The role of Big Data in the evolution of Platform based Ecosystems

A case study of an emerging platform-based ecosystem in the software engineering industry

Angelos Kostis
Abstract

Platform based ecosystems are becoming dominant models in the software engineering industry. ‘Big data’ has recently gained increased attention from both academia and practitioners and it is believed that big data affects every sector and industry. While an abundance of research focuses on big data and platform-based ecosystems, these two are typically approached as secluded spheres. This study aimed toward an investigation of big data’s role in the evolution of platform-based ecosystems in the software engineering industry. In the present thesis the influence of big data on the software engineering industry and more specifically, the impact of big data on the evolution of software ecosystems, is examined. A case study focused on a platform owner and pioneer in the software engineering industry has been conducted. This study identifies challenges and opportunities triggered by the advent of big data in context of platform-based ecosystems. Hence, considerable insight regarding the impact of big data on contemporary platform providers and the evolution of platform-centric ecosystems is gained. The findings illustrate that software ecosystems are affected by big data in a positive manner, but some identified challenges emerge and have to be tackled. Additionally, in this paper, it is suggested that both academia and practitioners should dig deeper into this relationship and identify how the evolution of platform-based ecosystems is impacted by the advent of big data.

Keywords: platform-based ecosystem, software ecosystem, big data
1. Introduction

The business landscape has recently been transformed dramatically due to the advent of various technological advancements. The emergence and explosion of cloud computing, social media, sensor technologies, wireless networks and an alarming increase of web traffic indicate that we are experiencing the “Age of the Internet of Things” (Brynjolfsson & McAfee, 2014; Chui, Löfler & Roberts, 2010). As a consequence, connectivity between devices and human actors has been improved and machine-to-machine communication has become a reality that takes place at an unprecedented rate. Everything is being changed incredibly fast and the up-to-date organizations all over the globe are being bombarded by tremendous amounts of digitized data. It is widely believed that this is only the beginning of a new age or in other words the tip of the iceberg. “There is no reason to believe that the waves of digitization are finished. In fact, what we have witnessed to date may well be just the early phases” (Holmström, 2013, p. 210).

Facing this constantly changing business environment requires innovative way of thinking. In order for an organization to survive in today’s competitive environment, a need for high-variety strategies adoption (Kahn, 1998) is emerged. Tiwana, Konynski and Bush (2010) raise the importance of ecosystems and underline that competition is shifting toward platform-centric ecosystems. Additionally, Cusumano (2011) advises the contemporary organizations to compete “on the platform within the industry and the ecosystem of competitors, partners and users”. It is widely believed that platform business model is a suitable structure for modern businesses in order to navigate the shaking earth of today’s business landscape on their feet (Gawer, 2015; Gawer & Cusumano, 2014; Tiwana et al., 2010; Cusumano, 2010; Eisenmann, Parker & Van Astyln, 2007; Beimborn, Miletzki and Wenzel, 2011). Following this idea, software ecosystems are gaining increased attention (Manikas and Hansen, 2013) and a growing number of contemporary firms adopt platform-thinking (Sawhney, 1998) and competitive principles (Bengtsson and Kock, 2000; Shahla and Ambra, 2012). What confirms the value of platform business models is that firms with breathtaking power like eBay, Google, Amazon, Twitter, Apple, Salesforce and Facebook are precisely based on platform ecosystems.

The software engineering industry is an illustrative example that confirms the aforementioned statement. Software development firms invest on platform-centric ecosystems to leverage the power of co-innovation (Janner, Schroth and Schmid, 2008) and the expertise of a diverse developer community (Tiwana et al., 2010). It is evident that in the software industry software ecosystems are gaining in popularity (Manikas and Hansen, 2013) and becoming the dominant model for software development (Tiwana et al., 2010). To this end, ecosystems such as Google Android, Firefox browser and Apple iOS show a great success.

While an abundance of research has paid close attention at various aspects related to platforms, the majority of the researchers underline that research activity and understanding of platforms’ emergence and evolution has a heavy and long road to climb (Cusumano, 2010; Holmström, 2013; Gawer, 2015; Tiwana et al., 2010). Embracing this opinion, Manikas and Hansen (2013) highlight that little research has focused on real-world ecosystems. Additionally, Holmström (2013) states that an abundance of opportunities and challenges regarding platform based innovation are understudied. The above-mentioned statements were the trigger point for my thesis.
According to Serebrenik and Mens (2015), the emergence of software ecosystems has led to an increase of digitized data. “Big Data” (McAfee & Brynjolfsson, 2012) seems to be a driver of challenges and opportunities (C&O hereafter) linked with software ecosystems (SECOs hereafter; acronym used by Jansen, Brinkkemper and Finkelstein, 2009). The introduction of the term “big data” highlights the fact that in contemporary times an unstoppable flow of vast amounts of structured and unstructured digitized data occurs (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Byers, 2011; McAfee & Brynjolfsson, 2012). It is underlined that every sector of the global economy has been affected by the advent of big data (Manyika et al., 2011; Mayer-Schönberger and Cukier, 2014). Big data is also gaining popularity in the software engineering industry (Anderson, 2015) and thus, in 2015, the first dedicated workshop (1st International Workshop on BIG Data SE, Firenze, Italy, May 23, 2015) took place.

Focusing on the software engineering industry, this thesis illustrates how big data affects the evolution of SECOs. In particular, the present thesis examines what C&O are brought on stage in context of SECOs due to the advent of big data. It is based on a single case study (Yin, 2009, p.19) focusing on a real world SECO, called CiRANO. Although, there are several actors that might be part of a SECO (Manikas and Hansen, 2013), the present thesis is focused on the C&O that have to be managed by the orchestrator (Jansen et al., 2009) or platform owner (Tiwana et al., 2010; van Angeren, Kabbedijk, and Popp, 2011) or else the SECO coordinator (Jansen and Cusumano, 2013). According to Jansen and Cusumano (2013), SECO coordinators are defined as “beneficiaries of SECO growth who have instruments available to influence the development of the platform or the surrounding ecosystem”. The SECO coordinator of the CiRANO SECO is an innovative company, called B-open, which operates in the high tech software engineering industry, offering software solutions as a platform provider. It was not until recently that B-open decided to open up their internal software development platform to external developers and created a SECO in order to expand platform’s expertise and capabilities. In what follows, this thesis explores what C&O are triggered by the advent of big data in association with CiRANO and have to be managed by the CiRANO coordinator.

1.1 Thesis Structure

Regarding the way this thesis is organized, the paper is following an IMRAD structure (Glasman-Deal, 2010). In particular, this manuscript is divided into five sections. The first part introduces the central aim of this study. Additionally, the research question as well as the expected contributions are described. Secondly, the ‘Related Research’ gives an overview of the existing literature and identifies a research gap. Thirdly, the ‘Methodology’ section describes the research approach, the method and the case of this study. Fourthly, in the ‘Results’ part, the empirical data are brought on stage. Fifthly, the ‘Discussion’ section connects the findings with previous research and the main implications of this study are highlighted. In addition, promising research avenues in context of big data and SECOs are presented.

1.2 Aim & Research Question

The central purpose of this thesis is to examine how the emergence and evolution of platform-based ecosystems are influenced by the advent of big data. After conducting a literature review and identifying a research gap, I came to the conclusion that the influence of Big Data on the
emergence and evolution of platform based ecosystems is understudied. Therefore, focusing on the software engineering industry, I aim to answer the following research question:

- What are the challenges and opportunities associated with the advent of big data in context of software ecosystems in the software engineering industry?

2. Related Research

In this section information about the main research streams related to platforms and big data are presented. What follows, includes definitions and related concepts of great importance that gained attention by the research community.

2.1 Research on Platforms

Three main streams can be identified in the platform literature: Product Platform (Robertson & Ulrich, 1998; Meyer and Lehnerd, 1997), Market Intermediary Platform (Caillaud & Jullien, 2003; Rochet & Tirole, 2002; Economides and Katsamakas, 2006) and Platform based Ecosystem (Cusumano & Gawer, 2002; Gawer & Henderson, 2007; Tiwana et al., 2010) or according to other papers, SECO (Messerschmitt and Szyperski, 2005; Bosch, 2009; Sandberg, Holmström and Lyttinen, 2013; Lungs, Lanza, Girba and Robbes, 2010). It is worthy to note that Thomas, Autio and Gann (2014) provide a systematic platform literature review and add one more research stream regarding a type of platform they call ‘organizational platform’. Regarding the first stream, a product platform results in a family of physical products and is linked with a single organization. In other words, a product platform is defined as “the collection of assets shared by a set of products” (Robertson and Ulrich, 1998, p. 20) or as “a set of common components, modules, or parts from which a stream of derivative products can be efficiently created and launched” (Meyer and Lehnerd, 1997, p.7). Secondly, the market platform unlike the product platform, is related to more than one firm; it actually “extends the concept of the product platform into markets operating through the platform” (Sandberg et al., 2013). Thirdly, the platform based ecosystem research stream gives another perspective to the notion of platform by presenting it as a “central point of control within a technology-based business system” (Thomas et al., 2014).

2.1.1 Research on Software Ecosystem

Let us now focus on the SECO stream and the main topics that have recently gained traction from the research community. In the digital era, the platform is inextricably linked with a software-based system. Hence, this stream is also presented as software ecosystem (Messerschmitt and Szyperski, 2005; Bosch, 2009; Sandberg et al., 2013; Lungs et al., 2010). SECOs are considered as subsets of digital ecosystems (Jansen and Cusumano, 2013). According to Messerschmitt and Szyperski (2005), a SECO “refers to a collection of software products that have some given degree of symbiotic relationships”.

Further, Tiwana et al. (2010) present an excellent state-of-the-art of platform ecosystems field and provide perfectly articulated definitions of the terms platform and platform-based ecosystem. “We define a software-based platform as the extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate” (Tiwana et al., 2010). Additionally,
according to Tiwana et al. (2010), the modules are “add-on software subsystems” that build on the top of a platform and add value to it. In turn, a platform-centric ecosystem is constituted by the platform and the modules connected to that.

Moreover, Lungs et al. (2010) define a SECO as “a collection of software projects which are developed and evolve together in the same environment”. Also, an alternative definition of what constitutes a SECO is given by Bosch (2009): “A SECO consists of a platform, products built on top of that platform and applications built on top of the platform that extend the products with functionality developed by external developers” (Appendix 1: SECO). All the above-mentioned definitions are widely adopted, but in this thesis a SECO is approached as defined by Jansen, Finkelstein, and Brinkkemper (2009):

A SECO is a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artefacts.

In this stream there has been shed light on issues related to platform leadership (Gawer & Cusumano, 2008), design and architectural issues (Bosch, 2010; Kazman, Gagliardi and Wood, 2012; Gawer & Cusumano, 2002; Baldwin & Woodard, 2008; Tiwana, 2008), the influence of industry-wide platforms on the innovation process (Gawer & Cusumano, 2014; Cusumano & Gawer, 2002; Holmström, 2013; Ahmad Ghazawneh, 2010;), the role of platform ecosystems for enterprise (Beimborn et al., 2011), the network effects (Tiwana et al., 2010), architectural control and ownership issues (Cusumano & Gawer, 2002; Gawer & Henderson, 2007) and the level of platform openness (Eisenmann, Parker & Van Alstyne, 2008; West, 2003; Boudreau, 2007; Gawer, 2015; Cusumano & Gawer, 2002). As far as platform openness is concerned, Eisenmann et al. (2008) examine thoroughly how, when and why managers should open or close their platforms. Moreover, the notion of generativity (Zittrain, 2006; Holmström, 2013, p. 204-211;) seems to be a topic of central significance inextricably linked with incremental innovation and the platforms’ evolution. Further, in the discussion pertinent to the evolution of platforms, there are some papers that focus on the coevolution of the platform architecture, governance, and environmental dynamics (Tiwana et al., 2010) and others that discuss what drives the decision of a platform owner to shift the platform’s boundaries (Gawer, 2015).

2.2 Defining Big Data

The Big Data movement is enormous nowadays. It is a buzzword that is being heard all the time when the discussion comes to innovation, problem solving, value creation, and every aspect of the contemporary business life. The Big Data research project of McKinsey Global Institute (2011) led by James Manyika and Michael Chui has been one of the most influential works for the Big Data research stream. In their paper, Manyika et al. (2011) present big data and raise its importance for both economy and society.

But, what is actually Big data? There is a myriad of definitions for this term, but one of the most widely accepted is the following: “Big data refers to datasets whose size is beyond the
ability of typical database software tools to capture, store, manage, and analyze” (Manyika et al., 2011). In addition, Rouse (2011) define big data as a “description of the voluminous amount of structured or semi-structured data a company creates or data that would take too much time to load into a relational database for analysis”. Further, big data is also approached as “extremely large sets of data related to consumer behavior, social network posts, geotagging [and] sensor outputs” (Johnson, 2012). Having an alternative perspective of what Big Data means for the contemporary organizations, McAfee and Brynjolfsson (2012) state: “Smart leaders across industries will see using big data for what it is: a management revolution”.

Moreover, in the literature related to the notion of Big Data, three main characteristics of big data are presented. These characteristics are widely acknowledged as the 3 V’s: volume, velocity and variety (Laney, 2001; Manyika et al., 2011; McAfee and Brynjolfsson, 2012; Gartner, 2012; Russom, 2011). Laney (2001) was the first one who talked about the 3V’s. In his publication “3D Data Management: Controlling Data Volume, Velocity, and Variety” he underlined: “E-commerce channels increase the depth/breadth of data available about a transaction (or any point of interaction)”. (Appendix 1: Big Data’s 3V’s)

### 2.2.1 Volume

By letting the numbers pertinent to today’s digital data speak for themselves, a terrific increase in the size of data that contemporary organizations have to access, store and analyze can be noticed. “As of 2012, about 2.5 exabytes of data are created each day, and that number is doubling every 40 months or so. More data cross the internet every second than were stored in the entire internet just 20 years ago” (McAfee & Brynjolffson, 2012). Additionally, Hortonworks’ CEO, Rob Bearden, in his keynote address at Hadoop Summit 2014 stated: “The data volume in the enterprise is going to grow 50x year-over-year between now and 2020. I think the most important thing to recognize is that 85% of that data is coming from net-new data sources.”

The figures presented above send us alarming messages and indicate not only an increase in the volume of contemporary data, but also a change in the way that data has been produced and distributed. The most easily recognized and obvious characteristic of big data is the overwhelming volume. Nevertheless, what makes big data a topic that generates the need for “the sexiest job of the 21st century” (Davenport & Patil, 2012) is not only the enormous data-pools, but also the characteristics of velocity and variety.

### 2.2.2 Velocity

Another big data’s characteristic is pertinent to the fact that data is generated in a breakneck speed (Manyika et al., 2011). In 2014, DOMO (https://www.domo.com/blog/2014/04/data-never-sleeps-2-0/) identified and quantified the amount of data that is created every minute. An infographic created by DOMO can be found in the Appendix 2: DOMO Infographic. The results are compelling and thus, it is significant to present some of these figures: Per minute, apple users download 48,000 apps, Instagram users post 216,000 new photos, WhatsApp users share 347,222 photos, Google receives over 4,000,000 search queries, Youtube users upload 72 hours of new video, Twitter users tweet 277,000 times and email users send 204,000,000 messages. All the previous numbers indicate a rapid data explosion that contemporary organizations are challenged to handle. The speed of data generation is
constantly increasing (Manyika et al., 2011) and the frequency of data flows into organizations is rocketed (Russom, 2011). Hence, having real-time information is more than valuable for the up-to-date firms. “Real-time or nearly real-time information makes it possible for a company to be much more agile than its competitors” (McAfee & Brynjolfsson, 2012).

The characteristics of volume and velocity have already been described. However, Big Data has gained so much attention because its complexity stems from the fact that today’s data is expanding at 3 fronts simultaneously (Laney, 2001; McAfee & Brynjolfsson, 2012).

2.2.3 Variety
The third characteristic of big data comes to the game to make the situation even more complex. Particularly, Big Data is characterized by great variety. Three issues are mainly linked with this characteristic. Firstly, contemporary data is generated by diverse data sources (Manyika et al., 2011). Secondly, a great number of big data sources are new (McAfee & Brynjolfsson, 2012). Thirdly and most importantly, the first two issues result in generation of inconsistent and multi-format data (Russom, 2011). “Big data is not just about giant data volumes; it’s also about an extraordinary diversity of data types, delivered at various speeds and frequencies” (Russom, 2011). Thus, problems linked with format inconsistency have emerged; In particular, data can be in structured as well as unstructured formats (McAfee & Brynjolfsson, 2012). For instance, big data refers to raw readings from sensors, images from satellites, voice recordings from cell phones, social network feeds, videos, GPS signals, traceable browser cookies, simple text files and pdf’s.

2.3 Gap in Research
To the best of my knowledge, the gap in platform literature is that scant attention has been paid towards big data issues. Specifically, there is not a clear connection between SECOs and big data. Also, research has not studied adequately real-world SECOs in software engineering (Manikas and Hansen, 2013). In addition, although there are some papers that address issues pertaining to software development in the big data era, a single study that connects SECOs with big data has not been identified. For instance, there are studies that explore challenges of large-scale software engineering (Nivoit, 2013) related to big data, but these studies are not focused on the structure of SECOs. For instance, although Bosch (2010) identifies architecture challenges for SECOs, these challenges are not linked with big data. Further, Jansen, Finkelstein and Brinkkemper (2009) identify challenges that SECOs are putting to software vendors, but they do not focus on how the SECOs themselves can be challenged by Big Data. Additionally, there are studies (Nagappan and Mirakhorli, 2015; Anderson, 2015) that discuss issues pertaining to Big Data in software engineering, but the notion of SECOs does not appear in their discussion. The only connection found is a study by (Serebrenik and Mens, 2015) who identify big data challenges in SECOs research, but these challenges are neither presented in detail, nor associated with SECOs themselves, but with challenges related to research on SECOs.

Therefore, I sense the research gap in the connection of SECOs with big data. Following this idea, Big Data is approached as a “management revolution” (McAfee and Brynjolfsson, 2012), since it can trigger a plethora of C&O for platform based ecosystems. Hence, the present thesis examines C&O that are triggered by big data in context of a SECO.
3. Methodology

In this part the research approach, the data collection and analysis method are described.

3.1 Research Approach and Method

After identifying the central aim and the research question that had to be answered in order to fulfil the main purpose, it was vital to select the most appropriate research method for the particular thesis. Therefore, it had to be considered what kind of data was required, what kind of information was needed and where this data could be traced. Keeping in mind that this study aims to examine the impact of Big Data on a platform ecosystem, it was a remarkably reasonable decision to adopt a qualitative research approach. The qualitative research methodology seeks to explore phenomena and answer questions like “how”, “what” and “why”. It helps theory building and provides an in-depth understanding of the phenomenon under discussion (Bryman, 2012). On the contrary, the common goals of quantitative research methodology are to quantify, generalize and test a hypothesis (Bryman, 2012), which are not even remotely close to this thesis’s objectives.

Taking into account that this thesis aims to explore C&O that are triggered by the advent of big data in context of SECOs, it has been illustrated that a qualitative single case study (Yin, 2009 p.19) of a platform owner suited perfectly with the purpose of this thesis. Moreover, in a SECO, different actors face different C&O (Manikas and Hansen, 2013). In order for the approach of the study to be crystal clear, it is significant to highlight that the present thesis focuses on C&O that have to be managed by a SECO coordinator (B-open in the particular case) and neither by external developers nor by end-users. Also, C&O are identified through a lens that examines aspects of both technical and business nature. By that way, questions regarding what C&O are triggered by big data for a SECO coordinator, why big data does matter and how a platform owner approaches these issues can be answered.

Furthermore, in order to ensure that our initial thought about the research method was right, it had to be taken into consideration what differentiates the case study from other research methods that could be used.

*What distinguishes the different methods is not a hierarchy but three important conditions [...] which consist of (a) the type of research question posed, (b) the extent of control an investigator has over actual behavioral events, and (c) the degree of focus on contemporary as opposed to historical events. (Yin, 2009 p.8)*

Regarding the first condition, the research question of this thesis is a “what” question, which is characterized as exploratory “likely to lead to the use of case studies, histories and experiments as preferred research methods” (Yin, 2009 p.9). As far as the rest conditions are concerned, this study focuses on contemporary issues and does not require control over events.

All the previous arguments militate in saying that a qualitative research centered on a single case study and in-depth semi structured interviews would be ideal for the present thesis. What strengthens even more our decision to conduct a qualitative case study research is that case studies have been used by a plethora of other studies in the field of Informatics. For instance, Cusumano and Gawer (2002) discuss topics related to platform leadership by examining how Intel, Microsoft and Cisco drive industry innovation. Additionally, Gawer and Henderson
an features.

of industry, to development B particular, employees

3.1.1 Research Case: B-open and the CiRANO Software Ecosystem

This section provides information about the case study. B-open (http://www.b-open.gr) is a software engineering enterprise founded in 2004 and the official headquarters of the firm are located in Thessaloniki, Makedonia, Greece. At the moment, the firm houses more than 15 employees and the firm’s future seems bright.

B-open offers innovative IT solutions for enterprises as well as for the public sector. In particular, the enterprise innovates through jPlatform, which according to firm’s CTO manager is “a modern design environment for the development of open distributed enterprise applications”. jPlaton started to be built in 1999 and has been in the market since 2004. The main purpose of the platform is to support all the software development phases and this is why B-open presents jPlaton as a life cycle development platform. “jPlaton is not another web editor; it is a complete application development platform, facilitating each phase of the development process from modelling, composition, and configuration to deployment and expansion of applications” (CTO). It is worthy to note that jPlaton is one of the top 10 PaaS platforms evaluated by 4CaaS in the “Analysis of the State of the Art” (2010) for the current PaaS solutions. After a couple of years, B-open decided to move jPlaton to the cloud in order to build an even stronger position as an open cloud computing pioneer.

In 2014, with the vision of becoming a world-class player of the software engineering industry, B-open decided to expand jPlaton’s boundaries and allow external programmers to use the platform’s features. The internal platform acted as a conceptual base for establishment of a SECO, called CIrANO. B-open became a SECO coordinator (Jansen and Cusumano, 2013) and external actors can now develop software applications in the cloud using the platform’s features. Also, external programmers can build on the top of the platform and add value to the ecosystem. B-open has also created their own add-on store, named jAgora. In order to gain insight into the characteristics that define CIrANO ecosystem, the classification model for SECOs provided by Jansen and Cusumano (2013) has been used. According to this model, the CIrANO SECO can be described as follows: CIrANO is based on a software service platform and is coordinated by a privately owned entity with an extension market to which participants can submit extensions for free or after making a payment (this is something that is still under consideration).

What makes B-open a highly suitable case for conducting such a research is that CIrANO is not only a platform for developing applications in the cloud for the cloud. CIrANO opens up an abundance of fruitful avenues for B-open to create value by opening the platform to external users and expanding jPlaton’s capabilities. By examining the CIrANO ecosystem, this thesis aims toward an understanding of the C&O that big data brings on stage with regard to SECOs. Last but not least, B-open has a sufficient philosophy regarding research, something that is illustrated by the fact that they have accepted more than 35 internships since 2004.
Before describing the data collection and analysis method, it is significant to present the components of this emerging SECO (Appendix 3: CiRANO SECO). jPlaton platform is the core. Through jPlaton an abundance of products have been developed. For instance, Comidor, Open business, jKteo and Bioassist are the basic products that built on top of the platform. By establishing CiRANO, B-open opens the gates to external developers in order for the latter to develop applications that extent the existing or develop entirely new products.

3.2 Data Collection

Regarding data collection, it was a multistep process constituted by: (i) ten in-depth semi-structured interviews (Creswell, 2013) conducted in three separate phases, (ii) a couple of seminars offered by b-open and (iii) an additional telephone communication with a B-open’s employee for further questions and remarks.

Before the interviews, attendance in two seminars was extremely helpful and almost required in order to get familiar with the philosophy of B-open. In these seminars, the researcher obtained an overview of the platform’s capabilities and advantages. The first seminar took place on 5th of July, 2015 in the headquarters of B-open in Thessaloniki. It was the first meeting with the people of B-open and basically, the CTO Manager gave a short but comprehensive interactive presentation of jPlaton. This first presentation acted as a pilot procedure for the upcoming process.

The second seminar was an official presentation of the CiRANO SECO and took place on July 24th, 2015 in Electra Palace Hotel, Makedonia II hall, Aristotelous Square, Thessaloniki. As far as the agenda of the seminar is concerned, there was a short introduction and presentation of B-open by the CEO of the firm, a live and interactive presentation of CiRANO by the CTO manager and a presentation by a sales manager and a software engineer regarding two software products of jPlaton, the Comidor Business Application Suite and the Bioassist. Further, two software engineers of B-open gave a live presentation of programming with CiRANO. Additionally, in both seminars some notes were taken in order for the researcher to review them later and prepare the interview guide appropriately. The interview guide can be found in the Appendix 5: Interview Guide.

All the interviews took place at the headquarters of the firm and during working hours. Conducting the interviews at the offices of the firm made the interviewees feeling more comfortable. It seems that the workplace has not affected their behavior, since they answered the questions in a natural way and their body language is judged as positive. The interviews were conducted in three stages. Although B-open’s people had no problem to conduct the interviews in english, we preferred to complement the interviews in greek. Greek is the mother tongue of the respondents and undoubtedly, it was more comfortable for them to be interviewed in their mother tongue. Additionally, after asking for respondents’ permission, all the interviews were recorded in order for the researcher to be able to transcribe later the audio. The interviews ranged from 30 to 52 minutes and the average duration of all the interviews was 39 minutes. The answers were uncomplicated, based on logical arguments and easy to understand. A table that presents information about the respondents, such as the positions they hold in B-open, their main duties, the years they work in the firm and their opinion about b-open can be found in the Appendix 4: Respondents data).
During the first phase, on 28th of July, four interviews were conducted with managers from the top and medium management level. Specifically, in this phase the respondents were the CEO manager, the Vice President, the CTO manager and a System-level Software Engineer. The second phase occurred on 6th of August, almost ten days later and after the analysis of the first set of interviews. This second phase was basically referred to programmers and marketers of B-open. The programmers of b-open are responsible for encoding information into a programming language and for developing software applications and modules. Additionally, they are responsible not only to develop and test that these applications run appropriately, but also to maintain them and ensure that any bugs are fixed immediately. Further, with the goal of achieving mutual beneficial relationships with a stable customer base, the marketers are focused on collecting data for potential and existing customers and executing marketing plans. The third phase of interviews took place on 23rd of October. In this phase, three more interviews have been conducted. The respondents have been a marketer, a programmer and once again, the CTO manager. The interview with the CTO manager was a complementary one, since we decided together to have a more detailed discussion about the C&O that big data brings on stage for software development and in turn, for SECOs. Of course, in this interview a lot of questions have not been posed again, since the main purpose was to dig deeper into the questions 18 to 23 of the interview guide, plus to discuss the question of how big data changes the way that software is developed.

The interview guide remained the same, but some questions were not posed to programmers and marketers, because questions pertaining to b-open’s vision, philosophy, future plans and platform governance could not be answered by them. On the contrary, programmers gave significant insights into existing processes, platform’s architecture and technical aspects. Also, some and them discussed how they feel about the establishment of the CiRANO ecosystem and the necessary collaboration with external developers. Further, the marketers gave me a lot of fruitful information regarding the difficulties of identifying appropriate data for customer understanding and issues related to data sources and collection of data. According to Sbaraini, Carter, Evans and Blinkhorn (2011), “By modifying the questions asked in data collection, the researchers fill gaps, clarify uncertainties, test their interpretations, and build their emerging theory”. Therefore, in the third set of interviews, questions pertaining to the way that software development is affected by the advent of big data have been added. From the analysis of the existing raw data, it has been highlighted that since the purpose of the SECO under examination is to develop software, it was vital to see also how big data challenges the way software is developed by CiRANO and in turn, how CiRANO itself is affected by big data.

3.3 Data Analysis

Regarding data analysis, a grounded theory approach (Strauss & Corbin, 1990; Miles & Huberman, 1994) has been adopted. Specifically, the researcher followed the guidelines provided by Strauss & Corbin (1990) and the analysis was based on open, axial and selective types of coding (Strauss & Corbin, 1990). “Coding is the pivotal link between collecting data and developing an emergent theory to explain these data. Through coding, you define what is happening in the data and begin to grapple with what it means” (Charmaz, 2006).
In a grounded theory study, data analysis takes place at the same time with data collection (Strauss and Corbin, 1990). Hence, the interviews were conducted in three stages to allow for intermittent data analysis (Sbaraini, Carter, Evans and Blinkhorn, 2011) and ensure that the gathered data are enough for theoretical saturation to be reached (Strauss and Corbin, 1990). Our decision to opt for a grounded approach stance from the fact that a grounded approach is best fitted for complex and uncertain research context (e.g., big data and SECOs).

Before the interviews, a review of the notes from the seminars took place. The most significant aspects of this case were identified and the construction of the interview guide drew on them as well. The interview guide has been created in such a way to categorize the questions and divide the interviews into parts facilitating at a certain degree the forthcoming analysis (Miles & Huberman, 1994). Re-listening to the taped interviews can be considered as part of the analysis, since valuable elements can be noted (Psathas and Anderson, 1990). Thus, after conducting the interviews, the audio files were listened and significant aspects related to the collected data were highlighted. In order to render respondents’ answers in an actionable format, word by word transcription occurred and transcripts were written in the same language with the conducted interviews. The three level coding process is described as follows. In Appendix 6, a table that gives an example of the coding process can be found.

3.3.1 Open Coding
Firstly, the researcher read through all the raw data several times and went through all the transcripts line by line to find some initially visible patterns. The questions at this stage were abstract and open. The goal was to get an overview of the respondents’ opinion and the researcher highlighted what initially considered as important. Hence, the researcher labeled relevant phrases in an abstract way and as a result, a list of 239 open codes was created.

3.3.2 Axial Coding
Afterwards, a second thorough examination of the transcripts, the collected documents as well as the initially formulated codes took place. The goal at this stage was to identify relationships between different codes and create a list of axial codes based on the identified open codes. This had a great implication on the analysis process, since a lot of codes were relevant and some relationships became obvious. Hence, the list of axial codes had notable differences from the previous list and the number of the initial 239 codes was limited to 27 codes (e.g., integration challenges, software challenges, hardware challenges, opportunity for business automation, predicting bottlenecks and projecting needs related to resources, achieving increased customization, opportunity for better customer understanding, behavioral analysis and network effects, efficient group targeting, facilitating innovation).

3.3.3 Selective coding
According to Strauss and Corbin (1998), “Selective coding is the process of integrating and refining categories”. Therefore, in the last phase of coding, the researcher went through all the codes generated in the previous steps, searched for deeper commonalities and differences between them and evaluated the already created families of codes. That was the most challenging type of coding, since identifying deeper relationships between the existing codes was a process of great difficulty. Finally, a lot of codes were dropped, due to the fact that new

13
codes were created by the combination of relevant codes. In particular, the constant interplay between data and codes resulted in a consolidation of relevant codes and creation of axial and selective codes, based on the open ones. In response to that, the number of 27 axial codes was decreased to 9 selective codes (e.g., IT infrastructure challenges, technical challenges, strategic challenges, talent and analytical challenges, security challenges, platform architecture challenges, marketing opportunities, opportunities for operational agility, strategic opportunities). Most significantly, a deeper understanding of the under discussion empirical data was achieved and “theoretical saturation was reached” (Holmström & Sawyer, 2010).

3.4 Ethical Considerations

While conducting a qualitative study, the researcher is required to take into account a number of significant ethical aspects such as privacy issues, anonymity and confidentiality (Bryman, 2012). In the Belmont Report (1979), it is underlined that the basic ethical principles are: 1) Respect for persons, 2) Beneficence and 3) Justice. In other words, anonymity should be under protection, respondents should not be harmed and there should be a fair balance between benefits and risks. “Whenever we conduct research on people, the well-being of research participants must be our top priority. The research question is always of secondary importance” (Mack and Woodsong, 2005).

I followed an informed consent process (Mack and Woodsong, 2005). In particular, two weeks before the interviews, a pre-meeting with the potential respondents took place. During that meeting, I gave them some information regarding the interview process as well as my perspective on ethical issues. It became clear that participation in the study should be voluntary and none of the respondents was going to be forced to participate. Moreover, I guaranteed anonymity of the respondents and protection of their privacy rights. Further, it was worthy to note that only if permission was given, would I be able to record the interview. Any interviewee could refuse to answer any question or even to stop the interview at any time.

All the aforementioned were also discussed before the beginning of each interview and the level of transparency was increased. Last but not least, the information gathered has been treated confidentially and used only for the purposes of this study.

3.5 Method Discussion and Critique

The process of identifying the strengths and limitations of a research method is essential in order for a researcher to ensure that she/he has an in-depth understanding of the method. Only by that way, can the researcher select the most appropriate research method and apply it in a correct manner.

As stated above, the central aim of this study and the nature of the research questions militate in saying that the selected research method is the most appropriate for the present thesis. Moreover, I underlined the fact that the most of research in the field of Informatics is based on qualitative research methods through case studies and in-depth interviews. Regarding the type of interviews, complementation of face-to-face semi-structured interviews was the most appropriate way to collect empirical data because it was possible to identify body language signals as well as emotions of the respondents. Last but not least, the decision to
adopt a grounded theory method, as presented by Strauss and Corbin (1990) is well justified, since this approach is ideal for studying new, complex and emerging phenomena.

However, the qualitative research method entails a couple of limitations. Firstly, it is not possible to make predictions and more importantly, it is not possible to quantify them (Atieno, 2009). Secondly, the pre-understanding of the researcher can more likely have an impact on the results of the study (Poindexter, 2002). Thirdly, it is not easy for the research findings to be generalized (Gable, 1994). As far as these limitations are concerned, all of them have been taken into account. Also, the purpose of this study is neither to focus on quantitative predictions nor on generalization of knowledge, but rather the present thesis aims towards a better and contextualized understanding of the power of big data in context of software based ecosystems through a thorough case study.

Additionally, drawing on the wisdom of learning from my own mistakes and omissions with respect to study design, it is important to mention that use of software tools that can increase the level of research efficiency is encouraged. For instance, a voice recognition software (VRS) could speed up the transcription process and could give me more time to focus on the actual analysis of the transcribed data. “VRS is computer software that automatically transcribes digital voice recordings without the need for typing” (Matheson, 2007). Moreover, use of software tools for data analysis, such as Atlas.ti (Holmström & Sawyer, 2010) could facilitate data analysis process.

4. Empirical Results - Findings

In this section, a presentation of the research findings takes place.

4.1 Marketing Opportunities

4.1.1 Opportunity for Better Customer Understanding

All the respondents, without an exception, stated that big data offers to the focal firm of a SECO opportunities for gathering real-time information about both end-users and external developers of the ecosystem. It seems that big data acts as a catalyst for understanding the customer needs and expectations. The CTO explains: “Big data offers us the opportunity for gaining real-time customer insight...[...]. By customers, I’m not talking only about the end users of our applications, but also the external developers that will connect to our SECO” (CTO). Also, a marketer highlights:

From my perspective, big data stands for understanding customer needs rapidly. All these numbers... all these figures... all these articles about big data are in place, because with big data we can understand better the needs and expectations of our customers and translate them into products or services. (MA)

4.1.2 Opportunity for Behavioral Analysis & Network Effects

One promising attitude of big data seems to be the capability given to the focal firm to collect valuable data for the behavior of potential members of the ecosystem as well as for their preferences. Subsequently, the focal firm can analyze their behavior, offer effective motives and meet a number of preconditions to build strong network effects. A marketer states:
Creating network effects is the main issue for a SECO. You need people in order to have a powerful ecosystem. The motivations of these people are unknown and until recently difficult to be discovered. Nevertheless, with big data analytics we can sense patterns in data regarding the motivations of external developers to enter CiRANO. We can now predict behaviors and reflect on these predictions appropriately. By that way, we can establish an ecosystem characterized by strong network effects and increased grow-ability. (MA)

4.1.3 Opportunity for Long-term Customer Retention
Further, it seems that big data acts also as a precursor for highly customized products, increased customer satisfaction and long-term customer retention. A marketer explains:

Three things can make an ecosystem successful in the long-run: [1] expectations and needs of developers’ to be met in order for them to have the will to participate in the ecosystem, [2] end users’ expectations and needs to be met in order for them to have the will to buy staff (participate in the ecosystem) and [3] internal resources to be managed in such a way that the previous two are most welcome. From my knowledge, the first two can be turned into reality with big data. When we can understand the customer needs, analyze his [/her] behavior in real-time and predict his [/her] intent rapidly, then it is very easy to create customized products and achieve increased customer satisfaction. As a result, increased customer retention rates can be ensured. (MA)

4.1.4 Opportunity for Efficient Group Targeting
Moreover, an opportunity identified in regards with big data is pertinent to more efficient group targeting. The respondents highlighted with a large degree of consensus that big data can be a great tool for identifying market trends and consumer intents. In response, a SECO can focus on more promising market segments. The CTO Manager explains:

It is easy to find people that are interested in our product and follow us on twitter or become friends with us on Facebook. However, it is extremely difficult to find such a target group that could satisfy us completely. It is difficult to find this group of people that from ‘leads level 1’ will become ‘leads level 2’, ‘leads level 3’ etc... and who finally will become customers. I see that big data as a great vehicle for better group targeting. (CTO)

Additionally, a marketer, responsible for b-open’s marketing campaigns, underlines:

As we said before, one of the big data magic things is that it can shed light on the intent of customers in real-time. Their motivations to buy our software products [end users] or participate in CiRANO [external developers] are becoming clearer and our target group more obvious. (MA)

4.1.5 Opportunity for Market Expansion
An opportunity triggered by big data in context of SECOs is also related to new tools and applications that have to be developed for firms that are interested in big data. A developer points out:
The reason why a SECO is built, is mainly because it is a structure that allows us [the developers] to produce software in an alternative, more collaborative and open way. Big data requires specialized software. We produce software and therefore, an opportunity that emerges for us is to expand our offerings and provide firms that are in need of big data analytical tools with such specialized applications and solutions. (PB)

4.2 Opportunities for Operational Agility

4.2.1 Opportunity for Customer-driven Software Development and Increased Customization

Some respondents underlined that huge amounts of high frequency data regarding customers are valuable sources for a SECO in terms of software engineering. It seems that engaging end-users of the applications is easier with big data. The Manager explains:

Another big mistake is being made in regards with software development for enterprises... The majority of business software applications are built by developers who do not have enough knowledge about business, are chosen by managers who are not users of these applications and finally, are used by people who do not participate neither in the design, nor in the development process. In B-open, we strongly believe that one valuable source of information is the end-user. The data of the end user should be guidelines for software development. And big data can give us the keys for the gates of these sources’ gates. (CEO)

Furthermore, following a customer-driven software development can result in shortening the software development lifecycle. The CTO manager highlights: “Involving the needs and expectations of our customers into our production process is of vital significance, not only because we increase the quality of the processes’ outcome, but also because we eliminate time-consuming processes and achieve shorter time to market.” (CEO)

4.2.2 Opportunity for Data-Driven Software Development

In the interviews, it has been stated that with big data ecosystems can automate their business processes. The CEO highlighted: “I see big data mainly as an opportunity to identify efficient processes and create patterns that can lead to increased business automation” (CEO).

In addition, it seems that the way software is developed today can become more data-driven. Hence, in the near future, the software development processes can be characterized by increased automation. A 7-years experienced programmer underlines:

If we take the most of big data, we will be at a position to develop software in a completely different way. In comparison with the way we develop software today, all tasks related to software development can become data-driven. By that, I mean that an ecosystem of developers can follow certain paths for certain situations under certain conditions. We can collect data pertaining to software development tasks in a systematic way. Wouldn’t it be great to have automatic processes for design and construction of our IT solutions? This can be achieved with big data analytical tools. It sounds frigging promising (PA)
Further, according to the majority of the respondents, big data holds a potential for predicting bottlenecks linked with software development projects. The CTO states:

Right now there is not a big data analytical tool that is precisely focused on how software is developed. For example, there is not a big data analytical tool that gathers information about the software development processes and can predict in advance a bottleneck [likely to be surfaced during the development process]. However, theoretically, there is a great potential for SECOs to be fed by big data with insights into how the application can be produced faster and more smoothly. With big data, we could put under examination hundreds of thousands of software development projects and find patterns in data pertaining to best practices, forms of data bases etc. (CTO)

As stated above, data-driven software development can become true if a lot of information about engineering processes is gathered. Another aspect raised by a less experienced programmer is the one of ‘code cloning’ and ‘repositories mining’. In particular, he claims:

In the near future, I see a great opportunity for developers participating in SECOs. Be careful, we talk about SECOs [SECOs], not open source. There is this issue... So many code lines in so many repositories. The previous generations should understand that code cloning is beneficial. Because of increased interconnectivity and the advent of big data, repositories are becoming transparent. This is an opportunity given by big data. I mean... now, due to the evolution of big data, people participating in such ecosystems will be able to mine repositories all-over-internet and can have access to lines of codes that we did not even know their existence in the past. (PB)

Last but not least, the same respondent mentioned a very interesting aspect linked with sustainability. Particularly, he says:

The above repositories-mining and code-cloning aspects will also have a great impact on ensuring environmental sustainability and creating ‘green’ SECOs. Fewer lines of codes will be written, the concept of duplication will be avoided, time consuming processes will be minimized and waste of resources will be decreased as well. (PB)

4.3 Strategic Opportunities

4.3.1 Opportunity for Product and Business Model Innovation

It seems that the advent of big data is a sound of opportunity knocking for product and business model innovation. All the interviewees highlighted the fact that big data facilitates improvement of innovation capabilities. More specifically, a marketer mentions: “Understanding customer needs and more importantly, predicting customer expectations are two vital steps to create innovative products. This is why big data facilitates innovation.” (MB)

Also, complementary innovation is taking place since SECOs are based on the notion of generativity. For instance, the external programming communities that will connect to CiRANO, will take advantage of the ready components of the platform and will be able to focus
on the business logic of their application. At the same time, all these external members will create “add-on software subsystems” that build on the top of the platform and add value to it. The CTO explains: “We need external developers to use CiRANO, use our ready components, focus on the business logic, develop their own applications and share them through jAgora [app repository]. Here the significance of generativity surfaces” (CTO).

Additionally, SECOs collect, store and hold significant data that can become key resources for novel business models. The opportunity for data-as-a-service business model is surfaced, since SECOs can also sell all this information and change the way they create value. The CTO Manager explains:

A platform ecosystem provider collects data regarding various aspects related to the ecosystem. When an ecosystem grows, the amount of data under discussion grows as well. Sooner or later, the value of this data will gradually become more obvious. At some point, platform ecosystem providers will sense opportunities for establishment of novel business models that could monetize data pertaining to software analytics. Data-as-a-service is always an option when it comes to big data and this is the case with SECOs too. (CTO)

4.3.2 Opportunity for data-driven decision making

Only the CTO highlighted that big data can have a great impact on decision making.

With so much data out there we can have a good insight on the performance of our organization, on how to treat people within an ecosystem, how to manage internal and external resources, how to make decisions regarding architectural choices and how to solve issues related to platform governance rapidly. Generally, big data can be a good tool for data-driven decision making. (CTO)

4.4 IT Infrastructure and Technical Challenges

Further, the interviewees underlined that the advent of big data brought on stage a couple of technological challenges for SECOs in context of both software and hardware. Hence, SECOs are challenged to invest on IT infrastructure in order to be able to access, store and mine value from big data in real time. The CTO Manager explains:

In order to be able to access, analyze and visualize this massive explosion of data efficiently, we are challenged to develop new software solutions. [...] But, it is not only the software. The performance of the existing hardware systems is also judged as inefficient for managing big data issues (CTO)

Following this notion, a software engineer highlights the fact that although real-time analysis is crucial for taking advantage of big data, SECOs lack of required computing power as well as effective software solutions. Thus, investments on IT Infrastructure are suggested as well. In particular, he points out:

Managing big data and making most of it means that we have to be at a position to work as closer to real-time as possible. This raises concerns regarding the way we access, store, analyze and visualize our data. In other words, big data questions our software as well as our hardware capabilities. Needless to say that
it is of great importance to produce software precisely based on big data and improve the existing computing power in order to be able to process data in a timely manner. (SE)

4.4.1 Software Challenges

Regarding the software, a platform based ecosystem is challenged to tackle the problem of data scalability in order for its developers to be able to build applications that are big data-friendly. A programmer explains:

The software development industry must find an answer to the following question: What kind of development techniques can the members of a SECO use in order to build their data in such a way that the ecosystem could be scalable in big data? Unquestionably, an ecosystem should create applications that are designed in such a way to be able to ‘host’ big data. (PA)

Additionally, the programmer claims that SECOs are challenged to provide their members with necessary tools in order for them to be able to build analytical applications specialized for big data. “Also, as software developers we have to come up with new analytics algorithms and develop software analytical tools focused on big data. This is more challenging than developing a simple application” (PA).

Moreover, it seems that real-time management of big data requires a move from the until recently used relational databases (that store data in rows and columns), towards non-relational databases (databases that do not store data in tables), such as NoSQL. The CTO explains:

Real-time access and analysis of big data cannot happen with the existing relational databases. The main issue is that the relational databases (MySQL Oracle etc) that are used for highly structured data cannot meet the requirements of these huge, heterogeneous and unstructured datasets. SECOs are challenged to come up with new database models that can result in achievement of high scalability. Ecosystems like CiRANO should provide their members with that kind of techniques. I believe that the emergence of new databases, which are not so relational and not so strict regarding data consistency, is vital. To this direction, NoSQL databases became a new trend in this domain and perhaps... (I am not fully convinced yet)... these databases can give some valuable solutions for managing big data. (CTO)

4.4.2 Hardware Challenges

Although development of new database models and creation of new analytics algorithms are crucial, lack of the required computing power makes it hard, to manage big data effectively. It seems that a challenge triggered by big data in context of SECOs is related to the hardware used for big data analysis. SECOs are challenged to find solutions pertaining to hardware underperformance. In particular, the CTO Manager points out that big data raises the need for hardware virtualization in order for parallel analysis of data across a number of cloud-based virtual machines to take place.
Every SECO is challenged to invest on hardware virtualization and parallel computing. Can you imagine what specifications should a single machine have in order to be able to analyze these vast and ultra-scale datasets in a timely manner? And then again, it will take weeks or even months to analyze and process big data in a non-parallel way. Trying to build such systems, we have to focus on hardware virtualization and on construction of virtual machines. To this direction, investments on Cloud computing is of paramount importance. With hardware virtualization, SECOs will be prepared for running analytical tools in parallel across a plethora of virtual machines. (CTO)

The importance of the cloud in solving big data issues is highlighted also by a software engineer. He says: “An organization that is interested in big data cannot neglect the cloud” (SE).

### 4.4.3 Integration Challenges

Further, the CTO manager continued the discussion pertaining to hardware virtualization by saying that creating virtual machines is not a panacea, since a lot of other aspects should be taken into account. Specifically, he claims:

> We have to consider that this environment is based on virtual machines that interact with each other and are basically self-resourced. It is naive to overlook problems such as resources management and integration issues. […] Today’s business landscape is characterized by a necessity for integration of existing software solutions. (CTO)

The above-mentioned integration challenge has been mentioned by almost every respondent either in context of integration of the various software systems or in context of integration of the multiple big data sources. More specifically, some of them talked about integration challenges in terms of bringing together the various virtual machines and software solutions needed for the analysis of big data (like the CTO above). Other respondents mentioned that integration issues arise also due to the big data’s characteristic of variety. Therefore, they mentioned integration challenges in terms of connecting multiple and heterogeneous big data sources. For instance, a programmer says:

> One of the biggest issues with Big data is pertinent to the variety and diversity of data sources. These data sources create data in various and incompatible formats. SECOs are challenged to find a way to integrate all of these data sources and this is definitely not easy; not only due to the number of data sources, but mainly because of data format inconsistency. (PA)

Also, a software engineer says:

> What has dramatically changed is the way of connecting and interacting with people and business… The rise of the IoT created the need for searching… and collecting data from every possible data source in order to understand better the customers’ needs and translate them into applications. Also, the communication
channels have changed and we try to be connected with everyone and integrate data from multiple data sources. (SE)

4.4.4 Data Storing Challenges

Furthermore, it seems that big data creates data-storing challenges pertaining to where data is stored and how this data is transferred. It became clear that due to the advent of big data, SECOs are challenged to improve their storing capabilities and data transferring techniques. A software engineer says:

*The majority of the contemporary organizations face a lot of difficulties to store huge amounts of high-velocity, high-variety data and consequently, storing such data is also a considerable challenge... unless an organization has the capability to create its own data center and support it with hyper-scale ability etc... and these firms are not too many. (SE)*

In addition, the vice president of b-open highlights that big data triggers challenges regarding storing data only in the cloud. “Storing data in cloud facilitates real-time access from everywhere. One huge challenge that has to be addressed is pertinent to the capability of storing data only in the cloud” (VP).

Additionally, a programmer embraces the opinion that big data creates a need for storing data in the cloud and for analysing data in the cloud through parallel software techniques and hardware virtualization. In particular, he says:

*Storing big data is not easy because we are asked for storing data in such a way that facilitates near real-time analysis. It is not only the database models... when our analytical tools are running in the cloud, storing data in the cloud could speed up the process. I do not want to dig deeper into technical aspects, but imagine that... when data is stored locally, it has to be transferred with some protocols to the place where data analysis happens. When the analysis is happening in the cloud with virtual machines and parallel computing... which is vital for managing big data... transferring data “from here to there” makes the process even more complex. So, storing data in the cloud and minimizing the need for data-transfer protocols seems to be another challenge for SECOs. (PA)*

4.4.5 Data Visualization Challenges

Even if a SECO tackles issues related to accessing and storing data, the challenge of bringing data in an actionable format or in other words the challenge of data visualization is emerged. The CTO manager mentions:

*For real-time visualization we need powerful computers as well as high quality software tools. [...] People got used to talk about products, problems, solutions and generally to have free-discussions in environments such as social media. The most of the enterprises are interested in these datasets. One thing associated with social media is how the unstructured data will be visualized in order for value to be extracted and data to make sense. (CTO)*
Following this notion, a Software engineer states: “Managing unstructured data is without any doubt a significant challenge for any business all over the globe” (SE). In addition, a programmer says:

All these data should be meaningful and that means that unstructured data should be brought into actionable format. Converting unstructured and highly heterogeneous data into structured data is something that requires relevant software tools. Platform providers are challenged to run projects focused on development of big data analytics. Nevertheless, these tools require talented people that have analytical capabilities to use these tools in a correct manner. (PA)

To this end, it seems that big data triggers also analytical challenges for a SECO. The next section describes this kind of challenges.

### 4.5 Talent and Analytical Challenges

Moreover, through the conducted interviews it became crystal clear that identification, access, management and quality assessment of relevant data are not easy tasks. Thus, obtaining big data analytical skills is crucial. “We are challenged to think in the same way of thinking of any potential customer all over the globe. This is not always so easy...”(CEO). It seems that SECOs have to develop big data analytical skills as well as to hire specialized staff. A programmer explains:

We are at a point that the value of big data is obvious. But, some questions regarding data management and data validity arise. What data should we collect? What is relevant data? How could we know what data to collect? What can we do with the collected data? How can we sense patterns and extract value? The quality of data is a matter that needs to be examined as well. All these can be translated into a challenge for any organization and for any SECO respectively to improve their analytical skills. (PA)

In addition, understanding that SECOs are challenged to develop big data analytical skills and hire specialized manpower, the CTO manager underlines:

A challenge we try to overcome is related to gaining access to [valuable and relevant] data [...]. This requires not only technical knowledge, but also evaluation skills, analytical skills, skills for understanding people, specialized staff and much more. (CTO)

### 4.6 Security Challenges

Moreover, with big data, a platform provider can gather and combine data related to personal facts of the various members of the ecosystem. Thus, it seems that another challenge is correlated with the fact that the focal firm should develop security mechanisms that prevent data leak and provide increased security. A marketer explains:

It is true that when data about individuals is gathered, security concerns arise. Some data gathered from a platform provider is pertaining to external
developers and end-users of the platform’s features. Data, such as motivations, behavior, habits, reactions, personal characteristics etc... Therefore, this is a sensitive issue when it comes to big data... especially when the collected data can lead to metadata and increased understanding of personal facts. For instance, when I try to persuade someone to participate in the CIRANO ecosystem, I am always asked about privacy issues and I have to make it crystal clear that we are very careful with data we gather. (MB)

4.7 Strategic Challenges
A lot of the respondents pointed out that big data raises the need for corporate openness, because by that way, platform providers can gain access to third parties data. The same programmer who discussed the opportunity for repositories mining and data-driven software development, claims: “But in order to mine all these all-over-internet repositories, they have to be transparent for real” (PB). Also, a marketer points out:

_Gaining access to third parties data means that you are welcome to explore their datasets and vice versa...I mean you have to allow them to do the same... Without any doubt, this is extremely challenging. Big data requires collaboration between ecosystems across different industries or even within the same industry. As I see it, big data is like an enemy that you cannot fight if you do not collaborate with others. Otherwise, it is just data that you gather and store. If you cannot combine data from various sources, then practically... we do not talk about big data._ (MA)

In addition, the CEO manager highlights: “It’s not possible for the information to be disseminated and for development to be achieved, if information is not open. In B-open, we extend the notion of openness by adding the significant characteristic of transparency” (CEO). The openness with respect to inter-organizational collaboration is something that the majority of the respondents mentioned as a valuable step to take most of big data. It became clear that platform providers are challenged to reconsider the way they compete. Nevertheless, although the adoption of coopetitive principles would be ideal, it seems that high level concerns of data privacy put barriers to this. The CEO explains:

_It would be possible for similar businesses with conflicting interests to collaborate. It would be possible data produced by one source, to be shared with other sources. The precondition for this is data to be open. For that reason, I believe that we have a lot of work to do regarding B2B collaboration._ (CEO)

4.8 A Brief Presentation of the Results
The empirical findings points at how big data influences the evolution of SECOs in the software engineering industry to a great extent. In particular, the advent of big data introduces a plethora of C&O in context of platform-based ecosystems.

As far as the identified opportunities are concerned, the empirical data illustrate that SECOs in the software engineering industry can take advantage of big data in multiple ways. In particular, the advent of big data offers to SECOs opportunities in terms of: (1) Marketing; (2) Operational Agility; and (3) Strategy. Regarding marketing opportunities, it is noteworthy that big data is seen as a valuable tool for deep, instant analysis of customer behavior. This has a
great implication for better, real-time customer understanding. Additionally, with big data network effects can be triggered more easily, since the motivations of potential ecosystem members can be captured in advance. Reasonably, SECOs are offered opportunities for more efficient group targeting, due to easier identification of market trends and consumer intents. Also, SECOs in the software engineering industry are given an opportunity to expand their product offerings, since organizations from different industries rely on them for building big data specialized software. Regarding opportunities for operational agility, it has become crystal clear that big data acts as a catalyst for a change in the way that software is produced by SECOs. The reasoning behind can be traced to the fact that big data allows SECOs to gather significantly detailed data about any internal or external member of the ecosystem. Therefore, it is now easier for SECOs in the software engineering industry to engage their customers into the development process and follow a "customer-driven software development" model. Additionally, big data holds a potential to make software engineering a data-driven process; Big data allows SECOs to hold significantly detailed data regarding the process of software engineering itself. Hence, SECOs in this industry can extract value from big data and transform software engineering into a highly automotive process. Further, in context of software development projects, valuable information can be mined for code cloning, repositories mining and bottlenecks prediction. A significant issue here seems to be the role of big data in achieving high-level sustainability through avoidance of duplication and elimination of resources waste. Regarding strategic opportunities, with big data SECOs are given the opportunity to create innovative products faster, but most importantly, to innovate through their business model; The data a platform owner holds can at some point become more valuable than the main purpose of the ecosystem. Thus, a platform owner can potentially change the way it creates value and move towards a data-as-a-service business model. Moreover, another strategic opportunity offered by big data is pertinent to the way decisions are made. Big data can catalyze a change in decision making within an ecosystem under certain circumstances – decision making based on gathered data).

As far as the identified challenges are concerned, it became clear that SECOs are challenged to overcome issues pertaining to: (1) IT infrastructure and techniques; (2) Talent and Analytical skills; (3) Security; (4) Strategy. SECOs in the software engineering industry are challenged to build applications that can deal with big data scalability. Thus, ecosystems should provide their members with necessary relevant tools. Due to big data's characteristic of high-variety, non-relational databases are incapable of storing and managing big data, and hence, SECOs are challenged to move towards non-relational databases. Moreover, hardware solutions and existing computing power are judged as inefficient to manage big data. This underperformance issue challenges SECOs to invest on hardware, cloud-based virtual machines as well as in parallel software techniques. What is more, the underlined need for bringing together the various virtual machines and connecting multiple and heterogeneous big data sources, militates in saying that big data creates various integration challenges. Further, the empirical findings show that storing big data is a challenging process and SECOs have to come up with ways to solve not only data format inconsistency (Variety) issues, but also issues related to cloud-based data storing. Also, due to the fact that big data is highly unstructured, a challenge to build software solutions capable of bringing big data in an actionable format
emerges (data visualization challenge). But, even if IT infrastructure and technical challenges are tackled, SECOs are challenged to find specialized people who have such analytical skills to work with big data. Additionally, the results of this study underline that with big data an abundance of security concerns have been triggered. Hence, SECOs are challenged to protect sensitive data and develop security mechanisms that prevent data to leak out. Moreover, a significant finding is linked with strategic challenges a platform owner has to tackle in the face of big data. It has been illustrated that big data asks for competition to be at the heart of a platform-based ecosystem in order for transparency to be at high levels and access to third parties data to be gained. Therefore, it became clear that platform providers are challenged to reconsider the way they compete within an ecosystem or with other ecosystems.

5. Discussion

After describing the research method and analyzing the findings, it is significant for the researcher to pull back and discuss the research outcome. In this section, a step back is taken in order to become clear whether the results answer the initially formulated research question: “What are the C&O associated with the advent of big data in context of software ecosystems in the software engineering industry?” Also, it is examined how the outcome of the present thesis fits in the previous literature and judgments about the contribution of this thesis are made. Further, after identification of limitations of our research, future research avenues are suggested.

5.1 Theoretical and Practical Implications

Putting our findings into context and placing the thesis in the general field properly, it is required for the level of consistency with previous literature to be examined.

The present thesis extends to a reasonable degree the existing literature too, since it connects two phenomena (i.e. SECOs and big data) that had typically been approached as secluded spheres. By examining a real-world ecosystem, increased understanding of the way that SECOs are influenced by big data has been gained. In particular, it has been shed light on C&O that platform providers have to manage in the face of big data. This list of C&O is judged as the main contribution of this paper.

Studying the impact of big data on a real-world ecosystem is of great importance, since in previous literature it has been highlighted that there are few studies focusing on real-world ecosystems (Manikas and Hansen, 2013) and few empirical studies on assessing big data's power (Wamba, Akter, Edwards, Chopin and Gnanzou, 2015).

According to the results of the present thesis, big data offers platform owners opportunities for more efficient and sophisticated marketing. Manyika et al. (2011) underline that big data offers opportunities for better understanding and segmentation of customers. These opportunities have also been identified in the present thesis. In addition, it has been highlighted that for an ecosystem, big data can be a driving force of establishing strong network effects and achieving long-term customer retention.

Previous research has highlighted that big data offers organizations the opportunity to improve the quality of their operations and outcome by following data driven processes
(McAfee and Brynjolfsson, 2012). In this study, it has also been found that platform owners can take advantage of big data and seize opportunities related to operational agility.

The majority of the respondents stated that platform owners are offered by big data enormous opportunities for product as well as business model innovation. The connection between big data and innovation is also mentioned in previous studies (i.e., Brown et al., 2011). This thesis comes to add value on top of them by narrowing down and discussing how platform providers are offered by big data opportunities for innovation. Additionally, the results of this thesis increase the level of awareness of the influence of platforms on the innovation process (Gawer & Cusumano, 2014; Ghazawneh, 2010). Some of the interviewees took the discussion further and underlined that platform ecosystems facilitate innovation. Another identified strategic opportunity is correlated with better decision making. Previous studies underlined that big data can act as a catalyst for improved decision making (Wamba et al., 2015) and our study gives support to this finding.

Moreover, the findings illustrate that platform owners are challenged to use and develop more efficient database models in order to deal with big data. The respondents highlighted that one of the greatest challenges that big data creates is pertinent to a need for non relational databases and that structured databases are inefficient for storing and processing big data. This is exactly in line with what McAfee and Brynjolfsson (2012) state: “the structured databases that stored most corporate information until recently are ill suited to storing and processing big data”.

In addition, this thesis comes to strengthen the validity of the definition of big data provided by Manyika et al. (2011): "Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze". In their paper, Manyika et al. (2011) highlight that new technologies are needed for managing big data. This study empowers this opinion, since it has been highlighted that both hardware and software challenges are posed by big data. Consequently, investments on IT infrastructure are of vital importance in order for a SECO to be able to manage big data.

Further, the results of the present thesis illustrate that SECOs are challenged to invest on parallel computing and hardware virtualization in order to be able to extract meaningful value from big data. This is in line with studies that present big data challenges related to infrastructure and techniques, such as parallel data processing (i.e., Dobre and Xhafa, 2014). Regarding the process of bringing big data into an actionable format, this study concludes that it is a challenging process and it seems that other studies came to the same conclusion. For instance, Manyika et al. (2011) characterize big data visualization as a key challenge that has to be met.

Moreover, the empirical results presented above show that in order to extract value from big data, SECOs are challenged to have high-level analytical skills and therefore, platform owners are challenged to recruit staff that is specialized on big data analytics. The fact that big data raises talent shortage issues is also underlined by Davenport and Patil (2012) who raise the need for recruitment of data scientists; "people that can coat treasure out of messy, unstructured data". To this end, this thesis encourages people to study courses related to big data analytics and universities to design and offer big data related programs.
As previous research has shown big data is seen as a frontier for competition (Manyika et al., 2011) and the results of this study verify this statement. However, in this thesis it is argued that in a platform ecosystem setting, big data questions also the platform owner’s perspective on competition. Particularly, platform providers are challenged to open the gates to competitors and collaborate with them. Therefore, in this paper it is argued that co-opetition should be at the heart of a platform ecosystem in order for the platform owner to take advantage of big data.

Considering what the practical implications of the results are and how the findings can be used by practitioners, it is worthy to note that the present thesis encourages platform providers to pay close attention to big data and invest on big data analytics, since an abundance of opportunities are offered. At the same time, this study raises awareness of the challenges that can be emerged by big data and that platform owners have to tackle. To this end, valuable guidelines can be provided.

Being prepared for the big data evolution and dealing with high-variety, high-velocity data demands two fundamental things: the right skills and the right technology. Therefore, platform owners should realize that in order to unlock the power of big data they have to hire specialized staff or educate existing human assets. As it was mentioned before, to this direction it is strongly advised, universities to provide students with big data capabilities and offer big data courses. Also, capturing the full potential of big data requires investments on IT infrastructure in order for the ecosystems to be prepared for addressing scalability issues. Managers should take care of this issue and examine thoroughly the possibility to use techniques, such as hardware virtualization, parallel computing and data storing on the cloud. All these techniques could also be offered to external developers and maybe that could be a great motivation. Additionally, cloud computing seems to be a great vehicle for navigating the complex big data landscape and due to that, platform owners are encouraged to invest on cloud computing, allowing not only data storage on the cloud, but also data analysis. What is more, managers should stop using relational databases for big data, since they are inefficient for this purpose.

On the contrary, they should give their best to develop and use non relational databases that can store and process big data more efficiently. Further, platform providers need to leave fragmented platform structures, reconsider their competitive approach and build such inter-industry relationships that allow third party data access, data integration and metadata creation. It seems that although, the adoption of co-opetitive principles is a prominent step for data sharing and metadata creation, data privacy issues make it hard for co-opetitive platform based ecosystems to bloom. In consequence, platform owners are encouraged to work on ensuring privacy and security in order not only to build strong network effects, but also to allow data transfer to mushroom. Moreover, in order for SECOS to leverage innovation in the era of big data, software reuse should not be a taboo issue. One the contrary, platform owners should show the way and increase the transparency of their repositories and platform’s architecture.

Last but not least, I would like to point out that managers should not see big data from a narrow point of view, associating it only with vast amounts of data. This thesis suggests firms to interpret big data as a significant tool for strategic decision making and operational agility.
6. Conclusions and Suggestions for Future Research

The present thesis opens up certain avenues for future studies. Particularly, suggestions for future research are made and they are related to the following aspects: 1) how can a platform owner motivate externalities to participate in an ecosystem? 2) how platform owners can balance the level of autonomy and dependency between various participants? 3) what kind of platform strategies can a platform owner use in order to protect all the members of an ecosystem and overcome challenges associated with B2B data share? 4) what is the role of the platform architecture in the evolution of SECOs and in overcoming big data challenges? 5) why competitors participate in SECOs? what are their motivations? 6) what is the role of platform based ecosystems in terms of sustainability and avoidance of data waste, 7) how platform based ecosystems could be integrated with each other resulting in a "farm of ecosystems" and what could be the problems and prospects, 8) a further discussion about the influence of big data on the emergence of platform based ecosystems and a case study of an emerging ecosystem in another industry, 9) exploration of the strategic significance of big data; A number of our findings support a landing on the conclusion that there is a proportion of people that conceptualize big data as huge datasets, underestimating its significant influence on business strategies. This conclusion resembles the argument of Constantiou and Kallinikos (2015) and Yoo (2015) that the significance of big data is far beyond the vast amounts of data and by interpreting big data in such way, one can overlook the fundamental role that big data plays in strategic decision making. However, this was not a clear outcome and it was indirectly mentioned during the interviews. A study that focuses on the strategic significance of big data would be promising and is highly suggested. 10) A deeper understanding of big data and business model innovation through a case study. 11) Further, something that is still on the planning phase regarding CiRANO ecosystem is how the pricing model will be set. A future study could focus on different pricing models within a SECO. 12) Another point that it would be great to be made is that CiRANO is an emerging ecosystem based on an until recently internal platform. Hence future studies can examine CiRANO in order to shed light on issues pertaining to the emergence and evolution of platforms. For example, one could examine difficulties associated with transforming an internal platform into an ecosystem (i.e., how people felt about such a change) or to study critical success factors or the reasoning behind such a transition. Of course, it would be also interesting to see the impact of big data on the evolution process through a longitudinal case study.

All in all, it is judged as critical to list the most significant findings from this study:

- Big data influences the evolution of SECOs in the software engineering industry to a great extend.
- Due to the emergence of big data, platform owners are challenged to solve issues pertaining to IT infrastructure, talent and analytical skills, security and strategy.
- Big data offers to platform providers opportunities related to more effective marketing, operational agility and strategy.
7. Concluding Remarks

The present thesis has shed light on the impact of big data on the evolution of platform-based ecosystems and has answered the following research question: “What C&O are posed by big data in context of platform-based ecosystems in the software engineering industry?”

This study has strengthened the findings of previous research and has also broaden the horizons of related literature, since it has provided accurate understanding of understudied issues associated with the relationship between big data and platform based ecosystems. I would like to highlight that the main contribution is that various challenges and opportunities posed by big data in context of software ecosystems have been described in detail. The treasure of our research is located in this list of challenges and it should be mentioned that the presented practical implications are of vital importance for an up-to-date platform provider. Therefore, practitioners as well as researchers are advised to dig deeper into this relationship.

Acknowledgements

My first acknowledgement is to my loving family for their patience, encouragement and support. Mom and Dad, thank you for giving me the opportunity to study abroad and for supporting me with so much love.

Also, this journey would not be possible without the enduring support of my friends and Zoe. I am very grateful to all of you.

I would like to thank my professor, supervisor and mentor Jonny Holmström for giving me valuable scientific guidance.

I wish to express a sincere thank you to B-open for allowing me to conduct research at their company and a special thanks to the CTO of the firm for his hospitality and support.

Last but not least, special thanks go to my classmates who revised my thesis and gave me valuable feedback.
References


Hanssen, G. K., & Dybå, T. (2012). Theoretical foundations of SECOs. In IWSECO@ ICSOB (pp. 6-17).


Johnson, J. E. (2012). Big data+ big analytics= big opportunity: big data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. *Financial Executive, 28*(6), 50-54.


Appendix

Figure 1. Software Platform Ecosystems

Figure 2. The 3Vs of Big Data / Source: Diya Soubra, Data Science Central, http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data
Figure 3. CiRANO Software Platform Ecosystem
**Figure 4. Infographic from DOMO available at:** [https://www.domo.com/learn/data-never-sleeps-2](https://www.domo.com/learn/data-never-sleeps-2)
<table>
<thead>
<tr>
<th>Respondent’s position</th>
<th>Main responsibilities and duties</th>
<th>Years of employment in B-open</th>
<th>A word or phrase that describes b-open</th>
</tr>
</thead>
</table>
| Founder and Chief Executive Officer (CEO Manager) | - Setting strategy, vision and long-term goals  
- Overall Management of the firm  
- Creating and Promoting the organizational culture | since its foundation (11 years) | “We are not only interested in selling software. We want to give our customers the opportunity to grow. The increased growability of our customers is our success.” |
| Founder and Vice President (VP) | - Head of Human Resources Department  
- Product Manager of a plethora of applications  
- Application Support Director | since its foundation (11 years) | “Our philosophy is described completely by the title of the firm.” (B-open) [consider the pun: be-open] |
| Chief Technology Officer (CTO Manager) | - Designing the solutions that the company sets as the main goals  
- Team-Building and Organizing  
- Managing the development team in order for the available assets and resources to be fully utilised | 10 years | “Collaboration is the word that describes b-open’s philosophy. Collaboration between people is at the heart of b-open. For that reason, the main goal of our platform is to promote this collaboration between different developers.” |
| Software Engineer (SE) | - Software Development  
- Analysis of the customer needs in order to create quality IT solutions that meet the needs of our customers | 3 years | Openness, Stay close to the customer and provide services such as rapid development, rapid and efficient support |
| Programmer A (PA) | - Encoding Information into a programming language  
- Developing Software applications and modules  
- Testing the applications  
- Maintaining the IT solutions | 7 years | “B-open means working with creative and open people.” |
### Appendix 4: Respondents’ Data

<table>
<thead>
<tr>
<th>Respondent’s position</th>
<th>Main responsibilities and duties</th>
<th>Years of employment in B-open</th>
<th>A word or phrase that describes b-open</th>
</tr>
</thead>
</table>
| Programmer B (PB)     | - Encoding Information into a programming language  
- Developing Software applications and modules  
- Fixing bugs in application and modules | 2 years | “We are committed to creating sustainable IT solutions of high quality to ensure customer satisfaction” |
| Programmer C (PC)     | - Encoding Information into a programming language  
- Developing Software applications and modules  
- Fixing bugs in application and modules | couple of months | “A firm that gives hope… a great opportunity given” |
| Marketer A (MA)       | - Collecting Data for existing and potential customers  
- Planning and Executing marketing plans  
- Creating and Ensuring a mutual beneficial relationship with the customer base | 7 years | “B-open is more than a workplace to me. We are a family that loves creating IT solutions and serving people with the best quality.” |
| Marketer B (MB)       | - Working with social IT  
- Executing marketing plans | 1 year | “Hospitality, Openness and Friends!” |
Interview Guide

Respondent Background
1. What is your job title in B-open? What are the main responsibilities and duties of this position?
2. How many years do you work in B-open?
3. What are the main activities and responsibilities of your job?

About B-open and existing processes
4. Could you please tell me three words that in your opinion characterise the today’s business environment?
5. Could you please tell me a word or phrase that describes B-open’s philosophy?
6. What services does B-open offer?
7. What types of clients do you serve? sectors etc
8. Do you operate only in Greece or you have clients from abroad as well?
9. What are in your opinion the main advantages of jPlaton platform? What other advantages does the emergence of CiRANO bring?
10. What are the main products that have been developed through jPlaton? Are these products customizable and extensible? Do they allow integration with other systems?

Big Data & Big Data Analytics
11. On what data do you sit on? How do you currently collect the aforementioned data? What are the main data sources for B-open?
12. What difficulties do you identify in the process of accessing the appropriate data?
13. What is in your opinion the importance of big data for a platform provider?
14. Have you noticed any change in the way B-open operates today in comparison with the way B-open used to operate in the past?
15. Do you see social media as a source of data? Is the unstructured data a challenge for a contemporary organization?
16. Do you use any Big Data Analytical tool? If yes, which one & how people responded to the idea?
17. Have you developed any analytical module for your products?

Challenges, Problems & Opportunities
18. Do you see any challenges or problems that Big Data brings on stage and contemporary organizations are challenged to overcome?
19. How does B-open response to these challenges? How is B-open adjusted to this constantly changing environment?
20. Could you identify any opportunities that Big Data offers to the contemporary organizations?

**Platform Architecture, Openness and Governance**

21. During the seminar, collaboration and integration were highlighted. What is the importance of collaboration and of generativity for B-open? How does the architecture help in that direction?
22. From your perspective, what does platform openness mean? How does it fit with B-open’s philosophy? What are the advantages of platform openness?
23. What are the future plans of B-open? What are the short term and long term goals of the firm?

**Ethical Considerations**

24. How does B-open protect the various users and developers of the platform? How are developers protected in terms of intellectual property and parts of apps that they want to keep for themselves?
25. One of your products is associated with health informatics. What are the ethical considerations in this case?
26. Should organizations consider ethical aspects regarding sensitive personal data and metadata creation?
<table>
<thead>
<tr>
<th>Open Codes</th>
<th>Axial Codes</th>
<th>Selective Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- gaining real-time customer insight</td>
<td>Real-time insights for the whole ecosystem</td>
<td>Marketing Opportunities</td>
</tr>
<tr>
<td>- easier to know who is really interested in the offerings</td>
<td>Opportunity for Better Customer Understanding</td>
<td></td>
</tr>
<tr>
<td>- he believes that a SECO can learn what a customer wants rapidly</td>
<td>Opportunity for Behavioural Analysis &amp; Network Effects</td>
<td></td>
</tr>
<tr>
<td>- understanding end users and external developers</td>
<td>Opportunity for Long-term Customer Retention</td>
<td></td>
</tr>
<tr>
<td>- understanding customers’ needs</td>
<td>Opportunity for Efficient Group Targeting</td>
<td></td>
</tr>
<tr>
<td>- understanding expectations</td>
<td>Opportunity for Market Expansion</td>
<td></td>
</tr>
<tr>
<td>- sensing patterns in data regarding motivations of potential members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- predicting behaviours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- building network effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- creating customized products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- achieve customer satisfaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- increase customer retention rates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- better group targeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- offering new software products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- keeping the customers satisfied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- know where to focus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- leader 1 to be leader 2..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- know who is going to follow...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- understand the intent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- need for creation of new products for big data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- knowing about different potential customers from everywhere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- having real-time information; getting feedback immediately</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- getting help in understanding who is interested in what and most importantly when</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- they are challenged to develop new software solutions
- they feel that the existing hardware is inefficient for managing big data
- having concerns regarding the way they access, store, analyze and visualize our data
- important to produce software precisely based on big data
- lack of required computing power
- worrying about scalability
- need for big data designing techniques
- the relational databases are inefficient for unstructured data
- challenge for new database models
- from relational to non-relational databases
- concerns about scalability
- hardware virtualization
- need for building virtual machines and do parallel analysis
- cloud computing is important
- integration issues arise - integration of analytical tools - integration of heterogenous sources - diversity of data sources triggers integration challenges
- storing big data is also a considerable challenge - storing data only in the cloud; again cloud computing significant
- storing data in such a way that facilitates near real-time analysis
- need for powerful computers and high quality software tools
- challenge for bringing unstructured data in an actionable format - unstructured data to make sense

<table>
<thead>
<tr>
<th>Software Challenges</th>
<th>Hardware Challenges</th>
<th>Integration Challenges</th>
<th>Data Storing Challenges</th>
<th>Data Visualization Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Infrastructure Challenges</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Challenges</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Open Codes

- the end-user as a valuable source of info; guideline for software development; with big data we can have this info in real-time
- understanding customers and the needs are discovered in advance; this helps production
- with big data the needs can be involved in the production process rapidly
- customization becomes easier
- he feels that quality can be increased; increased automation in software development; minimize time consuming processes; big data analytics in SE sounds promising
- analysing the development processes with big data; sense patterns; perhaps predictions are possible
- more efficient repositories mining is possible… repositories are becoming transparent
- code cloning is beneficial; it seems that previous generations do not embrace this idea
- code cloning is easier; a developer can identify parts of code from other projects with big data analytics…then, he can use code from another developer
- SECOs are different from open source… an initial question …why?
- creating green SECOs… write less code lines… decrease waste of resources; duplication can minimized; … …

## Axial Codes

- Opportunity for Customer-driven Software Development
- Data-driven processes
- Achieving Increased Customization
- Applying Big Data Analytics in SE processes
- Opportunity for Business Automation
- Predicting bottlenecks and projecting needs related to resources
- Creating Green SECOs; Environmental Sustainability

## Selective Codes

- Opportunities for Operational Agility