BIG DATA PROJECT SUCCESS – A META ANALYSIS

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

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

Sven Selle
Erfurt University of Applied Science, Sven.Selle@stud.fh-erfurt.de

Follow this and additional works at: http://aisel.aisnet.org/pacis2014

Recommended Citation
Koronios, Andy; Gao, Jing; and Selle, Sven, "BIG DATA PROJECT SUCCESS – A META ANALYSIS" (2014). PACIS 2014 Proceedings. 376.
http://aisel.aisnet.org/pacis2014/376

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2014 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
BIG DATA PROJECT SUCCESS – A META ANALYSIS

Koronios, Andy, University of South Australia, Adelaide, Australia, andy.koronios@unisa.edu.au
Gao, Jing, University of South Australia, Adelaide, Australia, jing.gao@unisa.edu.au
Selle, Sven, Erfurt University of Applied Science, Erfurt, Germany, Sven.Selle@stud.fh-erfurt.de

Abstract

The past ten years has delivered breath-taking advances in technology. The vision of an ‘intelligent’, ‘instrumented’ and ‘interconnected’ world is emerging. This frenetic activity is leaving a huge vapour trail of data. So much data is generated and collected today traditional enterprise repositories are currently challenged and the trailblazers of this new revolution had to develop new ways to store and analyse it. The dawn of ‘Big Data’ has arrived. Although much has been written about different types of definitions and the benefits of Big Data; very little has thus far been published on how to succeed with Big Data. This paper reviews this gap and attempts to propose a preliminary set of factors that are critical for success with Big Data. The aim is to identify factors that are critical for a project’s success in this area. In order to identify these factors a number of published case studies from Big Data projects were collected and evaluated produced a preliminary list of factors. This set was further analysed using ABC Analysis to produce six critical success factors. Further research is proposed to validate this set of factors through case studies.

Keywords: Big Data, Text Analytics, Big Data Success, CSF, Success factors.
1 Introduction

Modern organisations are currently surfing the wave of technological innovation which is itself reaching an inflection point. The convergence of internet technologies, mobile applications and an explosion of sensor technology is creating sweet spots of innovation. Individuals and organisations are leveraging these technologies and in the process generating torrents of data. A new term has been coined to describe this deluge of data that is confronting organisations today. Big Data has now reached hype proportions.

It is often found that the term of “Big Data” means different things to different people, most agree that the high volumes of data generated by organisations and collected from external sources is challenging their ability to store it and to manage it, let alone exploit it for enhanced business value. This together with the complexity of the large variety of different data formats as well as the speed, at which it is gathered, often in real or near-real time, has spawned a new vocabulary describing the many attributes of Big Data. Everyone is talking about the four, five and even more ‘V’s of Big Data.

Although much has been written about the growth of Big Data and the opportunities for monetizing this resource, little research exists to how to start the journey of Big Data. Furthermore, although Big Data oriented projects are still emerging, it appears that many suffer similar problems as other typical IT projects with many failing to achieve the project management performance indicators such (on time, in scope and on budget) and more importantly fall short of achieving objectives of the benefit realisation (Kelly & Kaskade, 2013). This research thus investigates lessons learnt from reported cases and provides guidance on Big Data project success.

2 Research Background

2.1 Adopted Big Data Definition

Big Data definitions abound (Schroeck et al 2012), Carter, 2011), Russom 2011) with the most often proposed as the situation when the high volume (size), the variety (structured, semi-structured and unstructured) and the velocity (high pace of data flows) of the data used in the organisation becomes a problem and challenges traditional information systems (Douglas 2012).

2.2 Big Data Challenges

Many traditional IT projects are now rebranded as Big Data projects. Vendors are promoting technology as the solution to Big Data problems. Technology plays an important role but is not central to success in these projects. In similar ways to traditional IT project success, Big Data success will most likely be contingent on addressing all the aspects of the People, Process and Technology model, and will face multidimensional challenges. Sicular (2012, p. 9) claims that these 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”.

Sicular (2012, p. 9)

This paper will explore Big Data success factors from these People, Process & Technology perspectives.

People

Current literature suggests skills for Big Data project implementation will be a major issue in its success. Big Data experts, often called ‘Data Scientists’ are very difficult to find and expensive to hire. This is likely to become more acute as organisations increase investment in this area and due to the latency from university and other education and training organisations in meeting the skill demands. This exacerbated by the nature of the required skills for Big Data which require not only
computational and analytical but also domain specific communication and storytelling skills. (Davenport & Patil, 2012).

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). Data scientists usually have a background in software development and computer science, but they also do have basic analytical skills such as statistics and hypothesis testing (Giannikas, 2011, p. 9).

Furthermore, most Big Data projects will involve multidisciplinary teams rather than rely 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. More importantly, the manipulation of large data sets often sourced from a variety of sources can pose challenges in data security, data quality and trust as well as privacy and ethical issues. It is therefore necessary to have roles such as risk management and legal services early in the project stages. (Sicular, 2012, p. 27).

Process

Unlike traditional projects, Big Data projects typically have two major stages that are quite distinct and require different process approaches, exploration of data and exploitation of data. The stage is that of exploration after that data has been captured or collected from a variety of sources and the technical mechanics of collecting and storing it, for example internally captured operational and transaction data may be fused with externally available from social media sites to gain some understanding of customer behaviour. Exploration of data provides an understanding of the value and potential of the data. This may, for example, allow prediction of customer behaviour in certain environments. Although this activity is usually termed as business intelligence or business insight, it is about greater understanding and building new knowledge (Peppard et al, 2013).

The exploration activity of Big Data cannot be restricted to traditional project processes but requires similar ways to tackle it as those used in innovation projects. 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).

By definition the exploration of data will also reveal some unexpected results that themselves require creative thinking. Google for example was able to model the spread of the H1N1 virus before the health authorities through the Google searches for symptoms by the suspected patients. Some banks are utilising similar capabilities provide lead indicators of consumer demand before official statistics become available.

It is important that the results of data analysis are based on high data quality, as the actions resulting out of Big Data projects can have significant 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).

After initial exploration data could be used to Exploit the data and insights to take advantage of information asymmetries for creating business value and competitive advantage. This can be achieved through initiatives for operational optimisation, enhancing business decisions or even developing new business models.

Although the traditional data management process models are still relevant to Big Data projects, there are certain issues to be taken into consideration if these models are applied on Big Data projects. For example, the work by Sicular (2012) and Gualtieri (2013) points out that learning is an important factor in the planning process to have a foundation for future Big Data projects. Based on these work, the research has adopted the following Big Data process model when studying critical success factors. This model consists of a number of phases as shown in Figure 1:

- Business Phase: Conduct requirement analysis
- Data Phase: Collect data and prepare for analysis
- Analysis Phase: create hypotheses and perform analytics
- Implementation Phase: Implement actions based on analytics results
- Measurement Phase: Measure project success based on pre-defined metrics
- Learning: Collect feedback for future Big Data projects

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

Technology

The most significant difference of Big Data from traditional IT systems, and one which has been used as its de facto definition, is the technological challenges with Big Data. Indeed the original motivation for the original Yahoo’s paper for HDFS was the need to deal with massive data generated on the Web (Cutting & Cafarella 2005) and that type of data volumes challenged normal database systems resulting in the development of the distributed type storage and processing which became the basis of Hadoop, with HDFS and MapReduce as its core components (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).

Integrating data from different systems with different security levels, intellectual property and privacy settings into one system is causing significant security and privacy challenges 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).

2.3 Critical Success Factors in Big Data

Due to the emerging nature of this area and the dearth of application cases the success factors for Big Data projects is still developing. This research uses existing frameworks for success in the parent disciplines and begins to extrapolate to Big Data projects.

The use of Critical Success Factors (CSF) is a well-established approach to measure the performance of an organization ((Rockart, 1979), (Bullen & Rockart, 1981) and (Rockart & Crescenzi, 1984)). Rockart (1979) defined CSFs as

"the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization. They are the few key areas where 'things must go right' for the business to flourish."

CSFs can be valid for an entire industry, a single organization, a department within the organization or just for individual people within the organization fulfilling the organizational strategy. CSFs are powerful tools for managers to help them understand what must go right in order to be successful in achieving goals and objectives (Sim, 2003, p. 20).

In order to determine the CSFs of Big Data projects, success factors of traditional data management projects were analysed for this research. As data management projects are essentially IT projects, the success factors for projects in general and for IT projects may also apply to such projects. Based on a detailed analysis of current literature a preliminary Success Factors for Big Data projects has emerged. The use of empirics was subsequently collected to validate the literature findings.

As Big Data is linked to Data management, this area was also investigated. Data Management is defined by DAMA (Data management Association International) as “development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets” (Mosley, 2007).

While data management is often strongly focused on structured data stored in traditional database systems, Big Data extends beyond this. To some extent, Big Data has to be seen as part of data management, but due to its variety of possible sources some properties are shared with IT-projects in a more general context.

3 Research Design

3.1 Research Aim and Methodology

This research aims on identifying CSFs for Big Data projects by adopting a qualitative research approach. Qualitative research methods, on the other hand, can be defined according to Strauss and Corbin (1990, p. 17) as “any kind of research that produces findings not arrived by means of statistical procedures or other means of quantification”. In other words the focus in qualitative research does not lie on numbers, but on the understanding of words, opinions and experiences (Taylor & Bogdan, 1998).

This research used a qualitative research approach to extract a set of CSFs from published case studies from Big Data projects. Therefore, an understanding of the used concepts was essential in order to
determine which factors had the strongest influence on Big Data project success. Content analysis was the primary research method in this study. 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”. Text in this context is defined very broadly, for instance books, chapters of books essays, interviews, speeches, articles and so on. In order to be able to conduct the content of a text, the text has to be broken down into elements.

Table 1 shows the adapted, predicted and summarized CSFs for Big Data projects as a result of literature review (Data Management, IT management and Big Data). The CSFs have been deduced from the researcher from the general CSFs in general projects, IT-projects and data management projects literature. These CSFs have been adjusted according to the requirements arising from the Big Data challenges mentioned earlier. The Keywords column contains the keywords that cluster into the concepts of the individual CSFs.

It is acknowledged that the proposed CSFs may not be regarded “critical” as that may not play a vital / critical role in the actual Big Data project. Thus, a process / method of determining the significance of the final success factors is implemented, which will be discussed in the later section.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Success factor</th>
<th>Example of Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>• Educated analysts in text analysis methodologies are needed.</td>
<td>• Analytical, Statistical Skills</td>
</tr>
<tr>
<td></td>
<td>• Big Data technology skills within own staff or hired externally are needed.</td>
<td>• Technological Skills</td>
</tr>
<tr>
<td></td>
<td>• 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.</td>
<td>• Multidisciplinary team</td>
</tr>
<tr>
<td></td>
<td>• 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.</td>
<td>• Documentation</td>
</tr>
<tr>
<td>Process</td>
<td>• Identifiable value for the business is delivered from the project. The project must support the business needs.</td>
<td>• Identifiable Business Value, ROLTICO</td>
</tr>
<tr>
<td></td>
<td>• Creation of a clear vision and implementation strategy for the software and hardware components of the new analysis platform.</td>
<td>• Big Data Strategy</td>
</tr>
<tr>
<td></td>
<td>• Creation and disseminating of documentation needs to be a fixed part of the project.</td>
<td>• Documentation</td>
</tr>
<tr>
<td></td>
<td>• The growth of analysis data has to be forecasted to be able to react early when used systems need to be expanded.</td>
<td>• Forecast Data Growth</td>
</tr>
<tr>
<td></td>
<td>• Processes need to be established in order to ensure high data quality of the documents used for the analysis.</td>
<td>• High Data Quality</td>
</tr>
<tr>
<td></td>
<td>• A measurable project outcome has to be defined and in fact measured to evaluate the success of the project.</td>
<td>• Measurement, Defined Outcome</td>
</tr>
</tbody>
</table>
Strong leadership has to be provided throughout the project. The top management level of the organization needs to get involved.

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.

The size of the project has to be determined carefully.

Technology

- Awareness of recent developments in the Big Data analysis area, including evaluation, selection and implementation of new software methodologies.
- Document collection is a significant problem, only with the right documents provided the needed answers can be found within the data.
- Evaluation of the used hardware to make sure it meets the requirement of the growing data amount also in future.
- Developing, integration and application of logical data models for the new analysis platform.
- Performance has to be monitored continuously and the system needs to be tuned on a regular basis to meet operational goals.

Top Management Support, Executive Level Support, Leadership

Project Goal, Milestones, Deadlines

Clear, Manageable Project Scope, Adequate Size

Investment in new Technologies, Analytical Tools

Document collection, Access to Sources

Adequate Hardware, Storage, Real Time

Integration of New Solution, Coexistence

Fast Performance, Fast Delivering of Results

Table 1: Research Model as Basis for Coding of Data Collection (Developed for this research)

3.2 Data Collection

This research is based on secondary data from a variety of published sources. According to Berg (2001)

“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”

This research involved the collection of 60 case studies, which document experiences and learning 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 (the sum of all responses of all selected ones), 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.

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).
The collected secondary data is text based and collected from various sources in different file types. From the perspective of variety the used data for the research itself fulfills the requirements for the classification as Big Data analysis (unstructured text analysis from various sources). Even though the volume cannot be considered as high and the neither is the incoming velocity. Table 2 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. It is noted that the weighting of secondary data types is a natural result of the selection process without any special adjustments.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of Documents</th>
<th>% of Documents</th>
<th>Provider Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Case Studies</td>
<td>60 / 47%</td>
<td></td>
<td>Cloudera, IBM, SAS</td>
</tr>
<tr>
<td>Public Survey Reports</td>
<td>14 / 11%</td>
<td></td>
<td>IBM, McKinsey, TDWI</td>
</tr>
<tr>
<td>Guidelines</td>
<td>32 / 25%</td>
<td></td>
<td>Forrester, Gartner, McKinsey</td>
</tr>
<tr>
<td>Blog Entries</td>
<td>21 / 17%</td>
<td></td>
<td>Forbes, techrebulic, zdnet</td>
</tr>
</tbody>
</table>

*Table 2: Summary of Document Collection - created for this research*

### 3.3 Data Analysis

Based on the research findings, this research further uses the *ABC analysis* to categorize the extracted CSFs according to their importance for the project’s success. This content analysis was able to extract a range of potential critical success factors. The weighting of individual success factors was not considered in this study. 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.

The *ABC analysis* is an adopted method originally used for accounting from Dickie (1952). The method itself is based on the *Pareto Principle* also known as the 80/20 principle. Adapted to project terms, this rule claims that only 20% of the influencing factors are responsible for 80% of the project’s success (Koch, 1998). Hence, the other 20% of the project’s success are composed by the other 80% of the CSFs. This categorization of the CSFs is considered useful to put a special focus on the 20% of the CSFs that influence the project success the most.

By contrast to the two categories suggested by the *Pareto Principle* the *ABC analysis* introduces a third category, in order to allow a higher granulation. 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

To be able to assign one of the categories to the CSFs, a ranking among the extracted elements is needed. This research will rank the CSFs based on the number of projects that consider the individual factor as important for the project success. The underlying assumption is that those CSFs, which are associated more often with a success for a single project, are also more likely to be critical for a general success in Big Data projects. According to the *ABC analysis* concept, the A-level category contains the most important factors out of the extracted potential critical success factors.

It is acknowledged that there are alternative and more sophisticated ranking methods to determine critical success factors among success factors as compared to the ABC method. However, this study has a unique aim to investigate when a particular factor becomes “critical” or requires special attention during a particular Big Data project phase. In other words, a success factor may not be considered critical in the overall project, but may be considered critical in one specific project phase.
It is believed that such findings may contribute more to the current body of knowledge. Therefore, this study has used a relatively simple ABC method to determine factors that are considered “critical” for the overall project.

The initial analysis identified 20+ factors that could potentially be important in the success of Big Data projects. This preliminary list of factors were categorised using ABC-Analysis to provide some indication of the importance of these factors. Those factors which fell in the A and B category were considered more significant for the success of Big Data projects as shown in Table 3 below.

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

Table 3: CSF and their Assignment to corresponding Project Phases (Developed for this research)

The factors were further categorised in a two dimensional matrix, according to the types in the PPT model, (people, process and technology) at one level and according to the project phases of Big Data.
projects, (Business, Data, Analysis, Implementation and Measurement). The relative number is shown in Figure 2 below.

![Figure 2: Number of Critical Success Factors identified through the ABC Analysis. (Developed for this research)](image)

The factors were also examined to identify which of these are too general, such as top management support, communication and participant involvement which although very important, in many cases critical were not very specific to Big Data projects. These general factors were however also included in the final list.

The final list is shown in Table 3 below

<table>
<thead>
<tr>
<th>Critical Success Factor</th>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information Strategy for Big Data</td>
<td>The organisations information management strategy is important to its success in general but critical if it is to effectively leverage the opportunities of Big Data. Developing a strategy that thus aligns IT with the business to ensure technology and people are focused on the business needs is important.</td>
</tr>
<tr>
<td>2. Identifiable Business Value</td>
<td>The project must have a clear vision of how it may contribute to the business to either improve or optimise what is already being currently done or to create entirely new business value by doing new things through business model innovation from insights harvested from the Big Data project. The demonstration of ROI would be very important in the development of the business case.</td>
</tr>
<tr>
<td>3. Top Management Support</td>
<td>It is not surprising that this success factor was identified by all the cases that were examined for without such support and commitment most projects will either fail or not even get off the ground. Like other projects, strong leadership and top level sponsorship is critical. This however comes if there exists identifiable potential business value. The business case with demonstrable ROI will thus be important.</td>
</tr>
</tbody>
</table>
4. Skills for Big Data projects
A number of specific skills include:
- mathematical modelling;
- data analysis and visualisation;
- Technical skills including some programming skills;
- Domain knowledge;
- Communication and storytelling

These skills are rarely found in one individual and thus multidisciplinary teams are required involving business and domain experts, technical experts and analytics professionals.

5. Information Quality, security and Integrity
Information quality, security and integrity as well as ensuring the privacy of people is a very important activity of any successful organisation. This information management activity should thus be integrated into the Big Data project from the beginning rather than being an afterthought after the project has been implemented.

6. Technological Capability
Big Data projects are dependent to a large extent on new technologies to handle the huge amounts of complex data at speed. This however should not be the starting point but should be considered when the business goals are clear. Nonetheless some technological capability development in the early stages, particularly around the visualisation and exploration stages would allow the organisation to find out more about things that it may not even know; the unknown unknowns.

| Table 3: Critical Success Factors for Big Data Projects (Developed for this research) |
| In summary, 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. |
| In spite of new multimedia data sources, data (especially unstructured ones) still remains mission critical for analytical project. Throughout the literature a gap was detected between the possibilities of traditional data analyses methods and the needs arising from Big Data challenges. Moreover, little comprehensive guidelines were found on what is important for Big Data projects. |
| This research aimed to fill this gap by providing a project model on how to approach Big Data projects, as well as highlighting factors that contribute to the project’s success. |

4 Conclusion

This research has sought to identify the factors that must be considered in order to implement successful Big Data projects. Although much industry based commentary on the benefits of Big Data has already been published, little guidance has thus far been given on how to start and execute Big Data projects. A literature review has been carried out in order find out the current situation of organisations dealing with these new phenomena, classified as Big Data.

The review of literature provided an understanding of the challenges of Big Data. A number of Big Data project cases was identified and the learnings from such projects were analysed to identify success factors related to Big Data projects. These were classified using ABC analysis. The initial analysis identified more than 20 potentially relevant factors which were subsequently grouped and prioritised to reduce them to six critical success factors. The research revealed that success is not only
about the application of the newest technology. It rather indicates many other factors must be considered for successful implementation of Big Data projects. It is also acknowledged that the final selected success factors (especially the A ranked ones) may not present any surprises to academics and practitioners. However, this study confirmed that similar factors still play an important role in Big Data projects. More important, by linking the factors to the Big Data project phases, this study have identified when a particular factor should be focused more in order to achieve the overall success. This can be a considered as a unique contribution in this study.

Although this study contributes to Big Data project success, a number of limitations exist. A large number of cases were identified, however not all Big data projects are reported; Furthermore, the quality of secondary data may be questionable. It is anticipated that future research will collect primary data to validate these preliminary critical success factors for Big Data project success.
References

Peppard J, Koronios, A. and Gao, J. 2013 How to Start with Big Data & Analytics Initiatives, In print