastructure layer. It includes HDFS: Hadoop Distributed File System. HDFS is part of the Apache Hadoop framework which is specifically designed to run on commodity hardware. The framework allows for the distributed processing of large data sets across clusters of computers [90]. This feature makes Hadoop very suitable for the purposes of big data analytics as large volumes of data need to be processed in relatively short periods of time.

Second, Apache Tomcat will be used as a web server for the platform application. Tomcat was chosen since it supports Java which in turn is the programming language chosen by Dialogues Technology for the development of the platform.

Finally, mongoDB has been chosen as a schema-less database by Dialogues Technology because of the combination of flexibility, power, speed and ease of scaling. High speed and ease of scaling are especially important for the purposes of big data analytics as already noted above.

Now that we have described the technical architecture on a high-level, we can turn our attention to a more detailed description of the two pilot services.

5.3.3. Detailed Pilot Service Specifications

In this sub-section we describe in more detail the structure of the two pilot services. We start with the Twitter sentiment analytics service.

Twitter sentiment analytics service

When customers start the service, they are prompted to enter search queries (maximum 5). The first atomic service executes search the queries in Twitter's Search API. The API then returns the tweets matching the queries and they are stored in mongoDB by another two atomic services (see Figure 23).

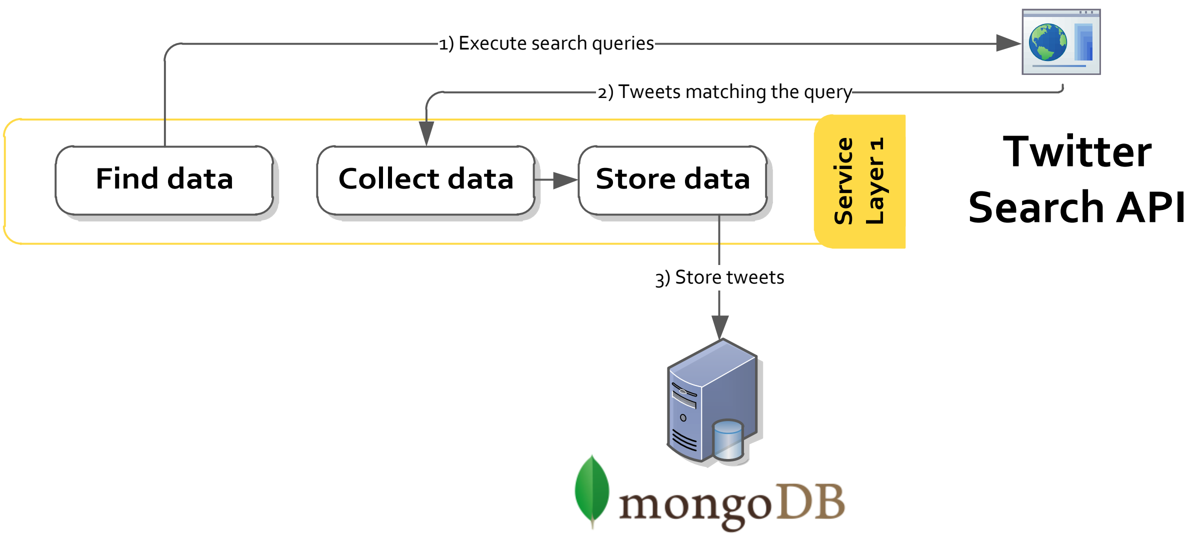

Figure 23: Twitter service architecture: layer 1

At the second layer, a service retrieves tweets from mongoDB and sends them to SNTMNT's API for sentiment analysis. The algorithm of SNTMNT uses semantic models, support vector machines and natural language processing classifiers as part of their sentiment analysis service. The algorithm they use analyzes tweet messages in a three step process [84]:

1) Pre-processing:
    - enriching unstructured content (hashtags, slang, abbreviations);
    - stemming;
    - detection of negations
2) Feature extraction:
    - NGrams;
    - Dictionaries;
    - Emoticons, pre-defined hashtags and product names
3) Machine learning model:
    - Support vector machines
    - Trained on 60,000 labeled sample tweets

The exact sentiment analysis algorithm is not disclosed by the company as it is their unique resource. The company claims an accuracy of 84.7% on a binary scale and 86.9% on a three point scale which is an outperformance of approximately 10% over the best generic algorithms [84]. Furthermore, their service is available only for tweets written in English. Support for Dutch will be added next as they are currently working on it.

SNTMNT's API returns the sentiment scores to the same service which saves them back in mongoDB. Another service then calculates the total counts of positive, negative and neutral tweets for each query and stores them in a CSV file for later usage (see Figure 24). MapReduce is used for faster calculation of these counts.

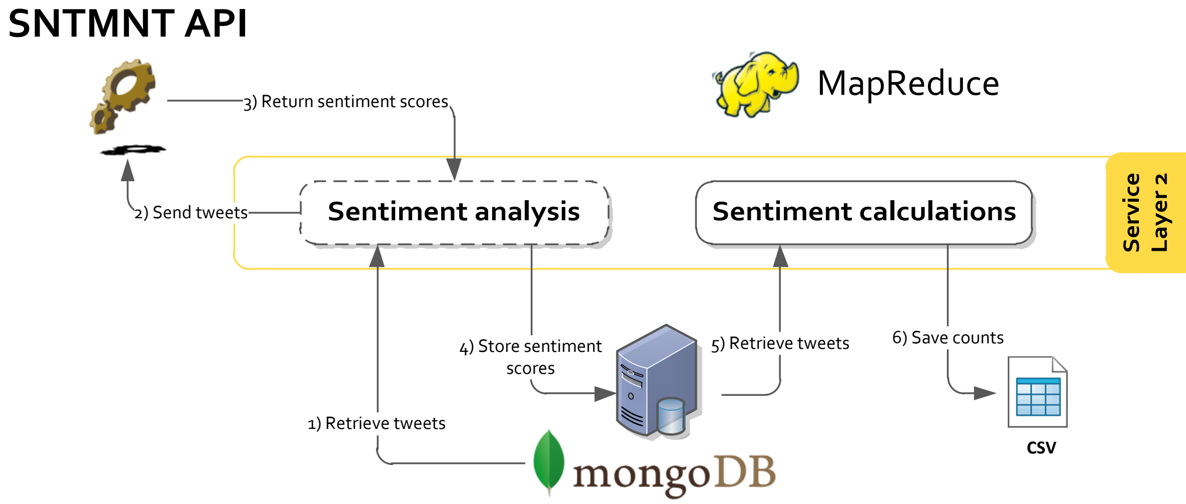

Figure 24: Twitter service architecture: layer 2

In the highest service layer, a reporting service retrieves the counts and calculates the averages for each search query. Then pie charts with the percentages for positive, negative and neutral tweets for each query are displayed on the same page for easier comparison by the service user (see Figure 25).

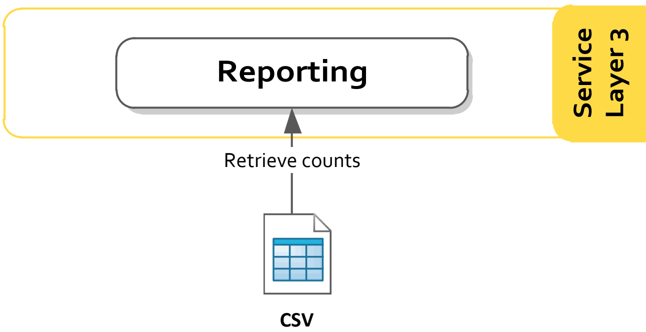

Figure 25: Twitter service architecture: layer 3

We describe the second pilot service next.

Weather analytics service

This service will initially collect data from two sources. First, open weather data from KNMI, the Royal Netherlands Meteorological Institute will be gathered. Second, SMEs will upload company performance data (e.g. sales by product/product category) from legacy systems that they are currently using. Eventually, more data types can be added to this service (e.g. flow of people or Twitter sentiments).

Both types of data will be pre-processed after collection and only relevant data will be stored to mongoDB (see Figure 26).

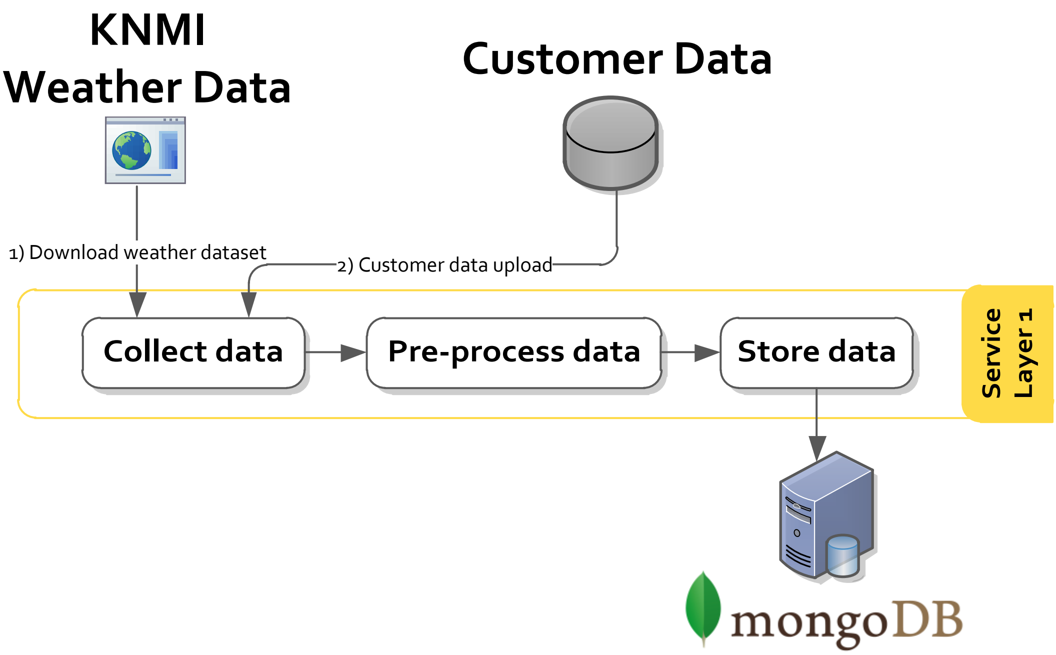

Figure 26: Weather service architecture: layer 1

At the middle level service layer, two services can be distinguished. First, data scientists will download the weather and company datasets from the platform in Excel or CSV formats. These are standard formats and most statistics software packages (e.g. SPSS, Stata, etc.) can recognize and open them. The data scientist can then build a predictive analytics data model based on the longitudinal datasets. The model will then be used for the development of a tailored predictive analytics service (see Figure 27).

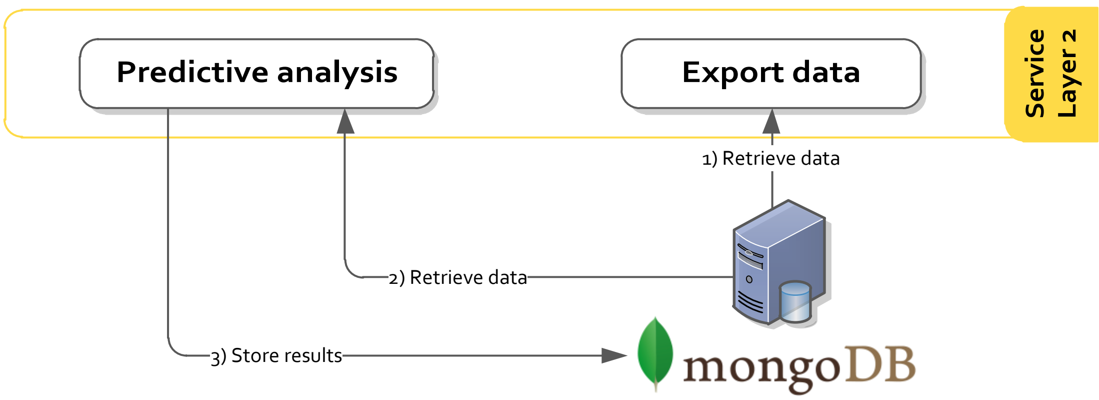

Figure 27: Weather service architecture: layer 2

In the highest service layer, two reporting services exist (see Figure 28). The first one is more descriptive in a sense that it plots raw weather (e.g. temperature) and company data (e.g. product sales) together on the Y axis of a line chart. Days or months will be plotted on the X axis respectively. In this way, trends could be spotted which might stimulate companies to contact data scientists to check whether there is indeed correlation between the variables. Also, if a correlation exists and a predictive analytics service has been developed, a reporting service shall display the results of this predictive analysis service for a user-specified period of time.

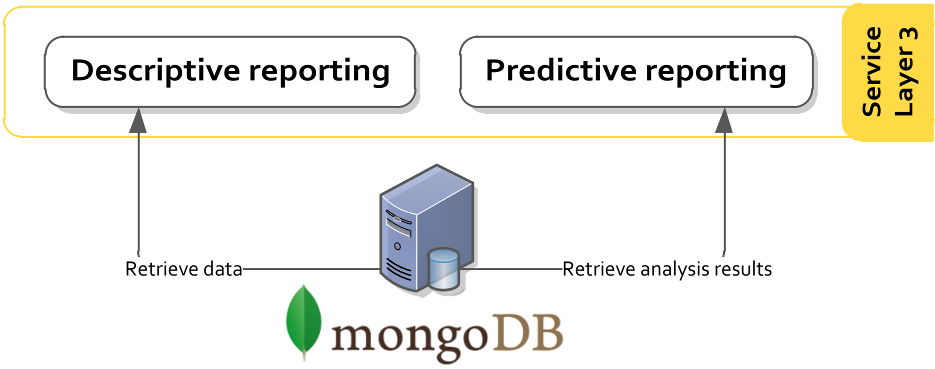

Figure 28: Weather service architecture: layer 3

The description of the two pilot services marks the end of this design stage. Finally, a section conclusion will summarize the activities that were performed by the researcher for the design of the big data SaaS platform

5.4. Conclusion

This chapter had the purpose of answering the second research question: *What are the goals, requirements and structural specifications for the big data SaaS platform?*

First, the goals of the big data SaaS platform were set by taking into account both the findings from the business needs interviews and the research domain. The following two goals were defined:

1) *To provide SMEs with access to big data services within different application service layers in order to meet different calibers of business needs*
2) *To facilitate the collaboration between SMEs and data scientists for the development of reliable big data analytics services*

Second, the functional requirements for the system were derived from the defined set of goals by means of desk research. Then, for the elicitation of user requirements, a second round of interviews was conducted. Potential users from the groups of SMEs and data scientists were asked about their requirements regarding user interface, pricing and general system aspects (see Appendix G for the list of user requirements). Contextual requirements were also defined by means of desk research.

Finally, we described the structural specifications that meet the set of system requirements. This includes the high-level technical architecture and the detailed pilot services specifications.

Based on the defined goals, requirements and structural specifications, a prototype of the system can now be developed. However, prototyping is not part of the scope of this project, thus all IS design activities that are part of this thesis are now concluded.

The goals, requirements and structural specifications described in this chapter are also used as input for the business modeling design in the following chapters. Attention is now shifted to the design of the business model and more specifically to outlining the business model behind the designed big data SaaS platform.

6

Business Model Outline

This chapter's position within the research framework can be seen in Figure 29. It has the purpose to answer the fourth research question:

How does the business model of the designed big data SaaS platform look like?

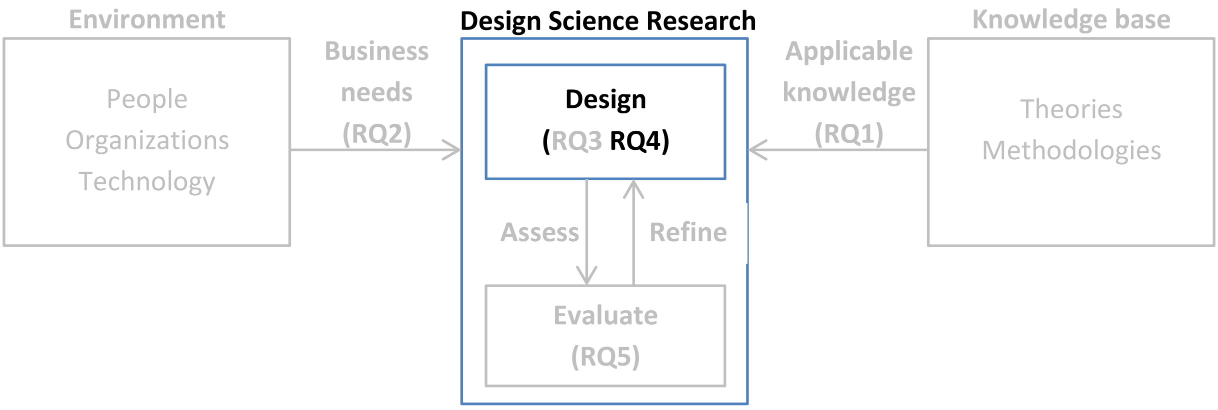

Figure 29: Positioning of the "Business Model Outline" chapter within the research framework

For the purposes of answering the fourth research question, the researcher will make use of the STOF business model framework and method [83]. The framework consists of four key domains, namely **S**ervice, **T**echnology, **O**rganization, and **F**inance. These four domains need to be designed in a careful and balanced way so that value is generated not only for the service customers, but also for the service providers [4]. The STOF method (based on the STOF model) helps designers with creating viable and feasible business models that create value for both customers and providers [83]. The different STOF method steps are depicted in Figure 30.

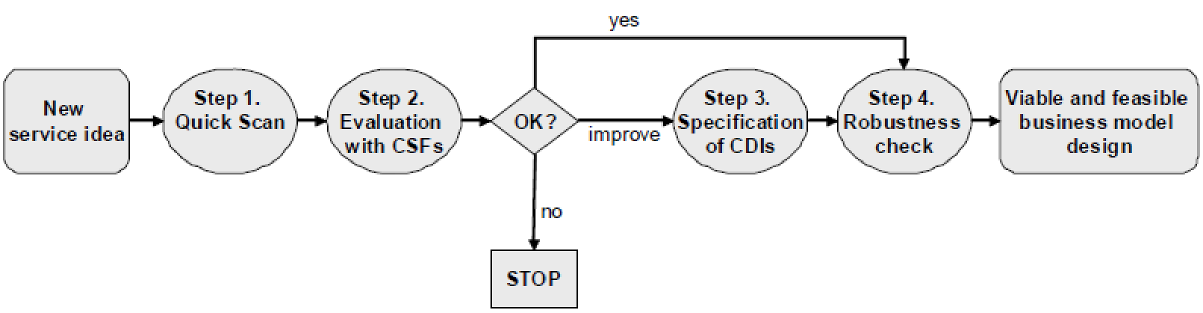

Figure 30: The STOF method steps [4]

Up to this point, no outline of the business model has been created. Therefore, in this chapter *we will create the first outline of the business model for the platform and its two pilot services* by performing the quick scan step of the STOF method. However, from the two rounds of interviews, we were able to learn about the needs and requirements of potential users of the big data SaaS platform. Consequently, we go deeper in the business model design by refining some of the domains. The knowledge that we have gained so far has implications mainly for the Service and Technology domains. We start with the description of the Service domain.

6.1. Service Domain

The starting point for the business model design is the value that the big data SaaS platform has to offer to its target customers. First, we make explicit who the target customers of the platform are.

6.1.1. Customers and End Users

We distinguish two customer groups for the big data SaaS platform: SMEs and data scientists. The former represent the demand side of the platform. The SMEs group is made up of business entities thus the end users will be employees within business organizations. Data scientists form the supply side of the platform. This group includes individuals who are also the system's end-users.

As two different customer groups exist, we have to make a distinction between them when we describe the rest of the business model components within the Service domain.

6.1.2. Two-Sided Market Dynamics

The artifact that is being designed in this thesis project is a two-sided platform. Consequently, when designing a business model from the platform provider's point of view, it is important to formulate a clear strategy on how to balance the demand and supply for the services available on their platform.

Starting the market dynamics on the platform would require building a customer base of particular size on one platform side which will in turn attract system users from the other. We have decided to build a customer base of SMEs first for several reasons. First, launching customers from the SME group can be attracted by developing services that do not require interaction with data scientists through the platform. Second, SMEs customers are the main revenue source for the platform. Securing cash flows as early as possible would allow for the further development of the platform through reinvestment. And, third, data scientists interested in receiving revenues would not join the business ecosystem if there is no demand for their knowledge and capabilities. After a customer base of SMEs has been secured and a mass market strategy has been adopted, the platform provider will try to attract data scientists to the platform.

Next, we narrow down the customer groups from each platform side by choosing target segments in the market.

6.1.3. Targeting

It can be argued that the customer group of SMEs is very broad - it includes companies below 250 employees from various industries in The Netherlands. However, one of the goals of the designed SaaS platform is to provide big data services to SMEs from different industries, i.e. a mass market strategy. However, the mass market cannot be reached immediately as the platform's technical capabilities have to grow with the growth of its base of complementary big data services.

Initially, customized big data services will be developed for SME customers from niche markets who aim at gaining competitive advantage over their rivals. As the platform's technical capabilities grow, a mass market strategy can be pursued by targeting one or two segments first (i.e. "beachheads"). However, big data services for the mass market should be more generic in nature. At this point, we have chosen to develop the two pilot services described in the previous chapter for this purpose. SMEs are segmented based on their industry sector. We assume that the Twitter sentiment analytics service will add value to SMEs from the retail sector, while the weather analytics service will be valuable to companies in the retail, food, and entertainment (e.g. cinemas, event organizers, etc.) industries. Thus, we identify SMEs in the retail and food industries as our two beachheads in the mass market.

A simple walk-through of how SME end-users will use the platform is presented in Figure 31.

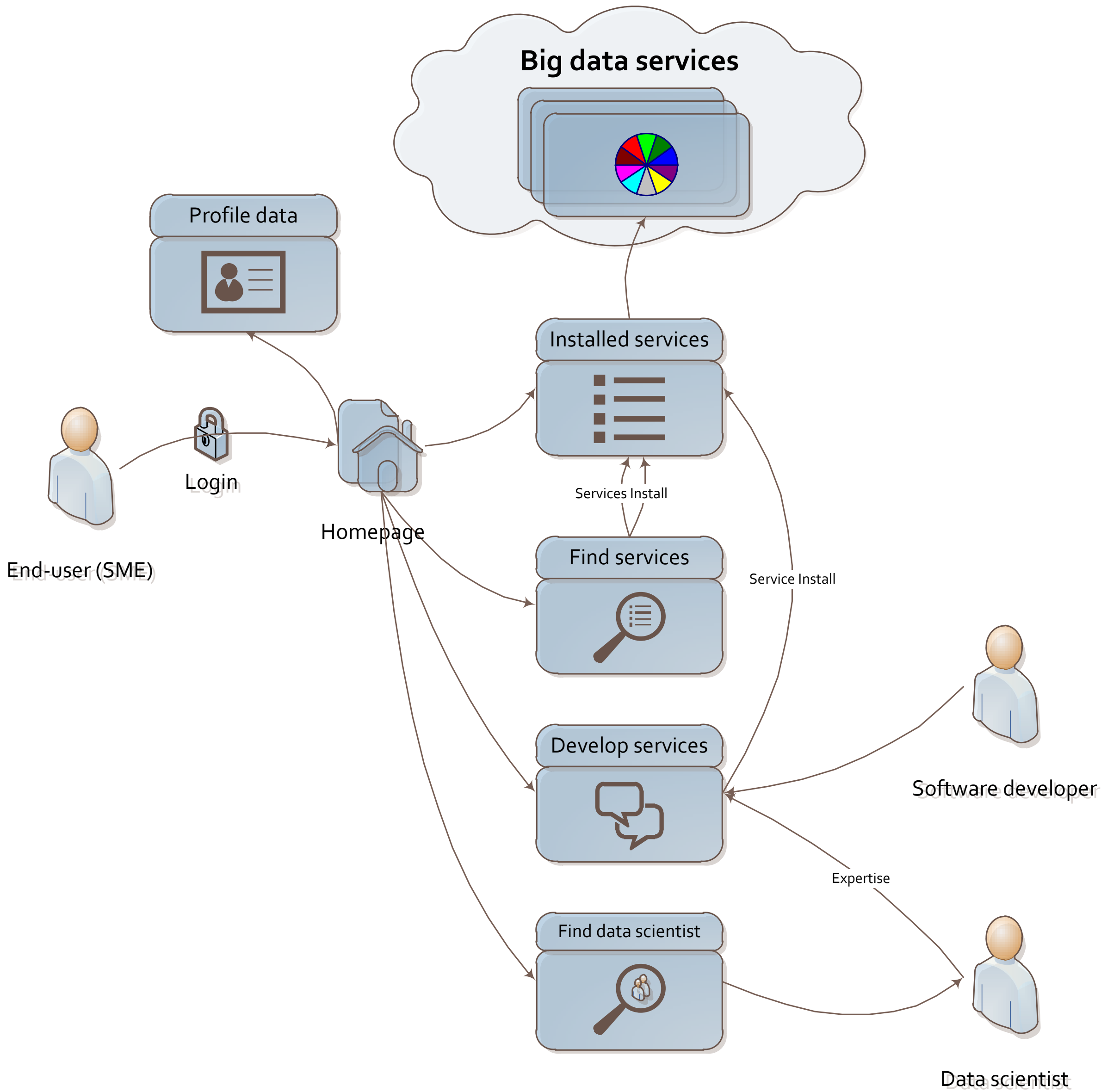

Figure 31: Walk-through of how SME end-users will use the platform

The second customer group of data scientists includes individuals with competencies in different data analyses techniques. We can segment this group based on the employment status of its representatives. We can distinguish between students (PhDs), freelancers, and employed professionals. Each of these segments can be targeted with a different value proposition.

Next, we shortly discuss the issue of platform openness.

6.1.4. Platform Openness

The platform is open for users on the demand side (i.e. SMEs) as this customer group provides the main revenue source. The platform is also open for users from the supply side (i.e. data scientists) of the market however certain screening procedures will be set in place. The platform provider will screen data scientists in order to ensure that they have some minimal level of experience in data analysis. This model is referred to as the walled garden model in the literature [4]. Finally, the platform will be closed for other users who can take the roles of platform provider or platform sponsor.

The following section describes the value elements and value proposition for the two customer groups.

6.1.5. Value Proposition

The perceived value of the platform depends on how its target users perceive core elements of the value proposition [4]. Different value elements can be distinguished for the two customer groups. For SME end-users the core value elements are:

- *Ease of use.* Available big data services must be easy to install, configure and use. Complicated interfaces and functionalities should be avoided as the targeted SMEs do not employ data analytics experts.
- *Speed.* The services will analyze large amounts of data and must show the results in a rapid manner. Long processing times might impair the perceived value of system users.
- *Trust.* The platform provider has to gain the trust of SME end-users if they are to upload sensitive company data for analytics purposes. The SME owners that were interviewed in the second round interviews indicated that they need to be assured that their data is secure and will not be shared with third parties (i.e. competitors).
- *Visibility of effects of use.* During the second round interviews with SMEs, one company owner indicated that he would like to see the effects of use of a particular service on his company performance.
- *Number of complementary services.* From platform theory we know that the overall value of a platform for demand side users is dependent on the number of its complementary services. Thus, the more big data services that a particular company can use, the bigger the value that the platform brings them.

For the customer group of data scientists several value elements are also distinguished:

- *Personalization.* During the second round interviews with data scientists, personalization was identified as an important value element. The list of available projects (from SMEs) that a data scientist can see needs to be based on his/her domain of expertise and interests.

- *Size of SME customer base.* Again from platform theory we know that the overall value of a platform for supply side users is dependent on the size of its customer base on the demand side.
- *Completeness of datasets*. Interviewed data scientists also indicated that the datasets that the platform provides to them should be full or with as less missing points as possible.

All value elements will be evaluated with end users from both customer groups once the prototype of the platform is created. We shall now describe the intended value proposition for SMEs and data scientist. The big data SaaS platform is intended to provide value to SMEs by:

- Providing generic big data services to them in a cost-effective way: non-customizable big data services can be provided to SMEs for free or at a per-usage pricing model as no tailoring is required. The Twitter sentiment analytics service is an example of such generic service application.
- Connecting SMEs to a pool of data scientists: SMEs facing particular data problems or opportunities that lack internal data analytics competencies can find and collaborate with data scientists on the platform.
- Providing a collaboration environment where new or customized big data services can be developed jointly by SMEs, data scientists and the platform provider. The know-how of Dialogues Technology in bridging the gap between business and IT shall be leveraged during new service development. The Dialogues Scrum methodology can facilitate this collaboration and provide a common language between the three stakeholder groups. An example service where such collaboration can occur is the predictive weather analytics service.

The two pilot services have the following intended value propositions for the selected SME segments:

- *Twitter sentiment analytics service (target segment: SMEs from the retail industry)*
  The service allows its users to compare sentiment results for several Twitter search queries. In this way sentiment about competitive brands or products can be displayed in a human friendly way on a single screen by using multiple pie charts. Service users can see how their brand or particular products are positioned in relation to their competitors or substitutes. Alternatively, system users from the retail industry can make choices on which brands to stock and sell in their shops depending on Twitter users' sentiments.
- *Weather analytics service (target segments: SMEs from retail and food industries)*
  The service allows its users to upload company performance data (e.g. product sales) and plot them together with weather data (e.g. weather conditions, air temperature, etc.) on the same line chart. In this way trends can be spotted by service users.
  In addition, a predictive analytics component can be added to the service. For this to happen, SMEs should commission to a particular data scientist the development of a predictive data model based on the available datasets. When the model is created, the platform provider can develop the predictive analytics service which will use weather forecasts as input.

At the other platform side, data scientists can be targeted with the following intended value proposition:

- *Access to open/paid datasets (target segments: students, employed professionals)*
  During the second round interviews with representatives from the data scientists group, the respondents identified that they would be interested in downloading datasets related to research projects they are involved in. Export services on the platform can structure longitudinal or cross-sectional data to particular formats (Excel and CSV) that can be easily imported into statistics software packages.
- *Revenue source (target segments: students, freelancers)*
  Data scientists can receive revenue from SMEs through the platform for their participation in the development of customized new services.

Moving on to the innovativeness of the designed system, the researcher was not able to find any existing platform in the big data domain that creates a similar business ecosystem. Thus, we argue that the designed platform is highly innovative when it comes to facilitating the cooperation between SMEs and data scientists in the big data domain.

The innovativeness of the two pilot services should also be discussed. First, the Twitter sentiment service is not very innovative as other companies provide similar services (e.g. Topsy). However, existing service providers are focused only on one source of data (i.e. Twitter) while the platform designed in this project deals with many data types and sources. Thus, given that the price and quality of this service are similar to the price and quality of existing ones, the designed platform can bring more value to its customers by providing a more diverse array of complementary services to them. Second, the weather analytics service is regarded as a highly innovative by the researcher as no similar services have been found.

6.1.6. Pricing

Pricing is another variable that, in the case of multi-sided platforms, should be viewed from different angles. As there are different value propositions for SMEs and data scientists, different pricing models can be set for them. We will first discuss pricing models for the customer group of SMEs.

We argue that several pricing models can be combined for SME customers:

- *Monthly subscription fee*
  First, SMEs will be charged with a small monthly subscription fee for covering customer support costs or any administrative costs that the platform provider incurs.
- *Monthly per-usage service fee*
  Second, for big data services, a per-usage pricing model can be applied. Depending on the type of service, SMEs can be charged for the amount of data used for analysis (e.g. number of tweets), number of analysis runs, or system resources used.
- *Per-hour fee for new service development/service customization*
  Third, for the development of new or customized services, SMEs will pay data scientists and the service developer (i.e. the platform provider) on a per-hour basis.

In our mass market strategy, we have chosen to attract SME customers by providing them with generic big data services. However, from the second round interviews, we found out that SMEs are not willing to incur initial costs as they cannot be certain that the platform actually adds value for them. Consequently, the SMEs from the mass market should be subsidized by the platform provider. First, a free 6 month subscription period will be set for new customers. Second, generic big data services will be provided for free below a certain usage limit. For example, for the Twitter service, a limit can be set for the number of search queries per service run or the number of downloaded tweets per month. In this way, customers can try out the services and get a perception on the different elements of the value proposition.

We identified the following pricing models for data scientists on the other side of the platform:

- *Commission fees*
  A small commission fee will be collected from the revenues of data scientists on the platform.
- *One time dataset download fee*
  Cross-sectional datasets will be available for data scientists to download for a fee.
- *Yearly license fee for longitudinal datasets*
  In addition, data scientists will be able to pay a yearly license fee for longitudinal datasets. That is, after the longitudinal license fee is paid, data scientists will be able to download the updated dataset within the license period.

Initially, for attracting data scientists to the platform, the commission fee can be waived. That is, the platform provider will subsidize data scientists and will allow them to retain all profits received through the platform. When a large enough customer base is created, the platform provider can start collecting commission fees.

6.1.7. Branding

The new big data SaaS platform can be branded by associating it with Dialogues Technology and their innovative Dialogues Scrum methodology. By applying the Dialogues Scrum in new service development projects, the gap between business and IT can be reduced. That is, companies and software developers communicate on regular basis which ensures that the end product reflects the needs and wants of the business.

For the case of the big data SaaS platform, besides business (i.e. SMEs) and IT (i.e. the platform provider), one additional stakeholder group is added: data scientists. Data scientists need to understand in detail the context of SMEs and the problems or opportunities that they are facing in order provide them with relevant expertise and data models. Thus, the new big data SaaS platform can be branded by associating it with Dialogues Scrum since the methodology has been successful in overcoming the differences in thinking of business and IT people.

6.1.8. Customer Retention

When it comes to customer retention, several strategies can be applied. First, a large catalogue of complements (i.e. big data services) can be created. In this way representatives of the SME customer group can identify and make use of several big data services that add value for them in different ways. In addition, the platform should support a large variety of data sources that can be included in new big data services.

Second, the platform will offer the possibilities for service customization and new service development to SME customers. In addition, these development projects should be managed effectively by the platform provider. That is, communication between the three stakeholder groups is crucial and has to be managed effectively in order to arrive with customized solutions that indeed add value to SME customers. Regular communication can create trust between the involved stakeholders which can serve as a retention mechanism.

6.2. Technology Domain

This section describes the Technology business model domain. The second domain includes the variety of technologies (e.g. network, devices, and applications) that enable the successful delivery of the intended value proposition to the platform's target customers [4]. The technical architecture of the platform was presented in Figure 22 in the previous chapter.

The designed big data SaaS platform can be classified as both an analytics and a collaboration information system. This classification stems from the main functions that the platform is designed to support. First, the system provides customers from the SME group with hardware and software for drawing insights from big data. More precisely, end-users from the SME group can subscribe to and use different analytics services running on the platform's IT infrastructure through a web interface. Second, the platform provides SMEs end-users and data scientists with a collaboration environment for the development of new or customized big data services. Thus, communication and interaction between these the two groups have to be supported.

Different technologies have to be leveraged in order to realize the two main functions of the big data SaaS platform and the value proposition elements. These technologies are described in turn in the following sections. First, we start by describing the platform's backbone infrastructure.

6.2.1. Backbone Infrastructure

The network and hardware infrastructure combined together form the backbone infrastructure of the big data SaaS platform. On one hand, system users will access the platform through a web interface. They can do so from any geographical location with existing Internet connectivity through a fixed line or a wireless/mobile connection.

On the other hand, the hardware infrastructure of the big data SaaS platform has to be easily scalable as big data analytics is associated with analyses of large volumes of data. Speed of the big data services has been identified as one of the core value elements thus the hardware infrastructure needs to allow for rapid scalability of storage and processing power.

6.2.2. Open Source Infrastructure Technologies

Several open source technologies need to be implemented in the platform in order to realize the intended value proposition.

First, the Hadoop open source framework will be leveraged for parallel processing of data across multiple clusters of computers [90]. The framework has been specifically designed to run on commodity hardware and it includes the Hadoop Distributed Files System (HDFS) and Hadoop MapReduce modules. The former is a files system that allows for high-throughput access to application data, while the latter is a system for distributed processing of large datasets. By implementing Hadoop, the speed value element is addressed as the framework can reduce significantly the time needed for analysis of large datasets.

Second, Apache Tomcat will be used as a web server for the platform application. Tomcat was selected since it supports Java which in turn is the programming language chosen by Dialogues Technology for the development of the platform.

Finally, mongoDB has been chosen as a schemaless database by Dialogues Technology because of the combination of flexibility, power, speed and ease of scaling. High speed and ease of scaling are especially important for the purposes of big data analytics as already noted above.

6.2.3. Service Platforms

Several components of the platform's technical architecture (Figure 22) do not have to be developed from scratch as they can be brought in from external service providers. First, data storage and processing power can be outsourced to infrastructure services like Amazon AWS or Microsoft Azure. IaaS providers allow their customers to scale their operations in an on-demand fashion, usually through an API. As already noted, this is an important requirement from the technical architecture that has direct impact on the speed value element.

Second, platform services will collect data from various data providers' APIs. As already discussed in the previous chapter, the most common API data formats are JSON and XML thus the platform shall support both of them. For the realization of the two pilot services, data from Twitter and KNMI needs to be collected by the platform.

Third, on different occasions, external big data services from third-party service providers can be integrated in the platform. For instance, the SNTMNT's brands sentiment API needs to be integrated in the Twitter analytics pilot service.

6.2.4. Devices and Interfaces

The platform has to be available for different client devices: desktop PCs, laptops, smart phones and tablets. These are two major requirements for these client devices: first, they must have means for Internet connectivity and second, the must have an installed web browser.

The 'intelligence' of the SaaS platform is located in its network infrastructure thus there are no specific hardware requirements set for the client devices. Moreover, synchronization of data between devices is not required as data are stored on the cloud. Perhaps one exception must be noted here. The platform is not going to provide a build-in environment for statistical analyses thus data scientists must perform these activities on their own devices (i.e. desktop PC or laptops). This requires that they already have installed software packages for statistical analysis like SPSS.

From the second round interviews, we were able to learn that system users would use different types of devices for different purposes. Mobile devices (smart phones and tablets) are especially suitable for communication purposes (for both customer groups) or for viewing results from real-time data analytics services (for SME end-users). On the other hand, desktop PCs and laptops are required when end-users from the SME category have to upload data to the platform or when data scientists have to perform statistical analyses.

6.2.5. Security

On certain occasions, SME end-users have to upload company performance data. For instance, for the weather analytics pilot service, company employees can upload sales figures per product or total turnover per day. During the second round interviews, we have learnt that companies are very careful when sharing such data as they do not want them accessed by their competitors. Thus, SME end-users need to trust both the platform provider and the implemented security mechanisms.

First, a security threat exists when end-users from the SME category transmit private company data to the platform. Thus, during data transmission, an encrypted connection should be created between the sender and the receiver through the SSL protocol. In this way, if a third party manages to retrieve the transmitted data, they will not be able to read it without breaking the encryption first.

Second, the platform will store the transmitted data for analytics purposes thus there is also a resource-related threat. Consequently, the platform provider must know where sensitive data are physically stored and make sure that the location is secure from both physical and network attacks. As data storage will be outsourced, the platform provider has to transfer the security requirements to the data storage provider and make sure that they are fulfilled.

Finally, user authentication needs to be implemented so that unauthorized parties cannot access the platform's services or the information shared by system users through their user profiles. User passwords must be encrypted before being stored in the system user database.

6.2.6. Management of User Profiles

System users from both customer groups will create accounts in the platform through a registration interface. For the case of an SME customer, the user account will be owned by the company, while employees will manage the profile information. On the other hand, data scientists will both own and manage their own user profiles.

End-users from the SME group will enter general company information in the system: company description, industry, size, contact information. This information will be shared with all data scientists. SME end-users will also enter more detailed company information: business model description, current data problems or opportunities. This profile information will be shared only with selected data scientists.

On the other side of the platform, data scientists will enter professional information in their user profiles: domain expertise, used software packages (e.g. SPSS), current availability, and contact information. End-users from the SME group will be able to see this information when they enter the profile page of a particular data scientist.

Unauthorized system users will not be allowed to access customer profiles on the system. Only after system users have authenticated themselves, they will be able to view other users' profiles.

6.2.7. Billing

Every month, SMEs will be presented with a bill that contains the fee for each of the services used by their account. Consequently, the system needs to track relevant usage parameters for each big data service. These include used hardware resources and amount of analyzed data.

Billing will be handled by the platform provider. Payment of the bills however will be outsourced to a third party service provider. In The Netherlands, iDeal is the most commonly used and trusted service provider for online payments. Thus, the platform will be integrated with iDeal.

6.3. Organization Domain

In this section, we describe the Organization business model domain. We start with a description of the roles, assets and actors that are required for the delivery of the value proposition.

6.3.1. Roles, Assets and Actors

Several roles are needed for the realization of the intended value proposition. The different roles are presented in Table 12. We have also identified the assets required from each role and the potential actors which can provide these assets.

Role	Assets (resources and capabilities)	Potential actors
Access network operators	Telecommunications and data communication networks	Internet service providers and mobile operators in The Netherlands
Infrastructure provider	Data storage and processing power	Amazon, Google, Microsoft
Open software providers	Software libraries and databases that enable big data analytics	Apache, 10gen Inc.
Big data platform provider	Agile software development, Dialogues Scrum know-how	Dialogues Technology
Customer support provider	Knowledge in the inner workings of the platform and its services	Dialogues Technology
Third party (big data) service providers	Specialized in particular type of big data analytics. Owners of specific (secret) algorithms.	SNTMNT
Data/content providers	Raw data or different types of content needed for data analyses	Twitter, KNMI
Payment providers	Secure environment for the consumed service	iDeal, PayPal
Investors	Financial resources	Angel investors, investment funds

Table 12: Roles needed for realizing the value proposition

We argue that several of the assets from Table 12 can be regarded as critical for the provision of the different value elements. The first critical resource is the hardware infrastructure: data storage and processing power. The provider of these resources must allow for rapid scaling (up/down) of resources. Second, the software libraries and databases are considered as critical as well. In order to enable big data analytics, these software packages must allow for distributed processing of large datasets. Finally, the know-how of Dialogues Technology in facilitating the dialogues between business and IT and their agile software development methodology are considered as critical capabilities of the platform provider.

6.3.2. Partner Selection

This section provides the selected partners for each of the roles that we have identified above (see Table 13). Actors are operating in The Netherlands as the platform will be launched in this country.

Role	Selected Actors and Selection Rationale	Actor type
Access network operators	Ziggo, XS4ALL (ISPs); KPN, T-Mobile, Vodafone (Mobile operators) The services of the big data platform will reach their intended customers through the communication networks of these providers. As no specific requirements are set for these networks, it is not necessary to select particular partners within this role.	Supporting
Infrastructure provider	IN2IP This company has been chosen by Dialogues Technology since a trust relationship had already been formed between the two actors in the past. IN2IP also provides managed security services that are designed specifically for secure handling of sensitive company or personal information.	Structural
Open software providers	Apache and 10gen Inc. Apache and 10gen Inc. are the providers of the Hadoop framework and the mongoDB database respectively. These two open source technologies have been selected by the main developer because of the presence of extensive documentation and developer community around them. In addition, both technologies are free for commercial use under the Apache License v2.0 and GNU AGPL v3.0 respectively.	Contributing
Third party (big data) service providers	SNTMNT SNTMNT has been chosen by Dialogues Technology as a partner for the development of the Twitter sentiment analysis service. The company specializes in financial sentiment analysis of tweets however they also provide a brand sentiment service. According to SNTMNT, their algorithm outperforms by 10% the best generic models out there. The algorithm itself makes use of semantic models, support vector machines and natural language processing classifiers.	Structural
Data/content providers	Twitter, KNMI Content from Twitter and KNMI needs to be gathered for the realization of the two pilot services. The KNMI institute has been chosen by the main developer as their weather data are being used by the shipping and aviation industries in The Netherlands. In addition, they provide open access to their weather data.	Structural
Payment providers	iDeal iDeal is a third-party payment service provider in The Netherlands. The researcher has chosen to integrate the platform with iDeal because of the high level of trust that the provider has in the Dutch market.	Contributing
Investors	No investors have been found at this stage.	Contributing

Table 13: Selected partners

In addition, we have also identified the type of the selected partner(s) for each role. Three actor types are distinguished: structural actors provide essential and non-substitutable assets; contributing actors provide specific services but are not part of decision making; and supporting actors provide generic services that can be obtained from different suppliers [4].

The selected actors form the value network that is depicted in Figure 32.

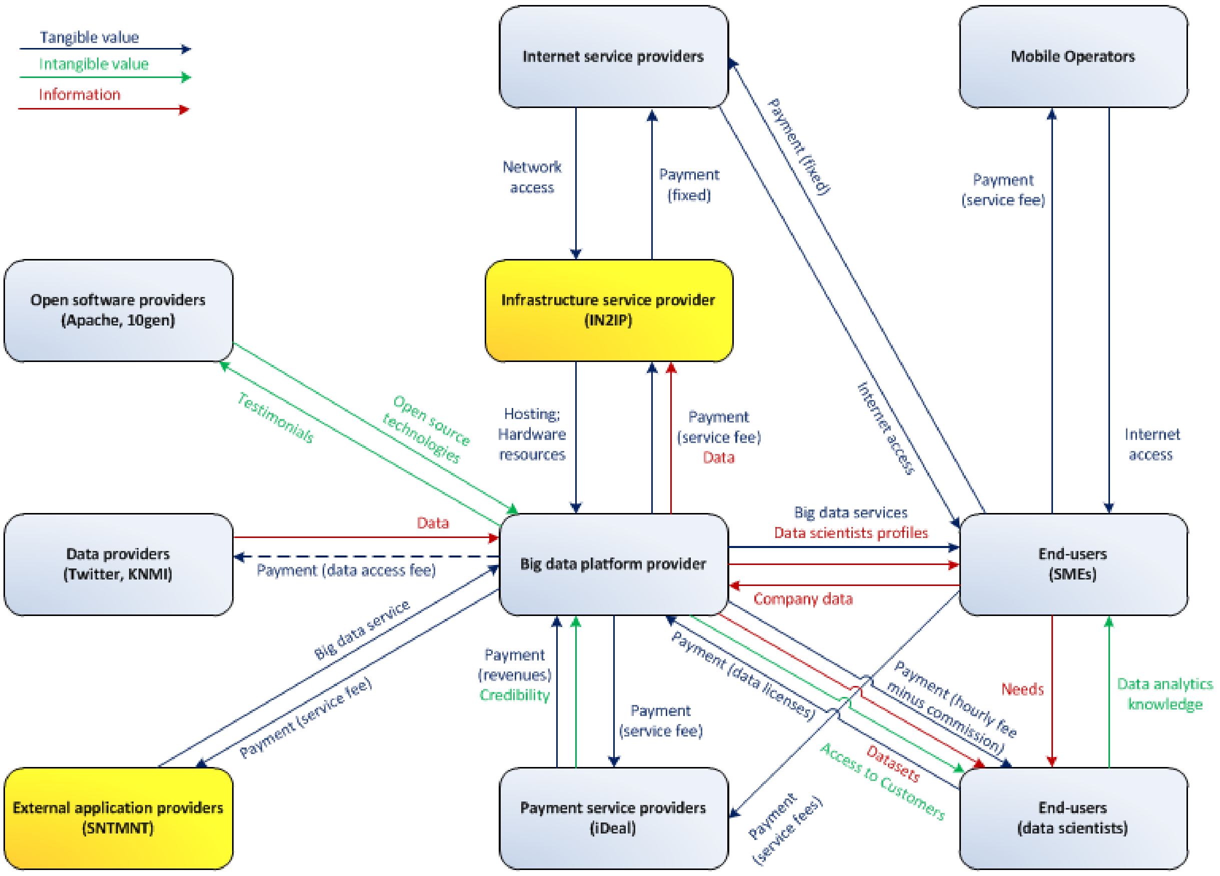

Figure 32: The big data platform's value network

Yellow blocks in the value network represent partners of the platform provider, while blue blocks represent other actors in the value network. Blue arrows represent tangible value flow, while green arrows represent intangible value flows and red arrows – information flows.

6.3.3. Platform Complementary Services

Initially, all platform complementary services will be developed in-house by the platform provider. The reason is that the technical infrastructure and these services will be developed simultaneously. The decision whether to open up the underlying technology by creating software development kits and APIs for other service developers will be taken at a later point when the platform's technical architecture is reliable and the market dynamics have been started. The decision of opening up the platform to a third market side (i.e. service developers) should be addressed periodically by the platform provider. Choosing the right moment for this could increase rapidly the customer bases on the other two sides of the SaaS platform.

6.3.4. Business Ecosystem Strategy and Governance

The platform provider, or Dialogues Technology, should take the role of a keystone as it provides the core technology blocks upon which different services can be created jointly with SMEs and data scientists. That is, by regulating the connections and stimulating dialogues between different ecosystem members, the platform provider allows for the creation of new services which add value to all members. On the other hand, a dominator strategy is not feasible as dominator try to take over the entire ecosystem by eliminating as many other members as they can [74]. The dominator strategy, therefore, is in direct conflict with the strategy selected for starting the two-sided market dynamics.

Initially, decisions on the technologies underlying the SaaS platform will be taken solely by the platform provider. Naturally, input from both SMEs and data scientists for each service will be taken into account. At a later stage, if the platform is opened for external service developers, their requests for new API features will be considered by the provider.

6.4. Finance Domain

The Finance domain is the final domain of the STOF framework. Costs, revenues, risks, investments and pricing are addressed in turn in this chapter as part of the Finance domain. The business model is designed from the platform provider's perspective thus the following sections are also described from this perspective.

6.4.1. Cost Sources and Costs

Costs are divided in two groups: fixed costs and variable costs. For the platform provider fixed costs come from three sources. The first source is labor costs for the development of the platform and its big data services. Within Dialogues Technology labor costs are calculated by using the FTE (Full-time equivalent) unit. 1.0 FTE is equivalent to a full workload week of 40 hours of one person. One 1.0 FTE software developer and one 1.0 FTE graphical designer will be involved for the development of the platform and the two pilot services. The second source of costs is customer support. Initially, customer support will be provided by the development team, but as the platform grows, a technical employee will be assigned with the support function. The third source of fixed costs is overhead, or the operating expenses per employee for office space and utilities.

There are also several sources of variable costs for the provision of the platform and its services. First, infrastructure costs for data storage and processing power will be incurred. The platform provider will not invest in building their own IT infrastructure and will use the services provided by IN2IP. In this way fixed costs for IT infrastructure will be transferred to variable costs. Monthly fees will be transferred to the IaaS provider based on the amount of consumed resources. A second source of variable costs is the usage of external big data services. For the Twitter pilot service, service fees must be paid to SNTMNT for their sentiment analysis service. The fees will be paid monthly on a per-usage model. Third, transaction costs are yet another source of variable costs. The payment services of iDeal are a source of transaction costs. We have decided to take this source of costs as iDeal brings credibility and trust to the platform. Finally, marketing costs are identified as another source of variable costs.

6.4.2. Revenues Sources and Revenue

The revenues sources of the big data platform come from its two customers groups. SMEs will be charged each month on a per-usage basis for the big data services that they have used. For the development of new or customized big data services, SMEs will pay on a per-hour basis. Finally, a small monthly fee for covering administrative and customer support costs will also be collected. The platform provider will start collecting this fee once the platform generates enough demand for this auxiliary service.

On the other side, representatives of the customer group of data scientists will be charged with a small commission on the revenues that they have earned through the platform. Data scientists will also be charged for datasets that are not available from other (open) data sources. For instance, aggregated datasets on company performance data can be sold to data scientists. Thus, another revenue source is licenses for cross-sectional and longitudinal datasets.

6.4.3. Risk Sources and Risks

We can identify several risks that threaten the viability of the business model.

First, a risk exists of the platform not reaching the mass market. The first customers from the SME group of the platform can be categorized as early adopters who are ready to adopt new technologies in order to gain competitive advantage in their industry. Early adopters however require customized solutions while adopters from the mass market require proven and reliable services. Thus, there is a risk that the platform provider will not be able to cross the 'chasm' between the early and majority markets.

Second, there is a risk of the emergence of competitive platforms in the market as the platform's IP cannot be protected be legal means. However, the unique competence of the platform provider is the know-how in agile product/service development and bridging the gap between IT and business. This competence could lead to a competitive advantage over new entrants. Also, increasing the installed base of the platform more quickly than competitors is crucial for introducing network effects and attracting more customers to both platform sides.

6.4.4. Investment Sources and Capital

IT infrastructure will be outsourced in order to reduce the need for initial investment capital for hardware. In this way, financial risks for the platform provider will be reduced as well. In addition, the platform provider will be able to remain lean and focus on its core competencies: agile software development and facilitating the collaboration between representatives of the different platform sides.

Investment capital is required mainly for software development. First, financial capital needs to be secured for the development of the system's prototype. After the prototype has been built its value proposition will be evaluated with potential end-users from both customer groups. Further investment will then be required for the development of the platform and its pilot big data services. Potential sources of investment are company capital from Dialogues Technology and capital from angel investors or investment funds.

6.4.5. Pricing

SMEs are regarded as a very price-sensitive target group. Therefore, a penetration pricing strategy is preferable to a market skimming strategy. The former is focused on increasing customer base (and market share) by keeping profit margins low, while the latter is focused on profit maximization by asking for a premium price.

The fee structure for SMEs is based on a combination between a fixed fee and variable fees. The fixed fee will be collected on a monthly basis for covering transaction costs (e.g. for iDeal payments) and the provision of auxiliary services like customer support. Initially, the fixed fee will be waved until these costs become significant and the platform provider cannot sponsor its customers anymore.

Monthly variable service fees for SMEs will be based on a per-usage pricing model (see *Service Domain*). We regard the cost-based pricing method as best suited for two main reasons. First, there is a lack of significant fixed costs for the provision of big data services to SMEs. Variable costs for consumed IT infrastructure resources can be measured very precisely by the IaaS provider. Second, any additional costs associated with data collection or external big data analyses can also be tracked precisely by the external (SaaS) providers (e.g. SNTMNT). Thus, a variable monthly price can be calculated based on the incurred service provision costs and a small margin for the platform provider.

However, at this point in time no concrete pricing agreements have been made with the selected infrastructure service provider (i.e. IN2IP) and external sentiment analysis service provider (SNTMNT). Thus, costs for service provisioning cannot be estimated yet. Additionally, potential clients of each particular big data service must be prompted for their willingness to pay (e.g. through conjoint analysis). Once this is done, a final decision on the pricing method for each service can be taken: stay with the cost-based method, or shift to another method (e.g. value-based method).

Furthermore, SMEs will be charged for customization or new service development on a per-hour basis. That is, the platform provider will be compensated for the hours used for software development, while data scientists will be compensated for hours spent for consultation and data analysis. During the second round interviews, both data scientists made it explicit that they prefer to be compensated on a per-hour basis as SMEs might be having difficulties when explaining their exact needs and goals.

6.5. Conclusion

This chapter had the purpose of answering the fourth research question: *How does the business model of the designed big data SaaS platform look like?* In order to answer the above research question we performed the quick scan stage of the STOF method. All four domains of the STOF model were outlined in turn. Insights from the two interview rounds were used as input when describing the first two domains: Service and Technology.

We have tried to describe the business model in as much detail as possible at this stage of the project. We paid specific attention to the additional CDIs that were identified in the Theoretical Background chapter. Table 14 below provides a summary of the 8 CDIs applied to this particular case.

Critical Design Issue	Application to the case of SaaS platform
Selection of interfaces (Technology domain)	Mobile devices (smart phones and tablets) were selected for communication between end users of the two customer groups. While, desktops and laptops are required when end-users from the SME category have to upload data to the platform or when data scientists have to perform statistical analyses on their own devices.
Selection of service platforms (Technology domain)	Selected service platforms: - Infrastructure service provider: IN2IP - External big data service providers: SNTMNT - Data providers: Twitter, KNMI - Payment provider: iDeal
Two-sided market dynamics (Service domain)	The customer base of SMEs was chosen to be built up first. The main reason for this choice is that SMEs form the main revenue source for the platform. Data scientists will then be attracted once there is enough demand for their expertise and competencies

Critical Design Issue	Application to the case of SaaS platform
Platform openness (Service domain)	Open for SMEs; Walled garden for data scientists;
Platform complementary services (Organization domain)	All (complementary) big data services will be developed by the platform provider initially. The reason is that the platform's technical capabilities (e.g. supported data sources) will be developed together with its first services.
Pricing (Service domain; Finance domain)	Pricing models for SMEs: - Monthly subscription fee to cover the costs for auxiliary services - Monthly per-usage service fee - Per-hour fee for new service development SMEs will be subsidized by the platform provider initially. The flat monthly subscription fee will be waived in the first 6 months. Additionally, big data services will be provided for free below a predetermined usage limit. Pricing models for data scientists: - Commission fee on earned revenue from the platform - One time dataset download fee - Yearly license fee for longitudinal datasets Initially, the platform provider will subsidize data scientists and will allow them to retain all profits received through the platform. After some period (not determined yet) the platform provider will start collecting these fees.
Business ecosystem strategy (Organization domain)	The platform provider has to take the role of a keystone player in its business ecosystem. That is, the platform provider has to allow for the creation of new services which add value to all members by regulating the connections and stimulating dialogues between different ecosystem members.
Business ecosystem governance (Organization domain)	Initially, decisions on the technologies underlying the SaaS platform will be taken solely by the platform provider. At a later stage, if the platform is opened for external service developers, their requests for new API features will be considered by the provider.

Table 14: Application of the eight additional CDIs to the big data SaaS platform case

In the following chapter we focus on evaluating the first outline both internally with the project team and externally with experts and practitioners. By doing so, we aim at identifying all areas of the business model that need to be refined further. The following chapter describes the activities performed as part of the final evaluation phase of this project.

7
Business Model Evaluation

This chapter's position within the research framework can be seen in Figure 33. Its purpose is to answer the fifth research question:

> ***To what degree the designed business model can be considered as complete, consistent, viable, scalable, and sustainable?***

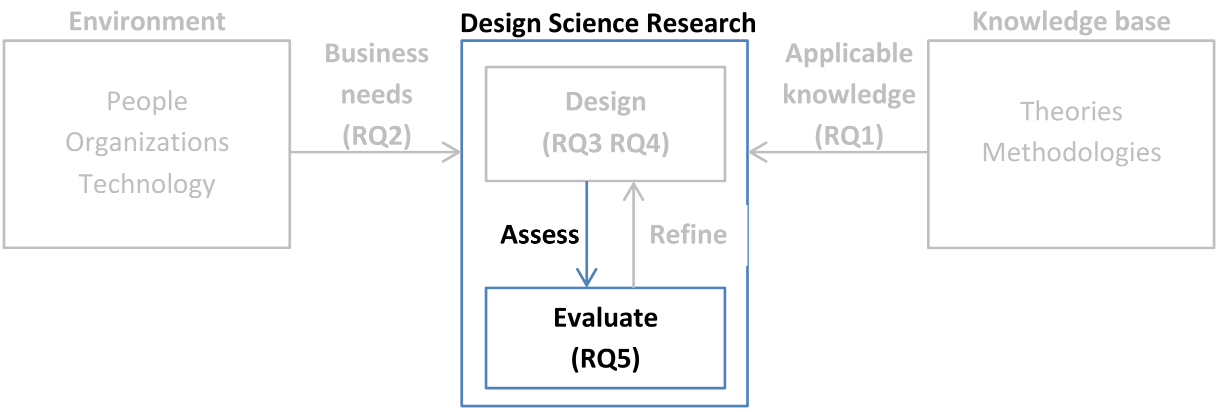

Figure 33: Positioning of the "Business Model Evaluation" chapter within the research framework

This is the final phase of the thesis as it closes the design cycle by evaluating the business model behind the designed big data SaaS platform. The business model evaluation is based on several criteria and their assessment is part of the final step of the STOF method, i.e. the robustness check [4].

In this chapter we first describe in detail the chosen evaluation methodology. This includes descriptions of the chosen set of evaluation criteria (i.e. completeness, consistency, viability, scalability and sustainability) and the selected respondents with their competencies. Second, we describe the evaluation results per criterion. Finally, we draw conclusions based on the performed evaluation.

7.1. Evaluation Methodology

The evaluation of the business model was conducted in two stages. In the first stage, the business model outline was assessed internally by the project team (the researcher, one software developer and one of the Dialogues Technology's directors). In this session, the outline of the business model was first introduced to the team by the researcher. After that, all evaluation criteria were discussed in turn. This internal assessment also served as a pilot evaluation session before the second evaluation stage.

In the second stage, we aimed at validating/evaluating the business model with people *outside the project team* in order to get their external perspective. Different evaluation criteria were discussed with each respondent depending on their area of expertise.

Next, we describe all business model evaluation criteria. After that, we give our argumentation for the selection of external respondents who assessed the outlined business model.

7.1.1. Criteria for Evaluation

The evaluation of the business model is based on the following criteria: completeness, consistency, viability, scalability, and sustainability.

Completeness

For evaluation purposes in this project, completeness is defined as the degree to which the business model was designed in sufficient detail. The completeness of each domain was assessed separately by discussing the following questions:

1) Are there important issues that are not addressed in the domain?
2) Are there issues that need to be elaborated further?

Consistency

The consistency criterion is included in the business model evaluation in order to check whether the relationships between the different business model domains are well balanced. Each domain can pose requirements to the other three thus a balance needs to be achieved. The following questions are addressed during evaluation of the business model consistency:

1) Are there conflicts between the four domains that necessitate changes in the business model?
2) Can the technology design deliver the intended value proposition?
3) Can the involved organizations deliver the intended value proposition?
4) Is there a balance between the costs of delivering certain value elements and the created customer value?
5) Have all technological functions been properly assigned among the involved organizations?

The questions above are provided in the STOF handbook [4].

Viability

The assessment of the business model's viability is based on evaluation of the CSFs within STOF. The CSFs indicate the degree to which the business model is capable of creating both customer and network value [4]. A viable business model is one that creates both types of value.

STOF provides a list of questions that can be addressed during the evaluation of CSFs. We have modified this list by taking into account the platform-related CDIs that were identified in the Theoretical Background chapter. New questions were added and several of the existing questions were modified. The additions and modifications are presented in Table 15: newly added questions are highlighted in green, while modified questions are highlighted in yellow.

Critical success factor	Questions
Compelling value proposition	• Is the value proposition sufficiently attractive to customer groups from each platform side? • Do additional complementary services make the value proposition more attractive to demand side customers? • Does a large customer base on the demand side make the value proposition more attractive to supply side customers?
Clearly defined target group	• Are the target groups for each platform side sufficiently clear? • Do we know enough about each customer group in terms of their needs, preferences, capabilities, available resources, and sensitivities? • Is the strategy on how to start the market dynamics between the different sides of the platform clear?
Unobtrusive customer retention	• Is subsidizing going to be used as a retention mechanism? Is it clear which customer groups should be subsidized and how?
Acceptable profitability	• Is it clear who will incur the customer subsidy costs if such exist?
Sustainable network strategy	• Is there a common goal to which all business ecosystem members can adhere and strive to?
Acceptable division of roles	• Is there a keystone actor willing and capable to lead the value network?

Table 15: Modified list of CSFs from the STOF method

The evaluation of CSFs in this project was based on the modified set of CSF questions.

Scalability

The scalability of the designed artifact and business model is evaluated by addressing the following two questions (from the STOF handbook [4]):

1) What will happen to the domain designs in case there is a huge demand for the platform's services?
2) Is the scalability of the platform's services aligned with its technical scalability?

Sustainability

Sustainability of the business model is assessed by addressing different aspects of the external environment by asking 'what-if' questions. These aspects include new technologies, changing market conditions, changing consumer and competitor behavior, and changes in relevant regulation [4]. The following questions are addressed (from the STOF handbook [4]):

1) What if new and cheaper technology becomes available? Is the technological architecture capable of absorbing it?
2) What if one of the partners ends their participation?
3) What if alternative services with comparable or better functionalities are brought to the market?
4) What if the market is expanded internationally?
5) How can changes in regulation affect the provision of the services?

Sustainability is the last of the selected evaluation criteria. Next, we give argumentation for our selection of respondents for the second evaluation stage.

7.1.2. Selection of Respondents

In the second evaluation stage we wanted to get an outside perspective on the outlined business model. We based our sample of respondents on two dimensions: area of expertise and type of knowledge. The first dimension consists of the following three areas: the research domain (i.e. data analytics and SaaS), platform theory and business models theory. The second dimension divides the respondents in two groups based on the type of knowledge that they have: experts or practitioners. The former is composed of academics and consultants, while the latter includes people with more practical, hands-on, experience. Table 16 visualizes the two dimensions.

	Research domain (Big data, SaaS)	Platforms	Business models
Experts			
Practitioners			

Table 16: Dimensions for the selection of respondents for business model evaluation

Ideally, an exhaustive evaluation of the business model is one where at least one respondent from each cell is interviewed. We were able to find a practitioner for both platform and business model cells. However, we were not able to find a practitioner from the research domain on time. Thus, four interviews were conducted in total. Table 17 shows the initials of the selected experts/practitioners. We have decided to hide their real names for privacy reasons.

<table>
<tr><th></th><th>Research domain (Business analytics)</th><th>Platforms</th><th>Business models</th></tr>
<tr><td>Experts</td><td>PB</td><td>FN</td><td>TH</td></tr>
<tr><td>Practitioners</td><td>-</td><td colspan="2">TC</td></tr>
</table>

Table 17: Selected respondents for business model evaluation

We have chosen to talk to PB as an expert in the research domain with more than seven years of experience in data management and business intelligence consultancy. We decided to exclude the evaluation questions related to completeness in this interview. The reason is that we assume that the person is unlikely to spot gaps in the different domains as he has no or very little expertise in business models.

FN is a PhD candidate at TU Delft doing research in the field of smart living services. FN is considered as an expert in platform theory and the domain of SaaS and SOA. Business model completeness, consistency and viability were discussed during this interview. Even though FN is not an expert in business models, the completeness criterion was discussed during the interview. The reason is that the respondent could advise whether important platform issues are omitted in the outline. In addition, consistency and viability were also discussed.

TH has done research in business model theory in the past. Currently, TH takes the position of principal advisor on innovation management and business models in a Dutch research institute. Completeness, consistency and viability were discussed in this interview since the researcher regards these three criteria as most important.

TC is a representative of the practitioners group with area of expertise in both platforms and business models. The respondent has practical experience in business development for two platform services (mysnooze.com and Gropometru.ro). We need to mention that he is involved with Dialogues

Technology, but *he has not been part of this specific project in any way* thus far. Therefore, we consider the respondent as an external practitioner. In this interview completeness, consistency, viability and sustainability were assessed. Scalability evaluation was omitted since the respondent lacks the technical knowledge needed to assess this criterion.

Respondent	Completeness	Consistency	Viability	Scalability	Sustainability
PB	-	X	X	X	X
FN	X	X	X	-	-
TH	X	X	X	-	-
TC	X	X	X	-	X

Table 18: Assessed criteria in each interview

Table 18 provides a summary of the evaluation criteria that were discussed during each interview.

7.1.3. Conducting the Interviews

At least two working days before the start of each interview, the business model outline was sent to the respondents. In this way, we aimed at securing more time for discussion during the meetings. Unfortunately, none of the respondents read the whole business model before their scheduled interviews. On two occasions, the respondents have not read anything, while in the other two interviews the respondents were not finished reading. The researcher handled this issue by presenting the missing points of the business model during the meetings before proceeding to the evaluation questions.

Again, all interviews were recorded with audio recording device after asking the respondents for their explicit permission. The evaluation results are described in the following section.

7.2. Evaluation Results

The results of the evaluation are described in this section. Each criterion is discussed in a separate sub-section by presenting the insights drawn from the different discussions. The source of each statement is provided in the text. The possible sources are: Project Team, PB, FN, TH, and TC.

7.4.1. Completeness

During the evaluation of this criterion, we were able to identify several aspects that are missing or should be elaborated further. We discuss these aspects per domain.

In the *service domain*, the big data services value proposition should be described either in terms of fundamental business goals like cost reduction/increasing current revenues (Project Team) or any other measurable goal (TH). In this way, business customers could be attracted more easily because the value proposition is directly linked to goals they can immediately relate to.

Additionally, a clear market strategy on how to kick-start the platform was missed by one of the respondents (TH). In his words, whether the platform will “fly” or not depends very much on the number of parties that can be attracted and sustained on each of its sides (TH). Or, as another respondent put it: “the most important thing for digital platforms is to build an installed base of users at both sides” (TC). Even though the issue of how to do this is addressed in the current outline (i.e. see *6.1.5. Two-sided Market Dynamics*), further elaboration is needed on segmenting the target markets (SMEs) more precisely (TH) and choosing particular niches to start with (TC).

Furthermore, FN indicated that platform openness can be defined in different ways: organization or technology openness. Thus, she suggested that it should be made more clear which is the case in the current business model. Also, FN suggested that the researcher should think about a strategy to prevent the issue of multi-homing. That is, end users might choose to use more than one platform with similar value proposition. On first sight, the issues of multi-homing and customer retention seem quite similar. However, there are some differences as customers can be retained for one particular big data service, but they might subscribe to another big data service on a competitive platform. Thus, we *recognize multi-homing as an additional platform theory CDI that can supplement STOF*.

Moving on, the *technology domain* does not address in sufficient detail what data sources can be integrated with the platform (Project Team). That is, what open or closed data sources can be integrated and how exactly this can be achieved (e.g. through search API, push API, upload, etc.).

Within the *organization domain*, a missing actor was identified. That is, a marketing/brand specialist needs to be brought on board (Project Team).

Finally, several gaps were noticed in the *finance domain* as well. First, several sources of costs are not covered. These include costs for hardware equipment needed by software developers and other employees (TC); marketing costs (Project Team, TC); and costs for acquiring closed data (Project Team). Second, a more systematic approach needs to be adopted for risk management. For instance, one risk that was not addressed and threatens the business model is the risk of using unreliable data sources (Project Team). TC suggested the researcher to conduct a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis. From this type of analysis risks that might not be identified otherwise can surface (TC). Third, it was suggested that generic big data services can be sponsored by advertisement and thus provided to SMEs for free (TH, FN), an option that has not been considered by the researcher up to this point.

The gaps discussed above must be addressed during the next step of the STOF method: refinement with CDIs. However, further refinement with CDIs is not part of this thesis and will be done after its completion.

7.4.2. Consistency

During the evaluation of this criterion no inconsistencies between the four domains were identified. A conflict between the different revenue sources described in the financial domain was noticed by one respondent (TC). He argued that instead of having several revenue sources, only one revenue source should be selected by the platform provider. In his opinion, the selected sources cannot be combined because they are contradictory. He suggested choosing one revenue source, see if it works and then change it if necessary. However, we do not agree that the listed revenue sources contradict each other. For the case of SMEs, the monthly subscription fee will be charged for the provision of auxiliary services like customer support and not for big data analytics services. For the latter, per-usage fee model was selected. Thus, there is no point of conflict. Nevertheless, we see an advantage of having a single revenue source: viability can be determined with higher certainty than when having multiple revenue sources. This in turn can help with securing financial investments or attracting partners. Thus, we shall start with one revenue source and add others later in time.

Furthermore, there are still some unclear areas related to consistency. For instance it is not yet certain if a balance exists between the costs of delivering certain value elements (e.g. speed) and the delivered customer value (Project Team, TH, FN). This can be assessed once a prototype of the system is created.

In summary, no inconsistencies between the four domains of the outlined business model were identified. However, we argue that consistency should be assessed again after the business model is refined further (again, refinement with CDIs is out of scope in the thesis).

7.4.3. Viability

As already discussed, a viable business model is one that creates both customer and network value [4]. The degree to which these two types of value are created in the current outline is assessed based evaluation of the eight CSF within the STOF method.

The evaluation results per CSF are discussed next. Each CSF is marked as *satisfactory*, *unsatisfactory* or *still unclear*, depending on the evaluation results. A CSF is considered as *satisfactory* only in case it was deemed clear and sufficiently elaborated in the business model outline by all respondents.

Compelling value proposition

During the pilot evaluation session with the project team, the participants agreed that this CSF should be evaluated together with potential customers of the platform. For this reason, the researcher scheduled another interview with the owner of baby products shop (in the first interview, he provided us with his user requirements for the platform). We could only schedule one meeting since the project time was nearly over. The main insights are described below.

When the respondent was asked if he finds the value proposition of the platform and its two pilot services compelling he replied negatively. In his words, "*I'm not going to be signing up for a platform I'm going to be signing up for solutions*" and "*The two solutions that you described* [i.e. the two pilot services] *are too thin to convince me to sign on".* He explained that these two services are fine, but a set of about 15 to 20 of such small services should be developed on the platform in order to attract customers. Thus, he recognized that the number of complementary services running on the platform indeed influences his perception of value. Later on, the respondent suggested that we should think about what questions SMEs have that can be solved by the platform's services and then quickly develop prototypes together with potential users. Furthermore, he indicated that he is willing to spend additional time with the project team for the development of such services.

We also discussed this CDI during different experts during the second round interviews. According to one of the respondents (TH), it is "way too hard" for SMEs to retrieve valuable information from data by themselves. If there is a provider that can do that, then that would be of value to SMEs. However, the question is "Are there enough complementary products?" to make it more interesting for an SME to go to the platform instead directly to another provider. Thus, complementary services definitely influence the value proposition (TH). Also, the more demand there is from SMEs, the more interesting it gets for data scientists to be associated with the platform (TH).

Based on the insights described above, we conclude that this CSF is *unsatisfactory* at this stage.

Clearly defined target group

There are two customer groups for the platform: one on each of its sides. Thus, clear target groups for both of them should be defined.

We observed conflicting opinions on the degree to which the demand side target groups are clearly defined. On three occasions they were assessed as sufficiently clearly (PB, TC, FN). In the rest of the discussions they were deemed as insufficiently clear (Project Team, TH). It was argued that the SME target groups can be more precise as retail and food industries are still too big of segments (TH). Furthermore, measuring the market size was also indicated as good to do since it can suggest whether the chosen segments are still too big (TH).

The target groups on the supply side were assessed as unclear in most interviews (Project Team, PB, and TC). It was suggested that they should be targeted by taking into account the type of expertise that SMEs (in retail/food industries) would need. That is, the type of data that the customer wants to look at determines the type of data scientist they can work with (PB). This is in line with the suggestion to segment and target data scientists depending on their domain of expertise (TC).

Finally, in all interviews, the strategy for starting the two-sided dynamics was considered sufficiently clear. However, based on the other insights above, we declare this CSF as *unsatisfactory*. Mainly, we must target data scientists based on their expertise and capabilities.

Unobtrusive customer retention

This CDI was deemed unsatisfactory (TC) and still unclear (TH) by the two respondents with knowledge in business models. According to TC, the described customer retention mechanisms are still vague and can be further elaborated. TH argued that retention is a next step that is done after attracting the customers. At this point it is unclear how it will be realized thus he marked the CSF as still unclear.

In addition, it was suggested that other retention mechanisms can be considered. According to PB, new (and relevant for SMEs) data sources should be added to the platform when available. In this way, possibilities for developing new big data services can emerge for existing customers. This topic was also touched upon by TH, who argued that from time to time, something new (e.g. data sources/services) has to be introduced on the platform. Furthermore, it was deemed clear to all respondents that customer subsidies will be used. It was also clear that SMEs are the customers that should be subsidized.

In summary, we regard this CSF as *unsatisfactory* as well. The main reason is that retention mechanisms were deemed vague in one of the interviews. We regard this as a signal for elaborating the related CDIs in more detail.

Acceptable quality of service

On several occasions this CSF was deemed as still unclear. Two reasons were mentioned. First, the quality of certain value elements (e.g. speed) cannot be assessed yet (Project Team). Second, quality of a given service strongly depends on the quality of the used data sources (TC). In other words, the quality of a given big data service is strongly dependent on the reliability of the used data (Project Team). For these two reasons, we have labeled this CSF as *still unclear*.

Acceptable profitability

Several respondents indicated that it is still unclear whether the revenue sources can sustain the business model (Project Team, PB, and TH). It was suggested that financial calculations need to be done in order to have a better idea if this is the case (PB, TH). The lack of financial calculations is also recognized as a gap in the business model outline by the designer. However, as no concrete pricing agreements have been made with the selected partners yet, we cannot calculate the costs for service delivery.

Furthermore, it was deemed clear that the platform provider should incur the subsidy costs for providing SME end-users with a free (but time-limited) access to big data services. However, it was discussed that it is not clear how the platform provider will do this (FN). That is, where the money will come from (investment, advertising, etc.).

Based on the insights, discussed above, we label this CSF as *still unclear*.

Acceptable risks

As already discussed during the evaluation of the business model's completeness (see *7.4.1. Completeness*), the major risk of having unreliable services due to unreliable data sources is not addressed in the business model (Project Team). In addition, there might be other risks that are not appropriately addressed like security, privacy and reputation risks (TH). In order to identify such risks and their impact on the business model in a systematic way, a SWOT analysis can be performed (TC). Thus, the CSF is *unsatisfactory*.

Sustainable network strategy

The interests and strategic objectives of the partners included in the business model are only assumed to be aligned at this stage. Dialogues Technology has current working relationships with IN2IP and a trust relationship with SNTMNT (both companies were part of ABN AMRO's incubation program). However, the interests and strategic objectives still need to be coordinated between the actors (Project Team). However, as TH has put it, "as long as you pay the other parties, they will work with you".

In some of the interviews, this was not explained to the respondents (i.e. PB, TC). Perhaps this might be the reason why they considered this CSF as satisfactory. Nevertheless, we deem the CSF as *unsatisfactory*. It may turn out that certain service providers (like SNTMNT) are not willing to be behind a platform since a collaboration like this reduces their margin (TH). Instead, they might prefer targeting end-users directly. Thus, we still need to ascertain whether all business models are aligned by talking to the selected service providers.

Acceptable division of roles

In all interviews, the respondents indicated that they consider all required business roles fulfilled. Additionally, it was clear to each of them that the platform provider should be the keystone actor which has to lead the value network as indicated in the business model outline. Consequently, we regard this CSF as *satisfactory*.

Critical Success Factors Evaluation Summary

The evaluation results of each CSF are presented in Table 19 below.

CSF	Evaluation result
Compelling value proposition	Unsatisfactory
Clearly defined target group	Unsatisfactory
Unobtrusive customer retention	Unsatisfactory
Acceptable quality of service	Still unclear
Acceptable profitability	Still unclear
Acceptable risks	Unsatisfactory
Sustainable network strategy	Unsatisfactory
Acceptable division of roles	Satisfactory

Table 19: CSF evaluation results

We can see from the table that almost all CSFs necessitate further refinement of the business model with their associated CDIs. In this way we want to assure that no CDIs was left insufficiently addressed. The first four CSF relate to creating customer value while the last four determine whether network value is created. We can see that, the business model is *not* viable since it does not create sufficient customer *and* network value yet.

7.4.4. Scalability

On all occasions, the scalability of the platform and its services was deemed aligned with its potential for technical scalability (Project Team, PB). The reason was attributed to the possibility of scaling the underlying infrastructure easily on demand.

Nevertheless, in the scenario of high (unexpected) demand for the platform's services several limitations were identified. First, financial limitations could prevent the platform provider from meeting this demand (Project Team). It is still unclear whether cash flow problems will occur since no financial and legal arrangements have been made with IN2IP and SNTMNT yet (Project Team). Second, additional human resources might be needed that are not available (PB). Customer service needs to be expanded and enough software developers must be available for the development of new big data services (PB). Finally, unavailability of data scientists might also pose a problem (PB).

In summary, the business model is technically scalable however financial and human resources might pose a barrier for meeting unexpectedly high demand. Being aware of these limitations allows the platform provider to devise a plan for action in case such demand is observed.

7.4.5. Sustainability

During evaluation of the business model's sustainability, the technical architecture was deemed capable of absorbing new technologies (Project Team, PB). This is particularly important as new big data (open source) technologies emerge every month.

The issue of new competitors offering similar or better functionalities can be tackled by investing more in promotion (Project Team). However, this has implications for the financial domain as it can be a significant source of cost for which investment is needed. In addition, staying agile and adjusting the business model were also mentioned as possible lines of action (Project Team). That is, by being agile the platform provider can find niche markets where it can specialize itself and maintain an edge over existing or new competitors.

Changes in privacy regulations on data were identified as another environmental factor that can affect the business model (Project Team, PB). It was argued that regulation on personal or transactions data could only get stricter (PB) which can effectively limit the possibilities for analyses of such data on the platform. Thus, if a service that deals with personal data is to be built, the risks of changes in privacy regulation must be addressed accordingly (Project Team). However, at this point in time no such services are envisioned.

Furthermore, in the event of a partner ending their participation in the value network, two options are available (TC). Either a new partner has to be found or the platform provider should acquire the resources and capabilities by themselves. The impact of such event has not been considered in the outlined business model. We regard it as highly unlikely however its consequences can be devastating. Thus, the platform provider needs to consider mitigation strategies on time.

With respect to the insights discussed above, we can conclude that the overall sustainability of the business model can still be improved.

7.3. Conclusion

This chapter had the purpose of answering the fifth research question: *To what degree the designed business model can be considered as complete, consistent, viable, scalable, and sustainable?* In order to answer this final question we assessed the outlined business model both internally with the project team and externally with experts and practitioners. The main *findings* and *conclusions* from the evaluation are summarized below.

The outlined business model of the platform and its two pilot services is not yet complete since several gaps were identified during the discussions. These include framing the value proposition of the pilot services in terms of (measurable) organization goals; addressing the issue of multi-homing; and adopting a more systematic approach for risk management among others (for the complete list see *7.4.1. Completeness*).

Moving on, no inconsistencies between the four domains were identified. However, as the business model is not yet complete and viable (see below) we cannot claim that it is consistent. Thus, after further refinement with CDIs, the business model's consistency must be checked again.

We found out that the proposed customer and network value are not yet sufficient to sustain the business model for the platform and the two pilot services. The reason is that only one CSF was deemed satisfactory (Acceptable division of roles). Two of the seven other CSFs were assessed as 'still unclear', while the rest were found to be unsatisfactory. Thus, we conclude that the outlined business model is *not viable* and further refinement with CDIs is to be undertaken.

Furthermore, the platform's business model was regarded as easy to scale in terms of technical scalability. However, financial and human resources were identified as potential barriers for meeting unexpectedly high demand. More precisely, maintaining positive cash flows in case of rapid growth of demand can turn out to be problematic. Timing of payments to service/data providers (e.g. IN2IP, SNTMNT) should be negotiated so that service fees from SMEs are collected first. If this is not achieved or not possible at all, cash flow problems will emerge. Consequently, we cannot claim that the outlined business model is financially scalable at this point.

Finally, we found out that the sustainability of the business model can still be improved. From a technical point of view, the business model is deemed sustainable as new technologies can be integrated with little effort because of the service-oriented architecture. However, the effects of events like partner exit and competitor entry should be considered in more depth (e.g. by conducting SWOT analysis).

Based on the evaluation findings listed above we were able to draw several *conclusions*. We saw that at an early production phase the platform on its own does not add value for SME customers. Its overall added value is the sum of the values of its initial big data services as no network effects exist yet. Thus, *until network effects are induced, the each big data service must have a compelling value propositions on its own*. However, we saw that in the current business model outline, the value proposition of the two pilot services is not yet compelling. Consequently, further validation and refinement of the two pilot services can be pursued. A starting point can be linking the services to measurable organizational goals as already discussed. If this turns out to be unachievable, then these service ideas must be abandoned. In the same time, other big data service ideas can be identified and developed with launching customer organizations. Eventually, by building up the base of complementary big data services the platform will create more value than the sum of the values of its services (as indicated during several interviews).

The next logical step would then be creating a customer base of data scientists on the supply side. The data scientists target groups were found to be insufficiently clear in the outlined business model. During the evaluation it was suggested on several occasions that data scientists should be targeted depending on what datasets and services are currently available on the platform. That is, for each service the domain of expertise of data scientists should be considered. At this point in time, this is not a big issue. Eventually, when a mass market strategy is pursued, targeting this customer group will be crucial for the platform's success in terms of adoption.

Finally, we argue that in the platform's early design phase *it can be beneficial to design separate business models for each complementary service*. In this way complexity in the platform's business model can be reduced and in the same time the additional business models for complementary services can be designed in depth. This *does not* mean that the platform should be isolated from its services. Rather, what we propose is a way of structuring: the platform's business model will focus mostly on creating economies of scale and scope, while specific service details will be addressed within each service business model. Therefore, the platform's business model encompasses the ones of its complementary services.

8

Conclusions, Limitations and Recommendations

With this research, we pursued an idea for bringing big data analytics closer to small and medium-sized (SME) organizations. Our aim was to achieve this with a software-as-a-service (SaaS) platform which will connect SMEs to data scientists in a common business ecosystem. Thus, the following research goal was set at the beginning of the project:

> ***To design a big data analytics software-as-a-service platform and the underlying business model from the platform provider's perspective***

In order to achieve this goal, five research questions were formulated and answered within different research phases. In the next section the findings and conclusions of these five stages are presented. In the second and third sections, we discuss practical implications and theoretical contributions of the research respectively. After that, limitations of the research are discussed and finally recommendations for future research are provided.

8.1. Conclusions

Each of the five research questions was addressed in a separate chapter in this thesis. In this section we provide the answers to these questions once again and present our conclusions. The first research question was set to:

Q1) What are business models, platforms, and business ecosystems and what critical design issues related to these concepts can be drawn from the existing knowledge base

This question was answered by conducting literature review. The concepts of business models, platforms and business ecosystems were defined for the purposes of this project.

First, a business model was defined as: *"a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational an financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors"* [83]. Second, a platform is "a shared, stable set of hardware, software, and networking technologies on which users build and run computer applications" [61]. And, third, a business ecosystem was defined as a loosely connected network of organizations that emerge around a core technology (e.g. a digital platform) [74, 76].

The main focus of the first question however is the identification of business model critical design issues (CDIs) from the three literature streams. A CDI is a pre-defined design variable which is considered of crucial importance for the viability and sustainability of a business model [4].

The most comprehensive list of CDIs was found within the STOF method for business modeling. We were able to supplement this list with eight additional CDIs that were identified in the literature. Two CDIs were added from the VISOR business model method, four were added from platform theory, and two more from business ecosystem theory. The list of CDIs and their positioning within the STOF method is presented in Table 20 later in this chapter (see *8.3 Theoretical Contributions*).

The second research question focused on the relevance of big data analytics and the envisioned platform for SMEs:

Q2) Are big data analytics and the envisioned SaaS platform relevant for small and medium-sized enterprises?

To answer this question we conducted four interviews with representatives from different SMEs. As all of them indicated particular business needs for big data analytics, we concluded that big data analytics is indeed relevant for SMEs. However, we found out that the SMEs have business needs for big data analytics from *different calibers*. Thus, in order to adapt the platform to different calibers of business needs, we made the design decision to *build the platform in a modular way*. On one hand, basic services (e.g. data collection, pre-processing, data storage, and data export) will be available to SME end-users. On the other hand, the same basic services will be re-used as “building blocks” for more complex end-to-end services (e.g. data analytics and reporting).

Moving on, the third research question addressed the information system design stages before prototyping:

Q3) What are the goals, requirements and structural specifications for the big data SaaS platform?

The goals of the big data SaaS platform were set by taking into account the findings from the business needs interviews and the research domain. The following two goals were defined:

1) *To provide SMEs with access to big data services within different application service layers in order to meet different calibers of business needs*
2) *To facilitate the collaboration between SMEs and data scientists for the development of reliable big data analytics services*

Furthermore, the functional requirements for the system were derived from the defined set of goals by means of desk research. For the elicitation of user requirements, a second round of interviews was conducted. Potential users from the groups of SMEs and data scientists were asked about their requirements regarding user interface, pricing and general system aspects (see Appendix G for the list of gathered user requirements). Contextual requirements were also defined by means of desk research.

Finally, we described the structural specifications (i.e. technical architecture) of the platform that meet the set of system requirements (see Figure 34).

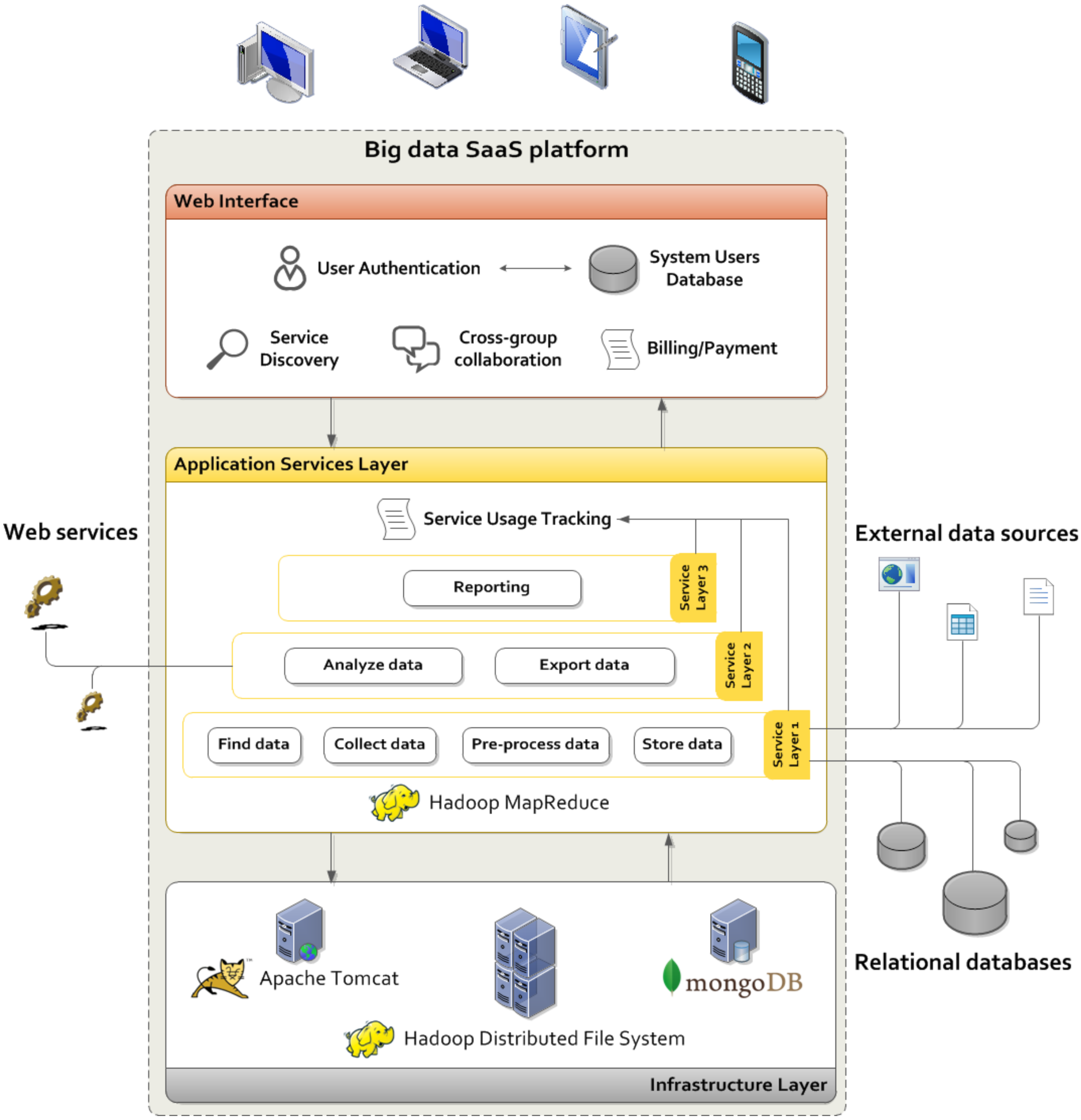

Figure 34: High-level technical architecture

In addition, two ideas for pilot big data services were provided by Dialogues Technology. The first idea is for a sentiment analytics service based on content from Twitter messages. SMEs end-users will be able to specify search terms (e.g. product/brand name) for finding relevant tweets first. After all tweets are found, content analysis will be used to determine the sentiment behind each search result (outsourced to SNTMNT). Finally, sentiment results will be aggregated and displayed in the form of easily readable graphs on the platform. By using this service SMEs can choose between competing suppliers, for instance. That is, they can compare sentiment analysis results of competing brands and based on that to determine with which suppliers to work or which products to buy/sell. The second service idea is for a predictive analytics service based on weather data (e.g. air temperature, cloudy/sunny/raining, precipitation, etc.) and internal data supplied by SMEs (e.g. sales figures, number of lead calls, etc.). This service will allow SMEs first to see how weather influences different business performance indicators and second to predict similar weather influences in the future.

The fourth and fifth questions address the business model design:

Q4) How does the business model of the designed big data SaaS platform look like?

To answer this question, the researcher performed the quick scan stage of the STOF method. All four domains of the STOF model were outlined. Insights from the two interview rounds were used as input when describing the first two domains: Service and Technology. Additionally, the eight CDIs that were identified in the Theoretical Background chapter were applied to the case of the big data SaaS platform. Reflections on the modified STOF method are presented in later in this chapter.

Q5) To what degree the designed business model can be considered as complete, consistent, viable, scalable, and sustainable?

To answer the final research question, the outlined business model was evaluated both internally with the project team and externally with experts and practitioners. The results of the evaluation led to the following conclusions:

The outlined business model of the platform and its two pilot services is not yet complete since several gaps were identified during the discussions. These include framing the value proposition of the pilot services in terms of (measurable) organization goals; addressing the issue of multi-homing; and adopting a more systematic approach for risk management among others. Moving on, no inconsistencies between the four domains were identified. However, as the business model is not yet complete we cannot claim that it is consistent. Furthermore, we found out that the proposed customer and network value are not yet sufficient to sustain the business model for the platform and the two pilot services. The reason is that only one CSF was deemed satisfactory (Acceptable division of roles). Thus, we conclude that the outlined business model is *not viable* and further refinement with CDIs is to be undertaken. Moving on, the platform's business model was regarded as easy to scale in terms of technical scalability. However, financial and human resources were identified as potential barriers for meeting unexpectedly high demand. More precisely, maintaining positive cash flows in case of rapid growth of demand can turn out to be problematic thus we cannot claim that the outlined business model is financially scalable at this point. Finally, we found out that the sustainability of the business model can still be improved. From a technical point of view, the business model is deemed sustainable as new technologies can be integrated with little effort because of the service-oriented architecture. However, the effects of events like partner exit and competitor entry should be considered in more depth (e.g. by conducting SWOT analysis).

In summary, we showed that big data analytics is relevant for SMEs and we identified concrete business needs. We also designed the big data SaaS platform in terms of goals, requirements and structural specifications. Based on this design, a prototype of the system can now be created. Next, we outlined the business model underlying the designed platform by taking the perspective of the platform provider. Finally, the business model was evaluated both internally by the project team and externally with experts and practitioners. Thus, we executed a whole design cycle which resulted in an initial design for the big data SaaS platform and its business model. Therefore, regardless of the evaluation results, we argue that *the research goal set in the beginning of the project has been achieved*. Nevertheless, further refinement of the business model (until viability is achieved) is needed and will be the subject of subsequent design cycles.

8.2. Practical Implications

The big data SaaS platform that we began designing in this thesis project has important practical implications for the customer group of SMEs.

We found out that SMEs have business needs for big data analytics from different calibers. On one hand, for some of the companies it is only feasible to use services for collecting, pre-processing, storing and exporting big data. The reason is that data security and privacy restrictions might prevent them from analyzing data in an off-premise fashion. On the other hand, other companies which do not use sensitive data for analytics purposes could also subscribe to data analysis and reporting services. Thus, *the platform will allow its SME customers to flexibly choose which parts of a big data analytics business process they want to outsource*. This is technically possible as we have designed the big data SaaS platform in a modular way by implementing service-oriented architecture (SOA).

Furthermore, the different big data services will run on the platform's infrastructure. In this way SMEs can focus on their core competences instead of building and maintaining their own *big data IT infrastructure*. Additionally, big data technologies are still relatively new (e.g. Hadoop and different NoSQL databases) and the pool of available specialists is not very big yet. We also argue that the big data SaaS platform will increase business agility for the SME customer group. That is, with relatively little effort and investment they will be able to try out and use the platform's big data services.

Another practical point that we make is that companies need to be aware of the inherent issues of big data analytics. First, people are oftentimes tempted to infer correlation, or even causality, when they see a pattern. However, statistical analyses *must* be performed in order to assure that such relationships indeed exist. Second, big data (especially social data) do not always represent the truth objectively. Datasets are the results of human design and as such they can be biased [39]. Biases can be introduced during data collection or even data analysis. Decision makers need to be aware that hidden biases exist and be always on the lookout for them. The platform will connect SMEs to data scientists who are trained at surfacing such biases and carefully drawing conclusions from data. Thus, we argue that the platform will be beneficial to SMEs in need of big data analytics solutions but lacking the necessary competencies.

8.3. Theoretical Contributions

In this thesis, we have devised an extension of the STOF method for business modeling. The proposed extension deals specifically with design issues related to business models for ICT (SaaS) platforms like the one designed in this research. The proposed modifications include:

- Addition of eight CDIs identified from business model, platform and business ecosystem theories. The CDIs are positioned within the STOF method in Table 20. These CDIs can be described either early in the quick scan stage or refined after evaluation with critical success factors (CSFs).

Critical Design Issue	STOF domain(s)	STOF Critical Success Factors
Selection of interfaces (VISOR model)	Technology domain	Compelling value proposition
Selection of service platforms (VISOR model)	Technology domain	Compelling value proposition; Acceptable quality of service; Acceptable profitability; Sustainable network strategy

Critical Design Issue	STOF domain(s)	STOF Critical Success Factors
Two-sided market dynamics (Platform theory)	Service domain	Compelling value proposition; Clearly defined target group
Platform openness (Platform theory)	Service domain	Compelling value proposition; Acceptable quality of service
Platform complementary services (Platform theory)	Organization domain	Acceptable quality of service; Sustainable network strategy; Acceptable division of roles
Pricing (Platform theory; STOF)	Service domain; Finance domain	Compelling value proposition; Clearly defined target group; Unobtrusive customer retention; Acceptable quality of service; Acceptable profitability; Acceptable division of roles
Business ecosystem strategy (Business ecosystem theory)	Organization domain	Acceptable profitability; Sustainable network strategy; Acceptable division of roles
Business ecosystem governance (Business ecosystem theory)	Organization domain	Sustainable network strategy; Acceptable division of roles

Table 20: Positioning of artifact-specific CDIs within the STOF method

- Addition and modification of several CSF evaluation questions based on the identified CDIs:

Critical Success Factor	Questions
Compelling value proposition	• Is the value proposition sufficiently attractive to customer groups from each platform side? • Do additional complementary services make the value proposition more attractive to demand side customers? • Does a large customer base on the demand side make the value proposition more attractive to supply side customers?
Clearly defined target group	• Are the target groups for each platform side sufficiently clear? • Do we know enough about each customer group in terms of their needs, preferences, capabilities, available resources, and sensitivities? • Is the strategy on how to start the market dynamics between the different sides of the platform clear?
Unobtrusive customer retention	• Is subsidizing going to be used as a retention mechanism? Is it clear which customer groups should be subsidized and how?
Acceptable profitability	• Is it clear who will incur the customer subsidy costs if such exist?
Sustainable network strategy	• Is there a common goal to which all business ecosystem members can adhere and strive to?
Acceptable division of roles	• Is there a keystone actor willing and capable to lead the value network?

Table 21: Modified list of CSFs from the STOF method

The modified STOF method was empirically illustrated for the case of the big data SaaS platform. We addressed all eight CDIs during the quick scan stage and we used the modified list of CSF questions (Table 21) during the evaluation of the business model outline. Next, we reflect upon the application of the proposed extension.

8.3.1. Reflection on the Proposed Extension of the STOF Method

The researcher finds the proposed extension useful as it explicitly addresses important issues related to business models for ICT platforms. That is, by using the modified STOF method, we were forced to think about issues which might have been neglected otherwise.

We saw that some of the additionally included CDIs have major impact on the designed business model. In our experience, decisions on starting the two-sided market dynamics between the different platform sides were core. That is, the decision whether to attract demand side or supply side customers first had important consequences for other business model aspects (e.g. targeting, pricing, etc.). In addition, the service platforms CDI forced us to think about what existing service platforms can be leveraged. For instance, the decision to outsource IT infrastructure reduced the inherent technological and financial risks for the platform provider. Pricing was also found to be important as customer subsidization was added to the basic pricing CDI within STOF.

We also saw that some of the newly added CDIs did not have many implications for the business model in an early launch phase. For instance, all platform complementary (big data) services will be developed by the platform provider initially. However, the platform could be opened technologically to external service developers who can create new big data services. If such decision is eventually taken, it will have major impacts on all business model domains. However, initially that is not the case in this project. Furthermore, it was somehow evident that the platform provider needs to take the role of a keystone player as it aims at creating a business ecosystem in which all actors must thrive.

Moving on, the modification of the list of CSF evaluation questions helped us with identifying gaps in the business model outline that necessitate further refinement. For instance, attention to clearness of target groups from *both platform sides* resulted in the conclusion that target groups within the supply side must be defined more clearly.

Finally, we were able to identify another platform-related CDI during the evaluation of the outlined business model. According to the expert in platform theory, the issue of *multi-homing* should also be addressed in platform business models. Multi-homing relates to end users using services of more than one platform with similar value propositions. Thus, finding a way to prevent multi-homing should be addressed appropriately.

8.4. Research Limitations

In just five months, from February until June 2013 we were able to execute a whole design cycle starting with business needs identification and ending with business model evaluation. However, conducting many activities came at the cost of depth. In this section we discuss the limitations of the research.

First, a small number of interviews were conducted in the business needs identification phase. Representatives of only four SMEs were interviewed thus we cannot claim that the conclusions drawn from these interviews are truly representative for whole SME population in The Netherlands. The same generalizability issue can be identified in the second round interviews as well. User requirements were gathered from only two representatives from each platform user group. Unfortunately, finding more respondents who are available or willing to participate in these two research phases proved to be hard as time was pressuring the researcher to continue with further design cycle activities. The researcher now recognizes that perhaps his ambition level was set too high which prevented him from conducting fewer activities but at a higher depth.

Another limitation that we can identify is the absence of a system prototype which we could have evaluated with potential system users. At the end of the research project, only a prototype of the Twitter analytics service was built. However, none of the platform's functionalities were prototyped yet. Thus, evaluation of the Twitter service prototype would not have yielded many useful insights.

A third limitation is that the business model underlying the big data SaaS platform was not found to be viable. Thus, further refinement with CDIs is required until viability is positively assessed. In addition, business models of partner organizations are only assumed to be aligned with the business model for the platform. It may be the case that external service providers (like sntmnt.com) would not want to collaborate with the platform provider. A possible reason can be that they want to have ownership of customers (i.e. SMEs) instead of accessing them through a platform.

Finally, the modified STOF method was applied in a single case. During the evaluation stage one additional CDI was identified with the help of the platform theory expert. Thus, we recognize that there might still be platform-related CDIs that missed in the literature review. By applying the modified STOF method to more platform cases such omissions can be noticed and corrected.

8.5. Future research recommendations

The limitations discussed in the previous section serve as a base for our recommendations for future research. They are presented next.

First, we propose that an online survey is distributed to SMEs from different industries in The Netherlands. We argue that the survey should focus on validating the conclusions that we have drawn from the first round of interviews. Respondents can be prompted whether they perceive particular data problems or opportunities that can be solved with the help of data scientists on the platform. Also, it will be useful for targeting what types of companies perceive off-premise data analyses as problematic.

Furthermore, as soon as a launching customer has been found, a prototype of the platform should be developed. User requirements collected in this research can be implemented in the prototype. However, not all of the user requirements should be fulfilled in this first prototype. As data scientists will join the platform at a later stage, developing their part of the system is not yet required. Thus, we propose that a minimum set of viable functionalities is built first. The prototype can then be evaluated with the launching customer and, if possible, with other potential customers.

The platform business model designed in this thesis was not found to be viable thus further refinement with CDIs should be undertaken until viability is achieved or the idea is deemed unfeasible. Finally, validation of the proposed extension of the STOF method should be pursued in more cases where multi-sided platforms are involved. A complete validation can focus on applying the method for different purposes like [4]:

- Testing the viability of a new platform service concept
- Describing the business models of existing platform services
- Designing a platform service based on existing services (e.g. a service mash-up)

Bibliography

1. Schilling, M.A., *Strategic management of technological innovation*. 2005: Tata McGraw-Hill Education.
2. Eisenmann, T.R., G. Parker, and M.V. Alstyne, *Opening platforms: How, when and why?* 2008, Harvard Business School Entrepreneurial Management Working Paper.
3. Järvi, K. and A. Pellinen, *Using Value Co-Creation to Redefine Business Models*, in *Technology Innovation Management Review*. 2011. p. 22-28.
4. Bouwman, H., et al., *Creating successful ICT-services: Practical guidelines based on the STOF method*. 2008, Enschede, Netherlands: Telematica Instituut.
5. Hevner, A.R., S.T. March, and J. Park, *Design Science in Information Systems Research.* MIS Quarterly, 2004. **28**(1): p. 75-105.
6. Mell, P. and T. Grance, *The NIST definition of cloud computing (draft).* NIST special publication 800, 2011. **145**.
7. IDC. *IDC Predicts 2012 Will Be the Year of Mobile and Cloud Platform Wars as IT Vendors Vie for Leadership While the Industry Redefines Itself* 2011; Available from: http://www.idc.com/getdoc.jsp?containerId=prUS23177411.
8. Bouwman, H., et al., *Chapter 2 Conceptualizing the STOF Model*, in *Mobile service innovation and business models*. 2008, Springer Verlag.
9. Verschuren, P. and R. Hartog, *Evaluation in Design-Oriented Research.* Quality and Quantity, 2005. **39**: p. 733-762.
10. Vaquero, L.M., et al., *A Break in the Clouds: Towards a Cloud Definition.* ACM SIGCOMM Computer Communication Review,, 2008. **39**(1): p. 50-55.
11. Wang, L., et al., *Cloud computing: a perspective study.* New Generation Computing, 2010. **28**(2): p. 137-146.
12. Buyya, R. *Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility*. in *Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium*. 2009. IEEE.
13. Armbrust, M., et al., *A view of cloud computing.* Communications of the ACM, 2010. **53**(4): p. 50-58.
14. Marston, S., et al., *Cloud computing — The business perspective.* Decision Support Systems, 2011. **51**(1): p. 176-189.
15. Dihal, S., *Mobile Cloud Ecosystems. Evaluating the feasibility and viability of smartphones as a shared resource pool*. 2011, TU Delft.
16. Keahey, K., et al., *Sky computing.* IEEE Internet Computing, 2009. **13**(5): p. 43-51.
17. Sotomayor, B., et al., *Virtual infrastructure management in private and hybrid clouds.* IEEE Internet Computing, 2009. **13**(5): p. 14-22.
18. Gonçalves, V. and P. Ballon, *Adding value to the network: Mobile operators' experiments with Software-as-a-Service and Platform-as-a-Service models.* Telematics and Informatics, 2011. **28**(1): p. 12-21.
19. Mitchell, D. *Defining Platform-As-A-Service, or PaaS*. 2008.
20. Azvine, B., Z. Cui, and D.D. Nauck, *Towards real-time business intelligence.* BT Technology Journal, 2005. **23**(3): p. 214-225.
21. Jourdan, Z., R.K. Rainer, and T.E. Marshall, *Business Intelligence: An Analysis of the Literature.* Information Systems Management, 2008. **25**(2): p. 121-131.
22. Negash, S. and P. Gray, *Business Intelligence.* Handbook on Decision Support Systems 2, 2008. **175-193**.
23. Elbashir, M.Z., P.A. Collier, and M.J. Davern, *Measuring the effects of business intelligence systems: The relationship between business process and organizational performance.* International Journal of Accounting Information Systems, 2008. **9**(3): p. 135-153.

24. Wixom, B. and H. Watson, *The BI-based organization.* International Journal of Business Intelligence Research (IJBIR), 2010. **1**(1): p. 13-28.
25. Hilmer, F. and J. Quinn, *Strategic outsourcing.* Sloan management review, 1994. **35**(4): p. 43-55.
26. Gartner. *Gartner Says Nearly One Third of Organizations Use or Plan to Use Cloud Offerings to Augment Business Intelligence Capabilities*. 2012 [cited 2012 25 Feb]; Available from: http://www.gartner.com/newsroom/id/1903814.
27. Gartner. *Gartner Executive Program Survey of More Than 2,000 CIOs Shows Digital Technologies Are Top Priorities in 2013*. 2013 [cited 2013 23 Feb]; Available from: http://www.gartner.com/newsroom/id/2304615.
28. Gantz, J. and D. Reinsel, *Extracting Value from Chaos*. 2011, IDC.
29. Laney, D. *3D data management: controlling data volume, velocity and variety*. 2001.
30. Gens, F., *IDC Predictions 2013: Competing on the 3rd Platform*. 2013, IDC.
31. ITU, *Measuring the Information Society*. 2012, ITU.
32. Murugesan, S., *Understanding Web 2.0.* IT Professional, 2007. **9**(4): p. 34-41.
33. DuVander, A. *6,000 APIs: It's Business, It's Social and It's Happening Quickly*. 2012 May 22nd, 2012 [cited 2013 23rd Feb]; Available from: http://blog.programmableweb.com/2012/05/22/6000-apis-its-business-its-social-and-its-happening-quickly/.
34. Atzori, L., A. Iera, and G. Morabito, *The Internet of Things: A survey.* Computer Networks, 2010. **54**(15): p. 2787-2805.
35. Gartner. *Gartner Says Business Intelligence and Analytics Need to Scale Up to Support Explosive Growth in Data Sources* 2013 [cited 2013 26 Feb]; Available from: http://www.gartner.com/newsroom/id/2313915.
36. McKendrick, J. *The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data*. 2011.
37. Gartner. *Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data*. 2011 [cited 2013 26th Feb]; Available from: http://www.gartner.com/newsroom/id/1731916.
38. White, C., *Using Big Data for Smarter Decision Making*. 2011, BI research.
39. Crawford, K. *The Hidden Biases in Big Data*. 2013 [cited 2013 4th June]; Available from: http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html.
40. Press, G. *What's to be Done about Big Data?* 2013; Available from: http://www.forbes.com/sites/gilpress/2013/03/11/whats-to-be-done-about-big-data/.
41. De Reuver, M. and T. Haaker, *Designing viable business models for context-aware mobile services.* Telematics and Informatics, 2009. **26**(3): p. 240-248.
42. Shafer, S.M., H.J. Smith, and J.C. Linder, *The power of business models.* Business Horizons, 2005. **48**(3): p. 199-207.
43. Zott, C., R. Amit, and L. Massa, *The Business Model: Recent Developments and Future Research.* Journal of Management, 2011. **37**(4): p. 1019-1042.
44. Timmers, P., *Business Models for Electronic Markets.* Electronic markets, 1998. **8**(2): p. 3-8.
45. Amit, R. and C. Zott, *Value creation in E-business.* Strategic Management Journal, 2001. **22**(6-7): p. 493-520.
46. Dubosson-Torbay, M., A. Osterwalder, and Y. Pigneur, *eBusiness Model Design, Classification and Measurements.* Thunderbird International Business Review, 2001. **44**(1): p. 5-23.
47. Weill, P. and M. Vitale, *Place to space: Migrating to eBusiness Models*. 2001: Harvard Business Press.
48. Osterwalder, A. and Y. Pigneur. *An e-business model ontology for modeling e-business*. in *15th Bled Electronic Commerce Conference*. 2002. Bled, Slovenia.
49. Hamel, G., *Leading the revolution. 2000.* Boston: Harvard Business School, 2000.
50. Magretta, J., *Why Business Models Matter.* Harvard business review, 2002. **80**(5): p. 86-93.

51. Chesbrough, H. and R.S. Rosenbloom, *The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies.* Industrial and corporate change, 2002. **11**(3): p. 529-555.
52. Teece, D.J., *Business models, business strategy and innovation.* Long range planning, 2010. **43**(2): p. 172-194.
53. El Sawy, O.A. and F. Pereira, *Business Modelling in the Dynamic Digital Space: An Ecosystem Approach*. 2013: Springer.
54. Robertson, D. and K. Ulrich, *Planning for Product Platlorms.* Sloan management review, 1998. **39**(4): p. 19-31.
55. Meyer, M.H. and A. DeTore, *PERSPECTIVE: Creating a platform-based approach for developing new services.* Journal of Product Innovation Management, 2001. **18**(3): p. 188-204.
56. Van Vuuren, W. and J.I. Halman, *Platform-driven development of product families: linking theory with practice.* Journal of Product Innovation Management, 2003. **20**(2): p. 149-162.
57. Meyer, M.H. and J.M. Utterback, *The Product Family and the Dynamics of Core Capability*, in *MIT Sloan Management Review*. 1993.
58. Cusumano, M.A., *The evolution of platform thinking.* Communications of the ACM, 2010. **53**(1): p. 32-34.
59. Gawer, A., *Platforms, markets and innovation*. 2011: Edward Elgar Publishing.
60. Evans, D.S., A. Hagiu, and R. Schmalensee, *A Survey of the Economic Role of Software Platforms in Computer-based Industries.* CESifo Economic Studies, 2005. **51**(2-3): p. 189-224.
61. Bresnahan, T., *New modes of competition and the future structure of the computer industry*. 1999, Berlin: Springer.
62. Liu, G. *Research on independent SaaS platform*. in *ICIME 2010 - 2010 2nd IEEE International Conference on Information Management and Engineering*. 2010. Chengdu, Sichuan, China: IEEE.
63. Roson, R., *Two-Sided Markets: A Tentative Survey.* Review of Network Economics, 2005. **4**(2): p. 142-160.
64. Caillaud, B. and B. Jullien, *Chicken & egg: Competition among intermediation service providers.* RAND Journal of Economics, 2003. **34**(2).
65. Eisenmann, T.R., G. Parker, and M.V. Alstyne, *Strategies for Two-*

Sided Markets. Harvard business review, 2006. **84**(10): p. 82.
66. Rochet, J.-C. and J. Tirole, *Platform Competition in Two-Sided Markets.* Journal of the European Economic Association, 2003. **1**(4): p. 990-1029.
67. Evans, D.S., *The antitrust economics of multi-sided platform markets.* Yale Journal on Regulation, 2003. **20**(2).
68. Cusumano, M.A., *Cloud computing and SaaS as new computing platforms.* Communications of the ACM, 2010. **53**(4): p. 27-29.
69. Cusumano, M.A. and A. Gawer, *The Elements of Platform Leadership.* MIT Sloan Management Review, 2002. **43**(3): p. 51-58.
70. Gawer, A. and M.A. Cusumano, *How Companies Become Platform Leaders.* MIT Sloan Management Review, 2008. **49**(2): p. 28-35.
71. Gawer, A. and M.A. Cusumano, *Platform Leadership*. 2002: Harvard Business School Press.
72. Dou, Y., D.J. Wu, and J. Chen, *Platform Pricing with Strategic Buyers.* 2012: p. 4535-4544.
73. Moore, J.F., *Predators and Prey: A New Ecology of Competition.* Harvard Business Review, 1993. **71**(3): p. 75 - 86.
74. Iansiti, M. and R. Levien, *Keystones and dominators: Framing operating and technology strategy in a business ecosystem*. 2004, Boston: Harvard Business School.
75. Iansiti, M. and R. Levien, *Strategy as Ecology.* Harvard business review, 2004. **82**(3): p. 68-81.
76. Den Hartigh, E. and T. Van Asseldonk. *Business ecosystems: A research framework for investigating the relation between network structure, firm strategy, and the pattern of*

innovation diffusion. in *ECCON 2004 Annual Meeting: Co-Jumping on a Trampoline, The Netherlands*. 2004.
77. Anggraeni, E., E. Den Hartigh, and M. Zegveld. *Business ecosystem as a perspective for studying the relations between firms and their business networks*. in *ECCON 2007 Annual meeting*. 2007.
78. Iansiti, M. and R. Levien, *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. 2004: Harvard Business Press.
79. Fragidis, G., K. Tarabanis, and A. Koumpis. *Conceptual and business models for customer-centric business ecosystems*. in *Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES*. 2007. IEEE.
80. Basole, R.C., *Visualization of interfirm relations in a converging mobile ecosystem.* Journal of Information Technology, 2009. **24**(2): p. 144-159.
81. Von Tunzelmann, N., *Historical coevolution of governance and technology in the industrial revolutions.* Structural Change and Economic Dynamics, 2003. **14**(4): p. 365-384.
82. Vos, E., *Business Ecosystems: Simulating ecosystem governance*, in *Department of Technology, Strategy and Entrepreneurship*. 2006, Delft University of Technology: Delft.
83. Bouwman, H., T. Haaker, and H. De Vos, *Mobile Service Innovation and Business Models*. 2008: Springer Verlag.
84. SNTMNT. *Consumer Sentiment API*. 2013 [cited 2013 Jun 14th]; Available from: http://www.sntmnt.com/brands/.
85. Sommerville, I. and P. Sawyer, *Requirements engineering: a good practice guide*. 1997: John Wiley & Sons, Inc.
86. w3schools.com. *Browser Statistics*. 2013; Available from: http://www.w3schools.com/browsers/browsers_stats.asp.
87. European Parliament and European Council, *EU Data Protection Directive*. 1995.
88. Zorrilla, M. and D. García-Saiz, *A service oriented architecture to provide data mining services for non-expert data miners.* Decision Support Systems, 2012.
89. Minhas, A. and F.H. Vogt. *Service Orientation for a Dynamic Enterprise*. in *Proc. Int'l Conf. Logistics and Supply Chain Management (ILSCM'06)*. 2006. Citeseer.
90. Apache. *Hadoop 1.0.4 Documentation*. [cited 2013 May 8th]; Available from: http://hadoop.apache.org/docs/r1.0.4/index.html.
91. Osterwalder, A. and Y. Pigneur, *Business model generation: a handbook for visionaries, game changers, and challengers*. 2010: Wiley.
92. Fritscher, B. and Y. Pigneur, *Supporting Business Model Modelling: A Compromise between Creativity and Constraints*, in *Task Models and Diagrams for User Interface Design*, D. England, et al., Editors. 2010, Springer Berlin Heidelberg. p. 28-43.

A

Preliminary Competitive Analysis

Name	Description	Differences with the envisioned platform
Informly	Platform that integrates data from several different sources (Google Analytics, AdSense, PayPal etc.) into a single dashboard.	No data scientists in the value network. Non opportunities for customization.
Geckoboard	Similar to Informly, Geckoboard integrates different analytics tools in a single dashboard. However, they have developed much more tools than Informly. Furthermore, Geckoboard allows its customers to connect new tools through a developer API which makes their platform a PaaS instance.	No data scientists in the value network.
Think Big Analytics	The company delivers both consulting expertise and big data analytics solutions to their clients.	Think Big Analytics takes the role of data analysts. Consultancy services and analytics solutions are aimed at large companies.

Table 22: Preliminary competitive analysis

B

Business Model Methodologies

Different business model types can be found in the literature due to differences in the classification of their components. Consequently, many business modeling methodologies and frameworks like the business model canvas [91], STOF method [83], and VISOR method [53] exist. These frameworks are described and compared next.

The Business Model Canvas

In 2010, Osterwalder and Pigneur [91] defined the business model canvas as a framework for describing and visualizing business models (see Figure 35). They argue that a business model can best be described by using nine building blocks which show how a company intends to make money.

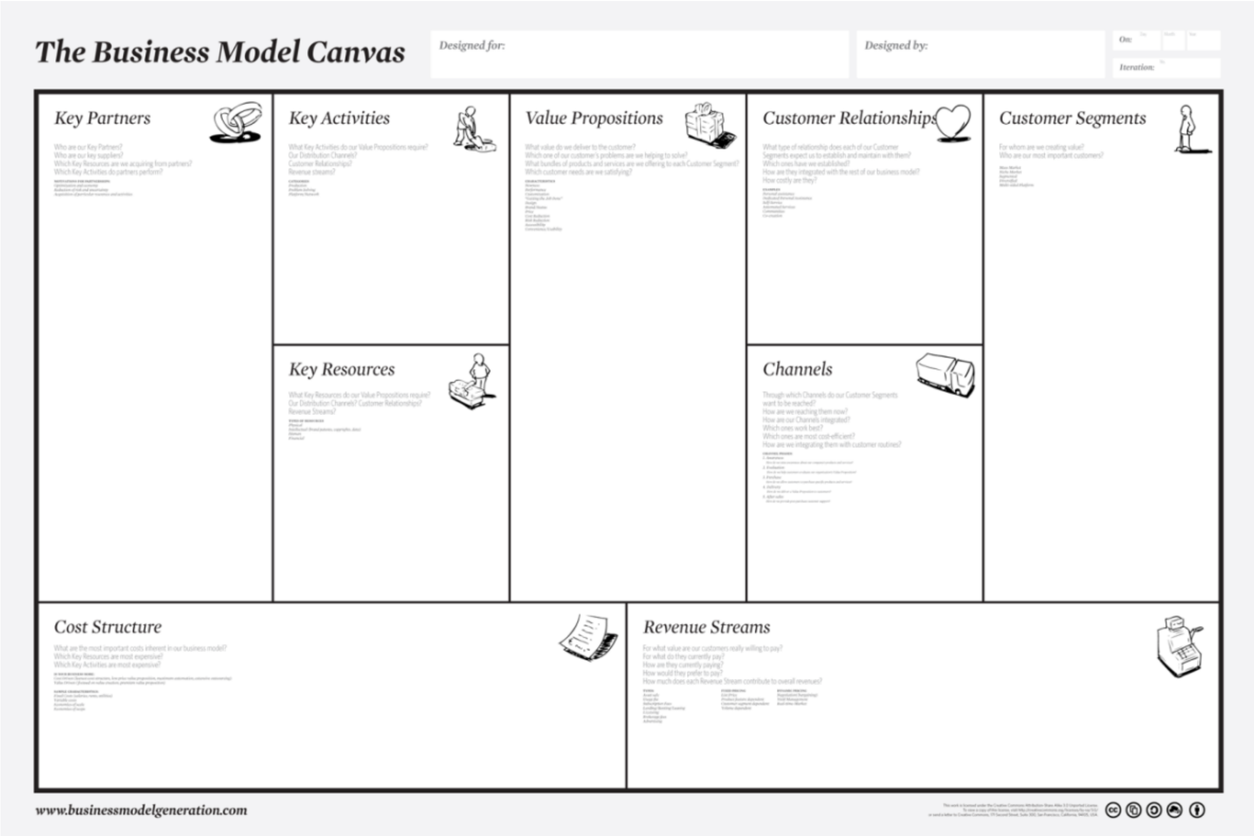

Figure 35: The business model canvas

The nine building blocks stem from four main areas of a business: customers, value proposition, infrastructure and financial viability. The nine building blocks are customer segments, value proposition, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure [91].

Fritscher and Pigneur [92] argue that the most interesting feature of the business model canvas is its ability to show the business logic of a company on a single page. The visual positioning of the nine blocks and the relationships between them also contribute to the power of the canvas [92].

The typical setting for designing business models with the canvas includes a whiteboard or a big piece of paper. An empty business model canvas must be drawn on the chosen working surface prior to start of the design activity. Participants in the design process can attach sticky notes with ideas written on them to any of the nine segments of the canvas. Additionally, notes can be stuck on a temporary area for later use or discussion. By using sticky notes, participants can easily add, remove or rearrange the different elements within each segment. When a strong relationship between different elements exists, a line can be drawn between them to visualize this relationship. By stimulating discussion between the participants involved in the business modeling activity, the canvas allows for the identifications of elements which might have been overlooked otherwise [92].

The STOF Method

Bouwman, Haaker [83] argue that business model designers should focus on four key domains, namely **S**ervice, **T**echnology, **O**rganization, and **F**inance (STOF) when drawing business models for ICT services. Figure 36 shows the four domains and the relations between them.

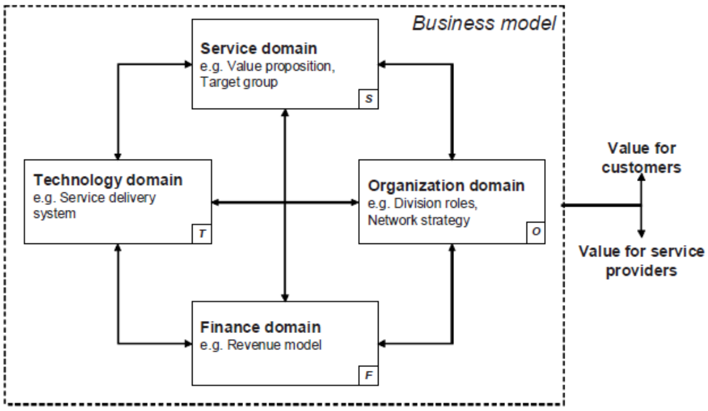

Figure 36: STOF business model domains [83]

The service domain should provide a description of the service offering, the value proposition and the targeted market segments among others. The technology domain is focused on the technical functionality that is required to realize the service offering. The organizational domain should describe the structure of the value network around the service and the position of the focal firm within that value network. And, finally, the finance domain should describe how a value network intends to generate revenue from the service offering including how this revenue, risk and investment will be shared within the value network. These four domains need to be designed in a careful and balanced way so that value is generated not only for the service customers, or end-users, but also for the various actors from the value network [4].

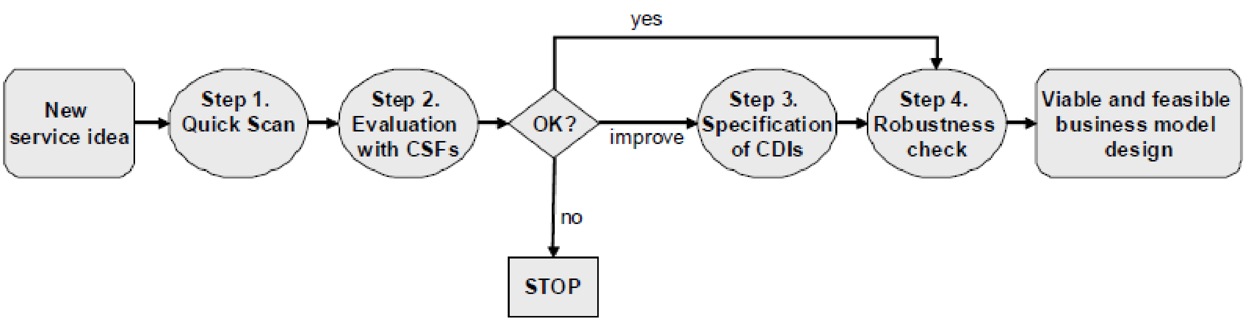

Figure 37: The STOF method [4]

A business modeling method, i.e. the STOF method, was developed on the basis of the STOF framework (see Figure 36). The method is a step-by-step approach for business model design that uses predefined design variables (or Critical Design Issues) which are considered as of crucial importance for the viability and sustainability of the future business model. Within the STOF method, an evaluation of a quick scan stage has to be performed, based on eight Critical Success Factors (CSFs). If one of these factors is assessed negatively, then CDIs related to that factor need to be further refined until the corresponding CSF is evaluated positively [4]. The STOF method allows for the design of viable and feasible business models.

The VISOR Method

El Sawy and Pereira [53] classify business model components in 5 categories, namely “Value proposition”, “Interface”, “Organizing model”, “Revenue model”, and finally “Service platforms” where explicit attention is given to IT platforms that can enable and support the business processes and relationships (e.g. IaaS platforms) thus improve the overall value proposition [53]. They argue that a successful business model is one that aligns these components so as to deliver the greatest value to the targeted customers and in the same time minimizes the service provisioning costs, as depicted in Figure 38 below.

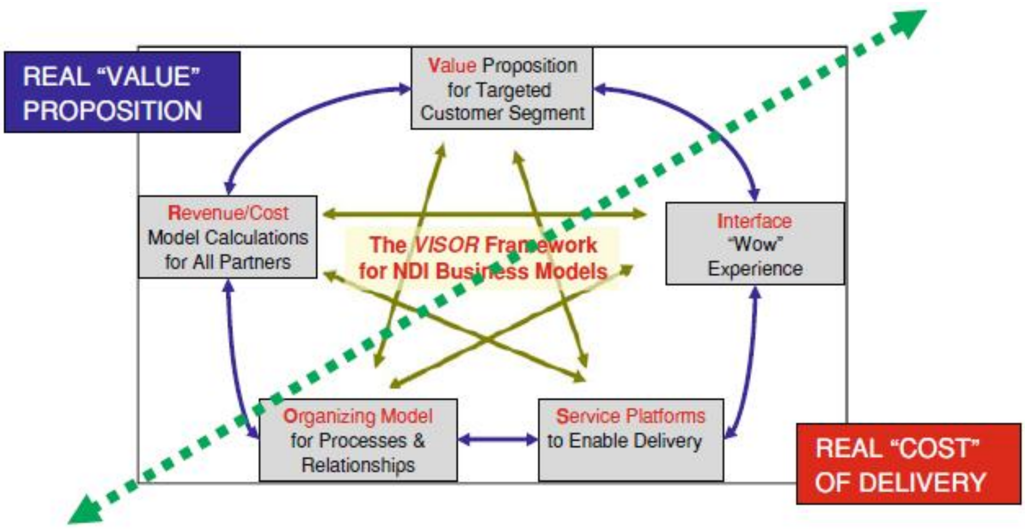

Figure 38: The VISOR model [53]

According to the VISOR method, the business modeling activity in the networked digital industry should be comprised of the following five steps:

1. *Value proposition.* A business model designer must identify why each customer segment would value an organization's product or service and whether they would be willing to pay a premium for them [53].
2. *Interface.* The user interface experience has important consequences on the success of delivery of a product or a service. Ease of use, simplicity, convenience, and positive energy should generate great customer experience with the product or service. A business modeler should assess how the unique characteristics of some new interface (e.g. smartphone, tablet, etc.) can enable the delivery of value [53].
3. *Service Platforms.* These are IT platforms that enable, shape and support the business processes needed for delivering products and services and for improving the value proposition. Service delivery in the networked digital industry depends on technology infrastructures thus organizations need to be aware of the existing platform ecosystems [53].
4. *Organizing model.* In the networked digital industry, business model designers must understand the new venture's dependencies on other firms (i.e. keystones or dominators). A company's competitive instinct may need to be balanced with an imperative to cooperate in order to improve the health of the whole ecosystem [53].
5. *Revenue model.* Finally, a business model designer should address the ways for a new venture to generate revenue. Revenues must exceed costs and be attractive to all members of the business ecosystem [53].

Business Model Frameworks Comparison

Table 23 provides a comparison between the three business model methodologies in terms of components and appropriate industry for application.

Business Model Framework	Components	Industry
Business Model Canvas	Customer segments; Value proposition; Channels; Customer relationships; Revenue streams; Key resources; Key activities; Key partnerships; Cost structure	Generic, not industry specific
STOF model	Service domain; Technology domain; Organization domain; Finance domain;	ICT
VISOR model	Value proposition; Interfaces; Service platforms; Organizing model; Revenue model;	ICT

Table 23: Business model frameworks comparison

The canvas is a generic tool in a sense that it could be used for business modeling in different industries. The STOF and VISOR methods, on the other hand, are specifically addressing business modeling for innovations in the Information and Communications Technology industry.

C

Design Science Research Guidelines

#	Guideline	Description	Adaptation to the Project
1	Design as an Artifact	Design-science research must construct a viable artifact in the form of a construct, model, method, or instantiation	The designed platform and its business model can both be considered as instantiations
2	Problem Relevance	The objective of design-science research is to develop technology based solutions to important and relevant business problems.	The relevance of the designed platform was pursued by conducting the business needs interviews at the start of the project
3	Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.	Expert opinions were used for the purposes of business model evaluation at the end of the project
4	Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, or design methodologies	During the design of the artifact and its business model, the STOF method was supplemented with critical design issues from platform and business ecosystem theory.
5	Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.	Research rigor was pursued by using the design science framework of Hevner, March [5]. In addition, an exhaustive set of expert views was pursued for business model evaluation
6	Design as a Search	The search for an effective artifact requires utilizing available means to reach a desired ends while satisfying laws in the problem environment.	By making use of design cycles we intend to iteratively search for an effective artifact that would solve existing problems for its end-users
7	Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.	The designed artifact will be presented to a large array of companies as soon as a stable set of services has been developed

Table 24: Design-Science Research Guidelines, adapted from [5]

D

First Round Interviews Protocol

This appendix contains the protocol for the business needs identification interviews. The overall interview is divided in 6 sections (Table 25) with the possibility to have a 7th one with backup questions.

Part	Time	Questions
1. Introduction	5 min	-
2. Business Needs Validation	10 min	Q1, Q2, Q3
3. Research Problem and Assumptions (A1-A4) Validation	10 min	Q4 – Q15
4. Solution Description	5 min	-
5. Solution Assumptions (A5) Validation	10 min	Q16, Q17, Q18
6. Finalization	5 min	-

Table 25: First round interview plan

Each section and its corresponding questions are described in detail below. There are 18 questions in total with 4 additional backup questions. Each question will be presented together with the rationale behind its selection.

Introduction

- Background of the researcher
- Interview objective
- Permission to record the conversation
- Interview structure
- Company size and turnover
- Responsibilities of the respondent and his/her department

Business Needs Validation

Questions 1 to 3 have the purpose of accomplishing the first goal of the interviews, i.e. to validate if SMEs perceive any real business needs for the envisioned SaaS platform.

Q1. Are you familiar with the big data concept and how do you understand it?
The first question is open in order to get initial impression on how the interviewee tends to answer such questions (i.e. shortly or in detail). It aims at understanding how the respondent grasps the concept of big data.

Q2. Do you see big data bringing value to your business in any way?
The ultimate goal of the second question is to find out whether the respondent's company recognizes any business needs that big data analytics can fulfill. At first, the question is asked

very broadly but if the respondent tends to give short answers or he/she has not thought about the subject, then more focused questions will be asked. For instance:

a. Is your company currently facing any problems that big data analyses could solve?
b. Has your company considered any new opportunities that big data analyses can create for you?

Q3. Why not? (Only if the respondent answered Q2 negatively. Else skip Q3)
Understanding the underlying factors that lead to the perception of the respondent that big data cannot bring value to his/her company is crucial. Factors could be either technical or business related.

Research Problem and Assumptions (A1-A4) Validation

The rest of the questions have the purpose of accomplishing the second goal of these interviews, i.e. to validate the assumptions made by the researcher upon problem statement formulation.

Q4. Do you currently use or plan to use any data analytics solutions?
Straightforward closed question which aims at understanding what big data analytics solutions the respondent's company uses.

Q5. Why not? (Only if the respondent answered Q4 negatively. Else skip Q5)
If the SME does not use any data analytics solutions then the reasons for that must be understood. Attention will be paid to any answers related to assumptions A1, A3, and A4, so that later questions that explicitly refer to these assumptions would be skipped.

Q6. Do you currently use any cloud-enabled services?
Straightforward closed question which aims at understanding what cloud-enabled services the respondent's company uses. If the respondent answers positively, attention will be paid whether A2 holds.

Q7. Why not? (Only if the respondent answered Q6 negatively. Else skip Q7)
The reasons for not using cloud-enabled services must be understood. Again, attention to any answer rejecting A2 must be paid.

The following questions aim at testing the five assumptions underlying the research problem. If one or more of the assumptions were already verified or rejected then the corresponding questions will be skipped.

Q8. Do you have the necessary resources to build or buy on-premise data analytics software?
Q9. Why not? (Only if the respondent answered Q8 negatively. Else skip Q9)

Q10. Do you regard the cloud-environment attractive for outsourcing IT infrastructure needed for business analytics?
Q11. Why not? (Only if the respondent answered Q10 negatively. Else skip Q11)

Q12. Does your company have the internal competencies to draw insights from big data?

Q13. Would you hire a business analytics consultancy company to help you with recognizing and drawing insights from big data?
Q14. Why not? (Only if the respondent answered Q13 negatively. Else skip Q14)
Q15. What is your perception on the costs of hiring external consultants? (Only if costs were not mentioned. Else skip Q15)

Solution Description

Our goal is to develop an online big data analytics platform that is based on the software-as-a-service delivery model (describe the SaaS delivery model if the respondent is not familiar with it).
We want the platform to be an intermediary between SMEs and data scientists or business intelligence experts (see the figure). We assume that by connecting these two stakeholder groups in a common environment, value can be created and captured by both of them.
First, SMEs will be able to access resources and capabilities that they do not possess on cost that is lower than the cost of obtaining data analytics solutions and hiring BI employees or consultants in the traditional on-premise way. Currently, we are thinking about having monthly subscription fees or pay-per-use pricing models for SMEs. Second, we think that data scientists can earn revenue by receiving parts of the service subscriptions of SMEs.
SMEs will be able to use different generic big data analytics applications that will be already available on the platform. However, they will also have the opportunity to collaborate with data scientists for recognizing new types of big data analyses that can bring value to their business. Then, new custom applications will be developed in an agile manner by the platform provider with the join participation of both SMEs and data scientists. SMEs will have the option to make such custom apps public in order to reduce or maybe even totally eliminate the application development costs.

Solution Assumptions (A5) Validation

The following questions will be asked after the SaaS platform and its value proposition are explained to the respondent.

Q16. What are your first thoughts on the platform I just described to you?
Q17. Do you think that the platform we intend to develop can bring value to your company?
Q18. Why not? (Only if the respondent answered Q16 negatively. Else skip Q17)

Backup Questions

The backup questions below are related to the SaaS platform's pricing model, exclusivity, trust, and acceptance.

1. What pricing model would you consider attractive for your company joining the platform?
2. Would you be willing to finance additionally the development of an application that you need in order to have exclusivity?
3. If not, would you then be willing to let other enterprises also use such an application?
4. To what degree would you trust the work of data scientists within such a platform?

Finalization

- Ask the respondent if he can read the interview report for validation purposes
- Thank the respondent.

E

First Round Interviews Reports

Each interview and its most important insights are discussed in more detail in the following sections. Light is shed on the SMEs' business needs for big data analytics solutions perceived by each respondent. Reasons for either verifying or rejecting the assumptions underlying the research problem are provided for each case as well.

Company A

Company A is operating in the financial services sector in The Netherlands. It is a payment service provider for one of the largest online retailers in the country. The company is specialized in the development of solutions in the area of credit scoring, customer database management and complex processes.

The interview respondent is currently taking the position of Interim Manager at Company A. During the interview, he did identify concrete business needs for the incorporation of big data analytics in the current solutions that the company provides. In his opinion, the company "could build a better picture of the creditability of their debtors" by adding unstructured data to their services. He gave two examples. First, changes in the relationship status of a person within social media have implications on their private environment. Second, shopping behavior data (from the online retailer partner) could be used for predicting spending patterns where, for example, a person who starts buying baby products probably had become a parent.

Some of the assumptions (i.e. A1 and A4) that were made during the formulation of the research problem are rejected for the case of Company A based on the interviewee's responses. The respondent indicated that the company operates in the financial services industry where legislation on financial processes would prevent the company from outsourcing data storage to the cloud for security considerations. Thus, A1 can be rejected since the company is forced to maintain high level of control over its IT infrastructure for storing personal financial data. In addition, Company A is part of a large holding from which financial resources could be obtained if necessary. Furthermore, A4, is rejected since the responded indicated that the company would hire a business analytics consultant "almost for sure" when they grasp the possibilities that big data can bring them.

All other assumptions (i.e. A2, A3, and A5) were verified. Assumption 2 was verified since Company A has been "has been 'cherry picking', based on cost and availability where cloud services will be used". Assumption 3 was also verified as the respondent indicated that the company's BI professionals have limited capabilities for drawing insights from big data. To follow up, he explained that the reason for this is that big data analytics are currently not in the agenda of the company. Finally, assumption 5 was also verified as the respondent answered positively the question if the described platform can bring value to Company A in the longer term.

Concluding, the interview with the respondent from Company A did lead to the identification of concrete business needs for the adoption of big data analytics solutions by the company. In his words, incorporating unstructured (i.e. big) data in the company's services "could help us in lowering risk and defining new propositions". However, whether and how the proposed solution will truly realize these business needs is still uncertain at this stage due to possible legislation constraints in the financial industry. It may well be the case that external services (i.e. services within the envisioned SaaS platform) can be called for analyses of data that is sent anonymously to the external world. Or data analysis can be done on the company's own IT infrastructure, but communication with data scientists and data collection could be provided by the platform's services. Consequently, deeper insights are needed in the user and contextual requirements for the system. These can be obtained later in this research during the second step of the design cycle.

Company B

Company B is a mortgage service provider and thus can also be placed within the financial services sector. The company is nested in the top SME segment with approximately 220 employees and yearly turnover of slightly less than €40 million. Company B provides mortgages to borrowers on behalf of a range of lenders, including banks and insurance companies. The company controls the entire mortgage life cycle from initial quotation to the total repayment of a mortgage, including mortgage application and collection of payments.

The interview respondent is the Chief Operations Officer of Company B. When asked if he sees how big data could bring value to his company, he replied: "*What we currently do is investigate if we can identify those clients* [i.e. who cannot pay back their loans] *before they are in trouble. And for that we can use the database in our system, but we might also use social media".* In his opinion, people get into trouble repaying their mortgages because they lose their jobs, they divorce or they get sick. He explained that if signs of these events can be detected on social media early then the clients can be approached with help before they get into serious financial problems. Therefore, it can be concluded that Company B do recognize particular business needs for adopting big data analytics.

In the course of the interview, the respondent indicated that they have their own IT infrastructure and business intelligence team to support the business processes associated with their mortgage services. As a consequence, A1 and A4 can both be rejected. The reasons for building their own infrastructure became obvious after the respondent was asked for further clarification. First, the company works with personal client data coming from their partners who prohibit them from storing the data on the cloud: "*The data we store are their* [i.e. banks and insurers] *data and we are not allowed to put those data in the public domain. The other way around is possible – we get data from social media, put it in our data in-house and use it for analytic purposes*". Second, working with personal client data and running predictive analyses are part of the core competences of Company B thus they are not willing to outsource these activities. Assumptions A2 is rejected. As the company is obligated to maintain high level of control over their clients' data, off premise data storage is not feasible for them. Assumption 3 and 5 are also rejected as the respondent indicated that their BI team has the competencies required for recognizing the right model and right types of big data for analysis.

In conclusion, for the case of Company B, big data should be matched to their current data and analyzed in an on-premise fashion. Thus, the SaaS platform could bring value to the company to a more limited degree – platform applications could collect, rank, aggregate and send relevant data from various sources to Company B's own data stores for further analyses done on their own infrastructure. Collaboration with data scientists on the platform and big data analyses are activities that are not relevant for this particular organization.

Company C

Company C is a media group comprising of several smaller companies in the newspaper, radio and online media business. Company C is also nested in the top segment of the SME category with 250 employees and yearly turnover of around €50 million.

The respondent takes the position of Manager Marketing Intelligence in the group. He indicated that his department is responsible for, among other things, maintaining the customer databases; analyzing whether customers are satisfied with the newspaper and their future needs; and reporting. Furthermore, the department does analyses mostly on internal data (e.g. editorial content, invoicing, etc.) but they also commission and use research from external companies (e.g. advertising expenditures of companies in different media). The respondent argued that the company currently is not engaging in big data analytics activities with the remark that big data analytics means to "crawl the Internet, get everything that we [i.e. Company C] can use and place it inside our own systems". In other words, their business intelligence initiatives focus mainly on data, generated by internally the organization.

Concrete business needs for using external data were identified during the course of the interview. Data from social media could be used for finding out the needs and wants of newspaper subscribers and providing them with more tailored news content. In addition, public data from corporate websites could be used for building profiles of Dutch companies that are richer than the ones the group currently has. Finally, deriving meta-data from recorded radio interviews was also said to be of high importance for the company.

Most assumptions underlying the research problem were verified based on the input of the respondent. According to him, the company does not have the necessary resources (mainly human resources) to build on-premise big data analytics solutions. In his word, "what we do is we listen to the business, and they come to us with questions, we collect data and we put it together in a single database… That's about it. We could not build a system that looks outside and brings all that data together". Therefore, assumption 1 can be verified. Assumption 2 was also verified. The respondent indicated that he is fine with using cloud-enabled services for business analytics as long as the data and services are available, secure, and cost effective. Assumption 3 was rejected in the case of Company C. The respondent indicated that there are professionals in his team who possess the competencies necessary for drawing insights from big data. Assumption 4 was also rejected as the respondent shared that occasionally they hire business analytics consultants. However, his perception on the cost of external consultants is that they are "very expensive". On several occasions, after the big data SaaS platform was described to the respondent, he indicated that cost effectiveness is most important in his view. That is, if it is financially more attractive for Company C to run big data analyses on a SaaS platform than in an on-premise fashion, then they would most probably do so (depending on final decision from the IT manager). In addition, having the data secure and not disclosed to other parties were also indicated as important prerequisites for Company C using the SaaS platform's services. However, the final assumption is still rejected as connecting company C to data scientists would not add a lot of value for the company.

As a conclusion, it could be argued that the envisioned SaaS platform can bring value to Company C but on a more limited scale. The business analytics experts that the company already employs could take the role of the data scientists (supply side) in the business ecosystem. In this way, the platform will add value to Company C by providing the necessary infrastructure and big data analytics services running on it (as the company cannot develop them themselves). Finally, cost effectiveness, availability, and data privacy and security were indicated as characteristics that would be used by Company C upon adoption decision.

Company D

Company D is the smallest company in our convenience sample with 25 employees. The organization operates loyalty management programs for their customers, some of which are major shareholders of the company.

The respondent indicated that he is responsible for "the database marketing team and customer insights". His team draws insights from transaction data as well as insights from research among purchasers (e.g. customer experience, customer satisfaction). He regards big data as large volumes of data that come from a variety sources. For the case of Company D, transaction data that is being generated daily from purchasers within loyalty programs and periodical digital communication with them through e-mail and websites are two examples of data that is being analyzed. Furthermore, the respondent indicated that they do not use real-time positioning data yet (based on purchaser's location) and that in his opinion "real-time is where the real big data is going to come" for influencing customer behavior.

It can be argued that Company D are already using big data for analytics purposes as they analyze different types of data as described above. However, we can make a distinction between data generated internally by the business (e.g. purchase transactions data, research among purchasers) and external unstructured data from social media, or other websites and APIs. The respondent indicated with certainty that they can see business needs for incorporating external unstructured data in their processes. The following scenario was provided as an example: one individual buys a product from the store of a customer of Company D and then tweets about their new purchase. It would be of value for Company D to link that tweet to a particular profile in their database. In this way information on the person's actions and interest in the product (sentiment analysis) can be combined and further steps (not specified by the respondent) could be taken. In another example scenario, some stores have customers who do not come to the shop regularly but when they do, they make big purchases. It would be of value for Company D if they could determine in real-time if such purchaser is about to come to a store based on his location for instance.

Later in the course of the interview, the respondent indicated that their shareholders (and customers at the same time) could provide them with resources needed for building or buying on-premise analytics solutions as long as there is a viable business case. Thus, assumption 1 is rejected. Related to assumption 2, the respondent indicated that at this moment it is very interesting for Company D to consider what they could do in the cloud as they are currently discussing storage of traditional data, its limitations and costs. Consequently, we can verify assumption 2. Furthermore, the respondent indicated that the company has access to a "flexible pool of externals" from where competencies that are necessary to draw insights from big data can be obtained. Therefore, assumption 3 is rejected. Assumption 4 is also rejected as the respondent indicated that they might hire external consultancy company, but they would most probably use the expertise of the people in the aforementioned pool of externals. Finally, for the same reason, assumption 5 can also be rejected. That is, the platform cannot add much value to Company D by connecting them to data scientists within the platform's business ecosystem.

In conclusion, we can argue that the envisioned big data analytics platform can add value to Company D on a limited scale. That is, the company does not recognize a need for collaborating with data scientists. However, in the presence of a viable business case, Company D would consider using services related to big data collection, processing, and exporting/analytics.

F

Second Round Interviews Protocol

All interview with respondents from the two user groups were conducted by following the protocol described in this appendix.

Introduction

- Background of the researcher
- Interview objective
- Permission to record the conversation
- Interview structure
- Background of the respondent

Solution Description

Description of the main functionalities of the big data SaaS platform

Pilot Services Description

Description of the two pilot services in order to reduce the level of abstractness of the designed artifact by providing context to the conversation

User requirements elicitation

First, we present the questions that were asked to the representatives of SMEs in the retail industry. After that, we present the questions asked to the data scientists.

SME-specific questions

<u>User interface</u>

- Profile information
    - Do you want to be contacted by data scientists with ideas for new big data services that are relevant for your business?
    - What company information would you like to share with data scientists?
    - How do you prefer to be contacted in such cases?
    - What information should be stored in the system for internal purposes like invoicing for instance?
- Existing service discovery
    - How would you like to search for existing big data analytics services on the platform?
    - Would you like to try out paid services for free for a limited time period?
    - How long should this try out period be?
- Finding a match
    - How would you prefer to find data scientists on the platform?

    Tip: search in a directory with all data scientists and send them private message / post a public message in a discussion board and wait for data scientists to show interest / both
  - What information would you like to see in a data scientist's profile page if you want to contact them for a new project?
- Collaboration tools
  - If you were to develop a new big data service together with a data scientist, what collaboration tools would you require from the system?
- Homepage
  - What information/features would you like to have on your homepage?

<u>Costs</u>

- What pricing models would you like to have supported by the platform? *(monthly/yearly fee; pay-per-resources-used; pay-per-analyses-run; combination of them)*
- Would like to have different pricing models for different big data services?
- Would you like to see a per-service breakdown of costs for a billing period?
- Which billing methods would you use? (*iDeal, PayPal, etc.*)

<u>System (general)</u>

- Device accessibility
  - From where would you access the system? (*home, work, mobile*)
  - What devices would you use to access the system? (*desktop pc, laptop, smartphone, tablet*)
- Compatibility
  - Would you want the services on the system to be compatible with any current information systems that you're using now?
  - Are importing/exporting capabilities between the two systems required?
- Services availability
  - What times of the day and days of the week do you expect to be using the platform?
  - How often can you tolerate system outages during these times?
  - How long can an outage last if one does occur?
- Security (Note: *for SMEs only*)
  - What is your most sensitive data?
  - If you were to upload that data on the system for analytics purposes what would be your security requirements?
- Customer Support
  - What type of customer support would you require from the system?

Data scientist-specific questions

<u>User interface</u>

- Collaboration
  - If you were to build a data model for a particular company, in what ways would you want the platform to support your activities? (*use the weather analytics example*)
    - Collaboration tools?
    - Access to datasets?
    - Data model testing environment?
- Finding a match

    - Would you contact companies if you have an idea for a big data analytics service that could be built on the platform?
    - How would you prefer to find and contact companies?
- Profile information
    - Do you want to have a profile page which companies can see when searching for a data scientist?
    - What profile information would you share with companies?
- Homepage
    - What would you like to see on your homepage?

Revenues/Costs

- Is a per-usage model attractive for you when it comes to monetizing on data models you've developed? (*e.g. number of analyses per period)*
- At what time interval would you want to receive the funds accumulated to your account?
- How would you like to receive these funds?
- Would you want to see a per-service breakdown of the funds accumulated to your account?
- Are you willing to pay a monthly fee in order to access the customer pool of companies on the platform?

System (general)

- Device accessibility
    - From where would you access the system (*home, work, mobile*)?
    - What devices would you use to access the system? (*desktop pc, laptop, smartphone, tablet*)
- Compatibility
    - Would you want the services on the system to be compatible with any systems that you're using now?
    - Are importing/exporting capabilities between the two systems required?
- Services availability
    - What times of the day and days of the week do you expect to be using the platform?
    - How often can you tolerate system outages during these times?
    - How long can an outage last if one does occur?
- Security
    - Do you have any particular security requirements for the system?
- Customer Support
    - What type of customer support would you require from the system?

Finalization

- Can you think of other requirements that you have as a potential user of the system?
- Ask if the respondent if they are willing to participate in a prototype evaluation session (*not part of this thesis project*)
- Thank the respondent!

G

List of User Requirements

The table below provides a list of the identified end-user requirements per interview. In addition, overlaps and conflicts with user requirements from the other interviews are also indicated in the final four columns (DS1, DS2, SME1, and SME2).

ID	Requirement	Overlaps and Conflicts			
	Interview 1 (data scientist, DS1)	**DS1**	**DS2**	**SME1**	**SME2**
UR1	Data scientists should enter information about their domain of expertise; the techniques and methodologies they know; and the software packages they can use;		**UR14**	-	**UR35**
UR2	The platform should suggest potential projects to data scientists depending on their domain of expertise		-	-	-
UR3	The homepage of data scientists should contain a list of suggested project (companies) and the available datasets		**UR16**	-	-
UR4	Datasets provided by the platform should be full with as at least missing points as possible		-	-	-
UR5	The platform should export datasets to formats readable by statistical analysis software (at least SPSS and Stata)		**UR13**	-	-
UR6	The platform should provide SMEs and data scientists with a private messaging feature		-	-	**UR33**
UR7	The platform should transfer revenues to data scientists immediately after they are received from an SME		**UR19**	-	-
UR8	The platform should be available for desktop PCs and laptops		**UR20**	-	-
UR9	Data scientists should be paid on a per-hour basis		**UR17**		**UR37**
	Interview 2: (data scientist, DS2)	**DS1**	**DS2**	**SME1**	**SME2**
UR11	Data scientists should have access to more detailed information about SMEs besides contact information: what is their business model and what major issues they are facing	-		-	-
UR12	Data scientists should be able to see the types of internal data that SMEs are willing to upload to the system and the software packages that they use for managing these data	-		-	-
UR13	The platform should allow data scientists to export datasets in Excel and CSV	**UR5**		-	-
UR14	Data scientists should be able to enter information about their domain expertise, current availability, their contact information and at what times they can be contacted	**UR1**		-	**UR35**
UR15	The platform should have a rating system for data scientists based on SME ratings	-		-	**UR35**

UR16	The homepage of data scientists should contain a list with the newest projects in their domain of expertise and new messages for ongoing projects	**UR3**		-	-
UR17	Data scientists should be paid on a per-hour basis	**UR9**		-	**UR37**
UR18	New service development projects on the platform should have milestones at which SMEs should pay data scientists (through the platform) for the spent hours	-		-	-
UR19	The platform should transfer funds to data scientists after payment from SME by means of bank account transfer	**UR7**		-	-
UR20	The platform should be available for desktop PC and laptops for downloading datasets	**UR8**		-	-
UR21	The platform should be available for smartphones and tables for communication purposes	-		-	-
UR22	The platform should create a secure connection after user login	-		-	-
	Interview 3 (SME1)	**DS1**	**DS2**	**SME1**	**SME2**
UR23	Sensitive company data uploaded by SMEs must not be shared with third-parties	-	-		-
UR24	SMEs should be able to create a wish list for services that they want developed in the future	-	-		-
UR25	SMEs should be able to try paid services for free for a six month period	-	-		**UR34**
UR26	The homepage of SMEs should show the outputs of installed services	-	-		-
UR27	The platform should provide SMEs with an estimation of service costs	-	-		-
UR28	The platform should allow for automatic billing	-	-		-
UR29	The platform should be available for desktop PC and laptops	-	-		**UR39**
UR30	The platform should have a FAQ section	-	-		-
UR31	The platform should provide SMEs with support over the phone	-	-		-
	Interview 4 (SME2)	**DS1**	**DS2**	**SME1**	**SME2**
UR32	Uploaded data needs to be sanitized in order to provide additional security of sensitive data	-	-	-	
UR33	The platform should provide internal messaging feature for private communication between SMEs and data scientists	**UR6**	-	-	
UR34	Paid big data services should have a trial period of at least a month	**UR25**	-	-	
UR35	The platform should provide a list of data analysts containing a photo, skills, user reviews and ratings hourly fee, and types of analyses they can do	**UR1**	**UR14** **UR15**	-	
UR36	The platform should allow SMEs to upload their company data	-	-	-	
UR37	Data scientists should be paid on hourly basis	**UR9**	**UR17**	-	
UR38	The platform should comply with data privacy regulation in The Netherlands	-	-	-	
UR39	The platform should be accessible through a browser on a laptop or desktop PC	-	-	**UR29**	

Only one conflicting set of requirements was identified: UR3 and UR16. We have decided to resolve this conflict by discarding one of the two requirements when developing the system prototype. UR16 will be used for prototype development while UR3 will be ignored. Later, during prototype evaluation, specific attention will be paid on what should be placed on the homepage of data scientists.