Towards A Process View on Critical Success Factors in Big Data Analytics Projects

Full Papers

Jing Gao
University of South Australia
Jing.gao@unisa.edu.au

Andy Koronios
University of South Australia
Andy.koronios@unisa.edu.au

Sven Selle
University of South Australia
selsy015@mymail.unisa.edu.au

Abstract

The research tries to identify factors that are critical for a Big Data project's success. In total 27 success factors could be identified throughout the analysis of these published case studies. Subsequently, to the identification the success factors were categorized according to their importance for the project's success. During the categorization process 6 out of the 27 success factors were declared mission critical.

Besides this identification of success factors, this thesis provides a process model, as a suggested way to approach Big Data projects. The process model is divided into separate phases. In addition to a description of the tasks to fulfil, the identified success factors are assigned to the individual phases of the analysis process.

Finally, this thesis provides a process model for Big Data projects and also assigns success factors to individual process stages, which are categorized according to their importance for the success of the entire project.

Keywords

Big Data, Critical Success factor.

Introduction

The amount of available information has been overflowing during the past years. Organizations generate, receive and store masses of data about customers, employees, orders, suppliers, competitors and many other aspects of their business. This situation is not just a current trend, but is considered to be an ongoing development and the data is expected to grow even faster in the future. Besides the enormous size of data, organizations are challenged by the number of various sources, which provide information. Sensors integrated in smart phones, cars, credit cards or general consumer electronics provide a new range of available information sources. Furthermore, social media provide valuable information about customers to companies. Due to these developments organizations now have to deal with more data sources than ever before. Even though the information is provided from different types of sources, text based data will continually play a major role for analytics (Russom, 2011, p. 7).

All the mentioned characteristics of the new age of data are classified as Big Data throughout the literature. The highest importance is assigned to topics arising around Big Data. Manyika et al. (2011, p. iv) for instance claim that "Big data (…) is now part of every sector and function of the global economy.

Like other essential factors of production such as hard assets and human capital, it is increasingly the case that much of modern economic activity, innovation, and growth simply couldn't take place without data". Gantz and Reinsel (2011, p. 2) even predict a change not only within companies, but also for society through Big Data: "The convergence of technologies now makes it possible not only to transform the way business is conducted and managed but also to alter the way we work and live".

Although data plays such a major part, organizations struggle to leverage from it. Kelly and Kaskade (2013, p. 2) report as one of the results of their survey on Big Data, which was conducted over 300 companies that "55% of Big Data projects don't get completed, and many others fall short of their objectives". The purpose of this research is to address the issue of failed projects and to deliver guidelines on how to succeed with Big Data analytics. It is created to provide useful insights into procedures on how to overcome obstacles in Big Data projects.

Overview Of Big Data Related Issues

The topic of Big Data emerged during the last years due to changes in using and processing information. Data warehouses have been the base for most of the data analysis tasks during the past decades and they are still playing a major role in this area. The possibilities of these systems however are limited. Information nowadays is captured in many different formats and stored in a big variety of sources.

Laney (2001) published the first paper discussing this change and the new challenges faced in data management by using typical Big Data characteristics volume, variety and velocity. This problem description has been used, modified and refined by different authors since. Gartner (2013a) for instance is using Laney's characteristics for their own official definition: "*Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*". However, Big Data problems are complex and can therefore be seen and defined from different perspectives. Gualtieri (2012) is defining Big Data with a business focus as following: "*Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information*". Howie (2013) defines Big Data from a technical perspective as "*the frontier of a firm's ability to store, process, and access all the data it needs to operate effectively, make decisions, reduce risks, and serve customers*". This research is following the Laney and Gartner definition of Big Data being characterized by the three dimensions volume, variety and velocity.

Due to the immaturity of the available Big Data solutions it is important to focus in projects not only on technologies, but to be aware of all challenges the project has to face. Sicular (2012, p. 9) claims that three major areas have to be taken into consideration in IT projects: "*Three equally important and interdependent components cause success or failure of an undertaking: people, process and technology*". Applying this categorization, Big Data projects have to face the following challenges:

- **People:**

  Big Data experts are very difficult to find and expensive to hire due to the novelty of this area. There are simply not a lot of people with the needed computational background and analytical skills. These experts are often referred to as *data scientists* (Davenport & Patil, 2012, p. 7).

  Data scientists gain domain expertise or industry knowledge, which is used later on in the Big Data analysis to gain the best results. They also develop a detailed understanding of specific business practices within the organization they are working for (Giannikas, 2011, p. 9). Furthermore, it is vital in Big Data projects to involve experts from the business right from the beginning of the project to ensure maximal success by focusing on business values (Sicular, 2012, p. 11). Data scientists usually have a background in software development and computer science, but they also do have basic analytical skills such as statistics (Giannikas, 2011, p. 9). The term scientist comes from the way these experts are looking for answers within data sets. They ask questions and set up experiments in research fashion (Giannikas, 2011, p. 9).

  In most of the Big Data projects it will be beneficial not only to relay on the skills of one single data scientist. Big Data projects require a various set of different skills best covered by a team of experts

from different areas. Sicular (2012, p. 10) claims that the bests result can be expected by creating a multidisciplinary analysis team, with the required skills needed for the respective project.

Another challenge arises from the huge security and ethical issues, which come along with Big Data projects. Therefore, risk management and legal experts should be involved in very early project stages, as well as compliance representatives (Sicular, 2012, p. 27).

- **Process:**

Big Data projects are innovation projects and require therefore a special treatment. The analysis team has to be able to react on changes during the analysis process. The team has to be organized in a way that allows all members to react flexible on these changes. Often changes are necessary due to revealed data management problems detected throughout the Big Data analysis. These problems require an immediate action and cannot be ignored (Sicular, 2012, p. 10).

Many organizations are challenged handling the ever growing large amounts of data. In order to ease this problem, organizations are likely to get into the situation to decide whether to keep all the emerging data or store only these parts that really benefit the business. If the decision is made in favor of determining upfront which data is relevant, it is important for organizations to establish processes to identify those documents that contain the information with the highest potential business value (SAS, 2013).

It is important that the results are based on high data quality, as the actions resulting out of Big Data projects can have severe consequences for an organization. As the data usually comes from a number of different sources from inside and outside of the organization, it is very challenging for the analysis team to ensure a high quality of the data provided from these sources (Sathi, 2012, pp. 4-5).

Due to the innovative character of Big Data projects, members of the analysis team need to be engaged to think in innovative ways and come up with creative ideas. Nevertheless, it is necessary to set some boundaries for the analysis team to ensure the project does not lose its focus. Within these boundaries, however, team members should be granted total freedom (Sicular, 2012, p. 29).

Supporting a well-defined business goal is vital for Big Data projects in order to deliver a value for the organization. Measurements have to be implemented to determine to what degree the business goals were supported by the project (Sicular, 2012, p. 27). Furthermore, the analysis has to follow the fundamentals of cost/benefit analysis, as well as core statistical science to succeed (Gopalkrishnan, Steier, Lewis, & Guszcza, 2012).

Another challenge arising from the novelty of Big Data is dealing with the expectations from the business towards it. The project team has to make sure to set boundaries for expectations coming from the business. These boundaries, however, have to be set in a way to keep the business still interested in the projects (Sicular, 2012, p. 26).

- **Technology:**

The major technical issues are the abstinence of end-to-end solutions, design patterns or universal architecture recommendations for Big Data problems due to the technical immaturity. That means that the used technology has to be individually assembled and customized to the purpose of the analysis project (Sicular, 2012, p. 9).

Due to the massive volume of data used in the projects, storage is going to be an issue sooner or later. This problem needs to be addressed in early planning stages of the project (Sicular, 2012, p. 20). Moreover, traditional database systems won't be able in some organizations to deal any longer with the ever growing amount of data. Bhashyam stated already in the year 2011 that Big Data is already 10 – 100 times as voluminous as typical data warehouses at this time. There is a need for new technology to address this situation ((Bhashyam, 2011)).

One of the aspects of Big Data is the transformation from large-scale, enormous-size, heterogeneous data repositories into well-structured data, which can be easily analyzed and interpreted. For some use cases the so transformed data has to be transformed even further to fit the requirements of typical

> *business intelligence* components such as diagrams, plots or dashboards. It is technically challenging to bring systems together that are so different in their nature and purpose of use. Big Data, however, is more than just data transformation into a structured form. Traditional database systems are built on data organized in a highly structured form in tables. Due to the various numbers of sources in Big Data projects, it is very challenging to extract the information in such a highly structured way. It is very likely that information gets lost during the transformation process of this highly unstructured raw data to the structured form required by data base systems(Gopalkrishnan, Steier, Lewis, & Guszcza, 2012) and (Cuzzocrea, Song, & Davis, 2011, p. 101).
>
> Merging the data from different systems with different security levels, intellectual property and privacy settings into one system is causing big security issues that Big Data projects are facing. The Big Data project teams need to find solutions to make sure sensitive data is only displayed to people who are supposed to have access to it (Tankard, 2012, pp. 5-6).

As discussed above, there are a number of different challenges to overcome in Big Data projects. People, process and technology perspectives have to be included in information management strategies. To focus on more than just technology is especially important in Big Data projects as the technology is still immature and new to many organizations. This circumstance has to be compensated by paying special attention to processes and people involved in Big Data projects (Sicular, 2012, p. 9).

The process of answering questions through Big Data analysis requires a certain set of skills that are preferably provided from a multidisciplinary project team. There are existing process models for business analysis from the age before Big Data, which have been proven to be efficient. Although these process models have proven to cover all the relevant facts of business analysis, there are certain issues to be taken into consideration if these models are applied on Big Data projects. Therefore, adaptations of these models are necessary to meet the Big Data characteristics volume, variety and velocity. Also, learning is an important factor in the planning process to have a foundation for future Big Data projects (Sicular, 2012, p. 11).

By adapting various project management methodologies, Figure 1 visualizes the process model that was composed by the researcher throughout the literature review.

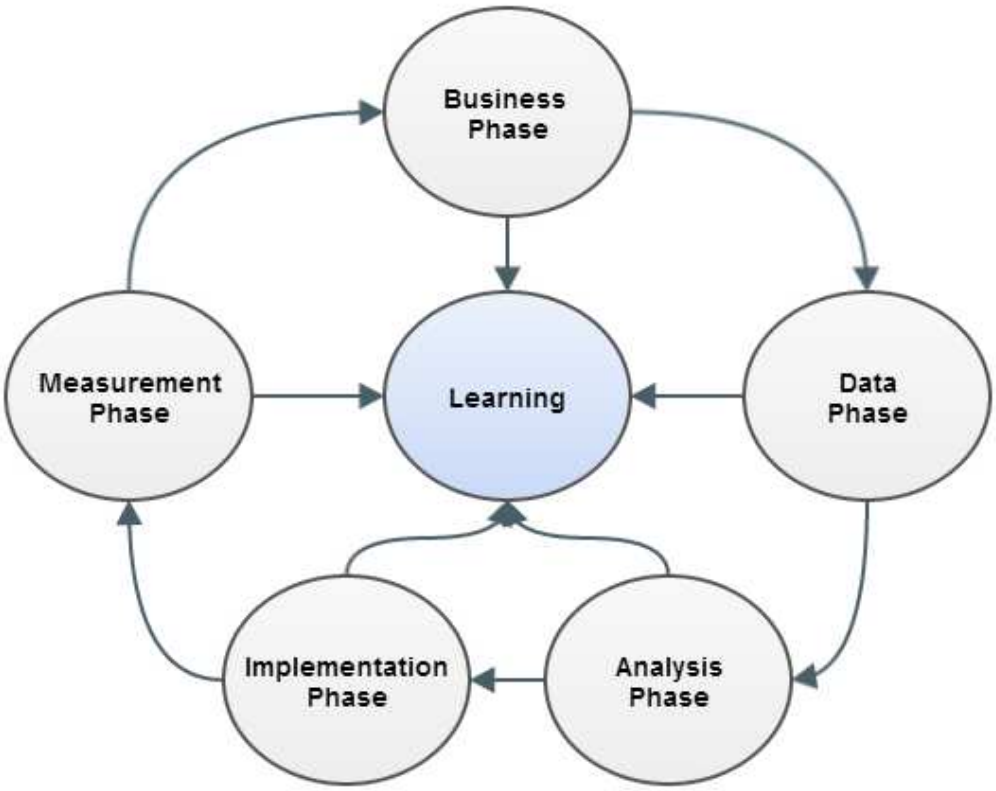

Figure 1: Business Analysis Process - created by the researcher based on (Sicular, 2012) and (Gualtieri, 2013)

Research Question and Design

In order to achieve the goal of providing guidelines on how to succeed with Big Data projects, the following research question will be answered throughout this document:

- How can organizations embrace success in Big Data text analytical projects?

To investigate the problem in a more detailed way, the research question will be subdivided into the following questions:

- Which process model can be applied for Big Data projects?
- What are critical success factors for Big Data projects?
- Which rule do individual critical success factors play at different project stages?

The research method used is content analysis. Berg (2001) defines content analysis as an objective coding scheme applied to data in order to make it amendable to analysis and systematically comparable. According to Busch et al. (1994 - 2012) it is a "*research tool used to determine the presence of certain words or concepts within texts or sets of texts. Researchers quantify and analyse the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts*".

This research is based on secondary data from different sources. According to Berg (2001, p. 214) "*secondary sources involve the oral or written testimony of people not immediately present at the time of a given event. They are documents written or objects created by others that relate to a specific research question or area of research interest*".

The cornerstone of this research is the collected 60 case studies, which document experiences from previously executed Big Data projects. These case studies deliver an insight into CSFs that have been considered in single projects in order to achieve the individual project goal. In addition to the case studies, surveys reports with a total number of 13727 responses, blog entries and guidelines from institution such as Gartner or McKinsey have been added to the data collection. These additional data was collected in order to add different perspectives on Big Data analytics. Each of the added document types adds a different advantage to the analysis.

Table 1 lists the collected data types, the corresponding number of documents for the individual types, as well as a few providers the data was extracted from.

Type	Number of Documents	Provider Examples
Published Case Studies	60	Cloudera, IBM, SAS
Public Survey Reports	14	IBM, McKinsey, TDWI
Guidelines	32	Forrester, Gartner, McKinsey
Blog Entries	21	Forbes, techrebulic, zdnet

Table 1: Summary of Document Collection

The survey reports add a wider view from a bigger number of participants to Big Data issues, based on other real scenarios. Blog entries allow this research to cover very recent developments in the Big Data area. A comprehensive view on Big Data, related to an organization, its strategy and its data management strategy is added by the inclusion of the guidelines.

The choice of basing the research on secondary data was influenced by the availability of the resource time. Secondary research data is usually available immediately, which is an important factor for this research. Another advantage of the analysis, based on secondary data, is the possibility to access a large amount of data covering a wider area, which may be unfeasible to collect as primary research data (Naveen, 2012).

Table 2 shows the adapted, predicted and summarized CSFs for Big Data projects. The CSFs have been deduced from the researcher from the general CSFs in general projects, IT-projects and data management studies. These CSFs have been adjusted according to the requirements arising from the Big Data challenges mentioned previously. The *Keywords* column contains the keywords that cluster into the concepts of the individual CSFs.

Dimension	Success factor	• Keywords
People	• Educated analysts in text analysis methodologies are needed.	• Analytical, Statistical Skills
	• Big Data technology skills within own staff or hired externally are needed.	• Technological Skills
	• The analysis team should consist of members from different business departments that are involved in the project, as well as members from the IT department.	• Multidisciplinary team
	• IT staff needs to be aware that creating and maintaining the documentation of the new implemented systems is vital. Especially because these are innovative systems.	• Documentation

Process	• Identifiable value for the business is delivered from the project. The project must support the business needs.	• Identifiable Business Value, ROI,TCO
	• Creation of a clear vision and implementation strategy for the software and hardware components of the new analysis platform.	• Big Data Strategy
	• Creation and disseminating of documentation needs to be a fixed part of the project.	• Documentation
	• The growth of analysis data has to be forecasted to be able to react early when used systems need to be expanded.	• Forecast Data Growth
	• Processes need to be established in order to ensure high data quality of the documents used for the analysis.	• High Data Quality
	• A measureable project outcome has to be defined and in fact measured to evaluate the success of the project.	• Measurement, Defined Outcome
	• Strong leadership has to be provided throughout the project. The top management level of the organization needs to get involved.	• Top Management Support, Executive Level Support, Leadership
	• A clear goal of the project has to be defined, and measurements have to be established. The project needs to have a clear start and end.	• Project Goal, Milestones, Deadlines
	• The size of the project has to be determined carefully.	• Clear, Manageable Project Scope, Adequate Size
Technology	• Awareness of recent developments in the Big Data analysis area, including evaluation, selection and implementation of new software methodologies.	• Investment in new Technologies, Analytical Tools
	• Document collection is a significant problem, only with the right documents provided the needed answers can be found within the data.	• Document collection, Access to Sources
	• Evaluation of the used hardware to make sure it meets the requirement of the growing data amount also in future.	• Adequate Hardware, Storage, Real Time
	• Developing, integration and application of logical data models for the new analysis platform.	• Integration of New Solution, Coexistence
	• Performance has to be monitored continuously and the system needs to be tuned on a regular basis to meet operational goals.	• Fast Performance, Fast Delivering of Results

Table 2: Research Model as Basis for Coding of Data Collection

Table 2 is used as base model for the coding of the document collection. The content analysis process itself can be seen as manual application of text analysis based on a data collection from various sources from the internet. The adequate method that was used is a combination of keyword extraction and concept extraction.

This research also uses the *ABC analysis* to categorize the extracted CSFs according to their importance for the project's success. The content analysis that is described in is believed to extract a range of potential critical success factors. It is acknowledged that these factors do not carry the same weight in importance. The purpose of the *ABC analysis* is to rank the extracted factors according to their significance and thereby identifying factors, which have the highest importance. Although, these factors have a different level of importance; they will be extracted to allow a broader coverage of Big Data aspect. There are no fixed thresholds for the A, B and C category. This research uses the following thresholds suggest by Lysons and Farrington (2005):

- A-level CSFs: **20 %** of the CSFs account to 70% of the project's success
- B-level CSFs: **30 %** of the CSFs account to 25% of the project's success
- C-level CSFs: **50 %** of the CSFs account to 5% of the project's success

Findings and Discussion

The content analysis result is summarised in Table 3 as shown below.

Phase	Category	CSF
Business Phase	A	Identifiable Business Value (1)
	B	Clear and Manageable Project Scope (2)
Data Phase	B	Identification and Access to needed Data Sources (3)
	B	Combine Different Data Sets (4)
	C	High Data Quality (5)
	C	Data Security and Privacy (6)
Analysis Phase	A	Innovative Analysis Tools (7)
	A	Adequate Hardware (8)
	A	Analytical Skillset (9)
	B	Technical Skillset (10)
	B	Integration of new Solutions (11)
	B	Fast Delivering of Results (12)
	C	Cloud-based Solutions (13)
	C	Flexible IT-Structure (14)
	C	Visualization (15)
	C	Virtualization (16)
	C	Adapt Architectural Principles (17)
Implementation Phase	A	Information Strategy for Big Data (18)
	A	Big Data as Strategic Instrument (19)
	C	Interpretation of Analytical Results (20)
Measurement Phase	B	Clear Project Goal with Deadline (21)
	B	Measureable Outcome (22)
Overall Phase	B	Top Management Support
	C	Multidisciplinary Teams
	C	Independent Business Unit
	C	Iterative process model
	C	Outsourcing

Table 3: CSF and their Assignment to corresponding Project Phases

The most important CSF (A-level) for Big Data projects, according to the collected data, is the investment in the needed, novel and innovative tools. Different studies state that the Big Data challenges cannot be

tackled with traditional data analysis solutions. Gogia (2012, p. 3) for instances states "*Big data is about having the technology (…) to allow firms to make sense of huge volumes of data in an affordable manner.*"

In order to benefit from the new emerging analytical tools, investments in hardware are also needed. Although storage is a necessity the data collection suggests that simply throwing more storage at the problem won't work for Big Data projects. A flexible IT-Infrastructure, which can easily adjust to the problem scenarios, is needed. Scalability is the most important property of the analytical systems. The system has to be able to adjust quickly to the changing velocity of the incoming data and scale accordingly. Another important aspect extracted from the collected data is focusing on the business value of the projects.

Simply purchasing the newest tools and hardware will not lead to a successful undertaking. The project necessarily needs to support a specific business case. Having a clear vision on how successful projects contribute to the business is vital. Otherwise the projects run into the risk of discovering hidden knowledge within the organizations data that does not contribute to the business success. Therefore, Big Data projects need to focus on delivering a return on investment (ROI). Projects won't be successful, if they are just started for the sake of being innovative. If there are several business cases within an organization that would benefit from a Big Data analysis, the case studies suggest picking the business case which is promising the fastest ROI.

An additional process focused CSF, the case studies suggest, is the strategic use of Big Data. As explained, Big Data projects need to deliver a ROI. Having a clear defined strategy on how to run a Big Data project is one step towards this. The collected data indicates that this strategy has to be part of the more general data management strategy. This seems legit as Big Data should not replace traditional analysis methods, but extend their possibilities. Analyses based on structured data are still vital for the business. The Big Data focused part of the strategy has to ensure, the project is aligned with the business. This supports the need of aligning technology and people involved in the project to support the business goals.

Moreover, an overall Big Data strategy will also help organizations to predict where meaningful discoveries can be expected. The strategy also aims on easing Big Data challenges at early stages, for instance by avoiding the storage of important information in insular data silos. The second part of the strategic use is integrating Big Data in the organizations strategy. Understanding the business strategy and choosing the adequate analytical toolset accordingly is one of the major keys to Big Data project success. Furthermore, it will help organizations to draw up more general plans on how to leverage from information and cause, ideally, a change in management and organization culture regarding the importance and availability of information.

Another major CSF deduced from the case studies is the availability of analytical talents within the organization. Even the earlier explained implementation strategy is useless if it is addressing the wrong issues. The same goes with the usage of the latest technology, if the wrong questions are analysed. In order to avoid these traps, the availability of the right people is crucial for the project success. Concerning velocity for instance it could be pointless to have the technology and processes in place to deliver the results in real time, if the analytical talent is missing to get to conclusions based on the result.

Error! Reference source not found. visualizes the collected and categorized CSFs. It clearly shows that more process and technology based CSFs could be extracted from the case studies. The highest number of A-level CSFs can be found in the process oriented category. Nevertheless, technology is an almost equally important factor as the same total number of CSFs could be identified.

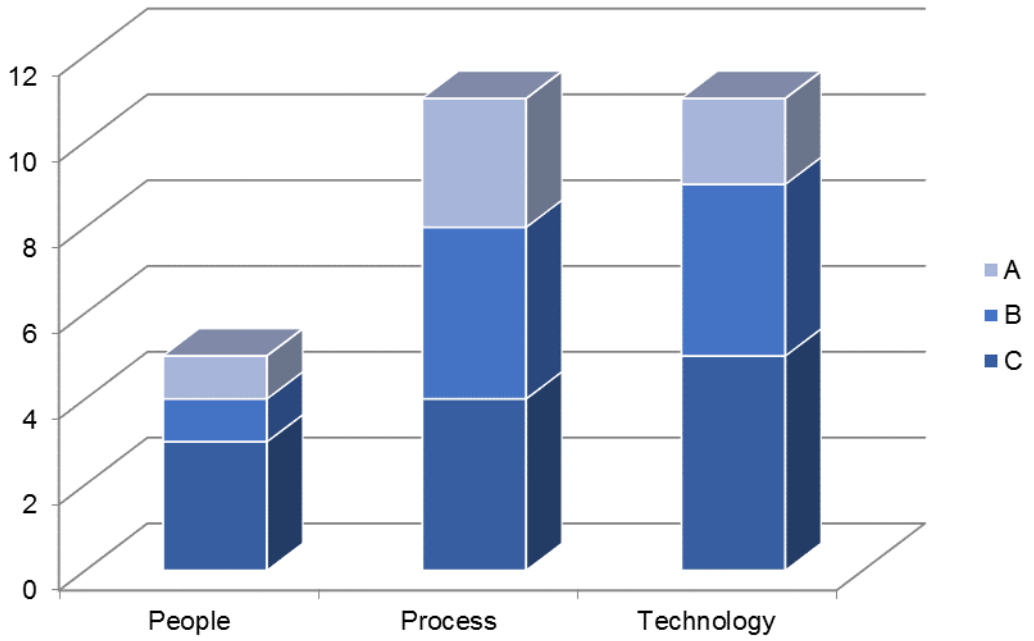

Figure 2: Distribution of A, B and C-level Success Factors in TPP Categorization

Figure 3 shows the general distribution of the CSFs over the runtime of the project, separated in the introduced categories *People*, *Process* and *Technology*. The graphic is not aiming on highlighting the importance of single CSFs within the project. It rather illustrates a general overview of the distribution of CSFs during the project's runtime, as well as their affiliation to a dimension. The size of the bubbles indicates the importance of the CSF. The biggest bubbles are A-level CSFs, the smallest C-level CSFs. The graph shows as expected that technical CSFs are most important in the middle of the project when the analysis is actually implemented and executed. People based CSFs are needed in that phase as well, but are also needed at the end for evaluating the results. Process focused CSFs are needed during the entire process to ensure project execution according to the schedule.

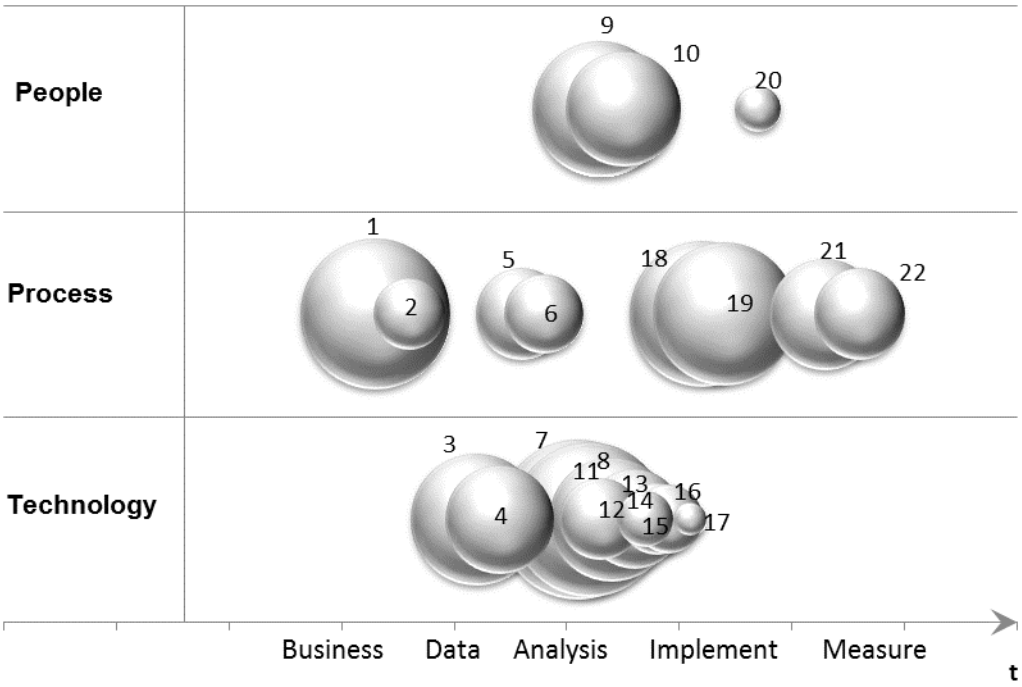

Figure 3: Correlation between Category, Time and Influence of the collected CSFs

Error! Reference source not found. 4 shows on a detailed level the CSFs in their categorization and their corresponding project phase.

Phase	People	Process	Technology
Business	-	Identifiable Business Value Clear and Manageable Project Scope	-
Data	-	Data Security and Privacy High Data Quality	Identification and Access to Needed Data Sources Combine Different Data Sets

Phase	People	Process	Technology
Analysis	Analytical Skillset Technical Skillset	-	Innovative Analysis Tools Right Hardware to Support Analysis Tools Integration of new Solutions Fast Delivering of Results Cloud-based Solutions Flexible IT-Structure Visualization Virtualization Adapt Architectural Principles
Implementation	Interpretation of Analytical Results	Information Strategy for Big Data Big Data as Strategic Instrument	-
Measurement	-	Clear Project Goal with Deadline Measureable Outcome	-
Overall	Multidisciplinary Teams Independent Business Unit	Top Management Support Iterative process model Outsourcing	-

Table 4: CSFs in individual Phases - TPP - Matrix

The findings also suggest that six CSFs have been identified that could not be assigned to a specific project phase, but affect the analysis project in general.

Top Management Support is one of those general CSFs. It is important to have a sponsor for the project on the executive level of the organization. Big Data projects are more likely to succeed when they address the core of the business strategy. The top level management is the position to decide about this overall alignment of an organization. Also, Big Data projects usually need to involve different business units. The top level management has the power to align all these units in order to achieve a successful project.

Having members from different departments is one aspect of another overall CSF: working in Multidisciplinary Teams. Besides the cooperation between business units, IT experts, analytical and statistical talents should get involved in the project team. General statements on the composition of the team are difficult, as this strongly depends on the purpose and the area of the project itself.

This analysis team can be either a temporary institution or an Independent Business Unit, which is specialized in analysis. The creation of such a designated analysis team is another suggested CSF from the case studies. Either way, the Big Data analysis should be an ongoing process, also when the team members are only temporary. In this case a process should be established to make sure to rebuild the team with adequate members every time this is needed.

Another C-level CSF, indicated by the analysed case studies, is the project execution according to an Iterative Process Model. As explained in the process model, the case studies suggest starting projects with a small scope, learning from the analytical results and getting bigger with every single iteration cycle.

Outsourcing is another overall C-level CSF suggested from the case studies to consider. In general, outsourcing has to be seen as a strategic partnership. Companies need to build up their own skills in order to succeed on a long term as Big Data is aiming for the overall organizational strategy.

Conclusion

This research addressed the issue of overcoming obstacles in Big Data projects. Although the importance of the availability of information is acknowledged by most of the organizations, they still struggle to succeed with Big Data analyses.

Besides this identification of success factors, this research provides a process model, as a suggested way to approach Big Data projects. The process model is divided into separate phases. In addition to a description of the tasks to fulfil, the identified success factors are assigned to the individual phases of the analysis process.

The research findings assigned success factors to individual process stages, which are categorized according to their importance for the success of the entire project. It is perhaps more important to know how the individual success factors contribute to the overall project success based on the process stage / phase. This understanding will help organisations better allocate resources in effectively managing current and upcoming Big Data projects.

REFERENCES

Berg, B. L. (1983). *Jewish identity: Subjective declaratIOns or objective life styles*. (Doctoral dissertation), Syracuse University, Syracuse, NY.

Berg, B. L. (2001). *Qualitative research methods for the social sciences (4th Edition)*. Needham Heights, MA 02494: Allyn & Bacon.

Bhashyam, R. (2011). Challenges of Handling Big Data. Retrieved 22.07.2013, from http://cdn.ttgtmedia.com/rms/pdf/Technology%20Challenges%20of%20Big%20Data.pdf

Busch, C., De Maret, P. S., Flynn, T., Kellum , R., Le, S., Meyers, B., Palmquist, M. (1994 - 2012). Content Analysis. Writing@CSU.

Cloudera. (2012). Ten common hadoopable problems. 16. Retrieved 11.08.2013, from

Cuzzocrea, Alfredo, Song, Il-Yeol, & Davis, Karen C. (2011). Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! *ACM (?)*, 3.

Dai, J., Huang, J., Huang, S., Liu, A., & Sun, Y. (2012). THE HADOOP STACK: NEW PARADIGM FOR THE BIG DATA STORAGE AND PROCESSING. *Intel Technology Journal, 16*(4), 20.

Davenport, T.H., & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century.

Gantz, J., & Reinsel, D. (2011). Ext ract ing Value f rom Chaos. *IDC iView*, 12.

Gartner. (2013a). Big Data. Retrieved 22.06.2013, from http://www.gartner.com/it-glossary/big-data/

Gartner. (2013b). Critical Success Factors. Retrieved 22.06.2013, from http://www.gartner.com/it-glossary/csf-critical-success-factor/

Giannikas, V. (2011). Six Tips for Students Interested in Big Data Analytics. ACM, 1.

Gogia, S. (2012). The Big Deal About Big Data For Customer Engagement.

Gopalkrishnan, V., Steier, D., Lewis, H., & Guszcza, J. (2012). Big Data, Big Business: Bridging the Gap. 5.

Gualtieri , M. (2012). The pragmatic definition of big data. Retrieved 11.08.2013, from http://blogs.forrester.com/mike_gualtieri/12-12-16-technopolitics_podcast_the_pragmatic_definition_of_big_data_explained

Gualtieri, M. (2013). The Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 2013.

Howie, T. (2013). The Big Bang: How the Big Data Explosion Is Changing the World. Retrieved 28.07.2013, from http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/04/15/the-big-bang-how-the-big-data-explosion-is-changing-the-world.aspx

Kelly, J., & Kaskade, J. (2013). CIOS & BIG DATA What Your IT Team Wants You to Know.

Laney, D. (2001). 3D Data Management: Cotrolling Data Volume, Velocity, and Variety. 4.

Laney, D. (2012). Big Data Strategy Components: Business Essentials. *Gartner*.

Lysons, K., & Farrington, B. (2005). *Purchasing and Supply Chain Management*: Financal Times Management.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity: McKinsey Global Institute.

Russom, P. (2011). Big data analytics. *Tdwi best practices report, 4*, 38.

SAS. (2013). Big Data – What Is It? Retrieved 02.05.2013, from http://www.sas.com/big-data/

Sathi, A. (2012). Big Data Analytics: Disruptive Technologies for Changing the Game. *1*. Mc Press

Sicular, S. (2012). No Data Scientist Is an Island in the Ocean of Big Data. *Gartner*.

Tankard, C. (2012). Big data security. Network Security. Volume 2012, Number 7 .