“Channel shift – using data analysis to improve service delivery at the City of Edinburgh Council”

Table of Contents

[1 Introduction 3](#_Toc424301480)

[1.1 Context 4](#_Toc424301481)

[1.2 Objective of the project 4](#_Toc424301482)

[1.3 Thesis structure 4](#_Toc424301483)

[2 Background 5](#_Toc424301484)

[2.1 Big data 5](#_Toc424301485)

[2.2 Artificial Intelligence 5](#_Toc424301486)

[2.3 The role of design 5](#_Toc424301487)

[2.4 Data-driven design 5](#_Toc424301488)

[2.4.1 Data-driven design methods 5](#_Toc424301489)

[2.4.2 Examples 6](#_Toc424301490)

[2.5 Human-driven design 6](#_Toc424301491)

[2.6 Business Intelligence 7](#_Toc424301492)

[2.6.1 Introduction to Business Intelligence 7](#_Toc424301493)

[2.6.2 Review of recent works 7](#_Toc424301494)

[2.6.3 Business Intelligence and User Centered Design 7](#_Toc424301495)

[2.6.4 Architecture 7](#_Toc424301496)

[2.6.5 Data analysis in Business Intelligence 7](#_Toc424301497)

[2.6.6 Structured and Unstructured data 7](#_Toc424301498)

[3 Description of the work undertaken 8](#_Toc424301499)

[3.1 CRM data 8](#_Toc424301500)

[3.2 Mosaic data 8](#_Toc424301501)

[3.3 IBM Cognos 8](#_Toc424301502)

[3.3.1 Introduction 9](#_Toc424301503)

[3.3.2 Working with IBM Cognos BI 10](#_Toc424301504)

[3.4 Preliminary activities 14](#_Toc424301505)

[3.4.1 Meetings in the Council 14](#_Toc424301506)

[3.4.2 First iteration (proof-of-concept) 14](#_Toc424301507)

[3.4.3 Problems 25](#_Toc424301508)

[3.5 Work undertaken 28](#_Toc424301509)

[3.5.1 Analysis objectives (questions to be answered) 28](#_Toc424301510)

[3.5.2 Implementation 28](#_Toc424301511)

[3.5.3 Problems 28](#_Toc424301512)

[4 Analysis or Evaluation 29](#_Toc424301513)

[5 Conclusion 29](#_Toc424301514)

[6 Bibliography 29](#_Toc424301515)

# Introduction

an introduction to the document, clearing stating the hypothesis or objective of the project, motivation for the work and the results achieved. The structure of the remainder of the document should also be outlined.

Over the last few years, the School of Informatics has been collaborating with the City of Edinburgh Council in the area of open data in initiatives such as the Smart Data Hack and the Council's EdinburghApps hackathons. In the context of Edinburgh Living Lab, this relationship has broadened into investigating other areas of data science, and new kinds of collaboration. My MSc project is taking place within this context, and is focussing on bringing analytic techniques to bear on Customer Relationship Management (CRM) data that has been collected by the Council over the last year.

As one of fastest growing local authority areas in Scotland, Edinburgh is facing an ever increasing demand for Council services, outstripping the funds available to meet this demand. There are a number of projects on-going in the Council that try to address the resulting challenges, one of which aims to improve the way that Council interacts with residents, particularly in terms of dealing with complaints and reports of problems. At the moment, citizens can communicate with the Council using multiple 'channels': email, web forms, mobile apps, phone, post and face-to-face conversation. So-called "channel shift" is the policy of encouraging residents to use web forms in preference to other communication channels. In order to promote channel-shift, the Council is exploring a number of ideas, which are being developed across two distinct units, namely digital services and business intelligence. The Council has been recently building capacity to collect data and use sophisticated tools for managing and integrating it. My MSc project is hoping to contribute to internal resources for extracting business insights from analysing this data.

At this stage of the project, I have managed to carry out a few proof-of-concept analyses. One of these focused on a specific set of interactions, namely reports by residents of missed bin collections. One interesting question is whether there are patterns in what kind of people reports such problems and whereabouts in the city they are located. The Council uses the notion of 'user persona' to group residents into a number of categories, based on a combination of indicators including demographics, lifestyles, preferences and behaviour, drawing on data such as the Census, the Electoral Roll, house sale prices, and the ONS annual Expenditure and Family Survey. By combining CRM data with these personas, I was able to make clear who are primary users of the 'missed bin collection' reporting service, relative to each channel.

I am currently investigating a couple of further issues. One of those include cases where users tried using a web form, but for some reason switched to face-to-face or phone channel instead.

More broadly, I hope that my research will help the Council to ensure that transactions initiated via digital channels are dealt with effectively, as well as contributing to creating “success stories” and know-how within the Council for implementing transactions for services which do not yet support digital channels.

## Context

* Context objective: the Council has been recently collecting data in a CRM system, but at this stage it is not being analysed or used internally. The purpose is to create know-how within the Council, success stories to guide the use of CRM.
* Use Mosaic personas in the process
* Use BI tools and design practices in the process
* Work with CRM data and IBM Cognos

## Objective of the project

* Using data analysis try to answer questions, generate recommendations for implementation of web-forms, new transactions, improvement of existing web-forms

## Thesis structure

# Background

background to the project, previous work, exposition of relevant literature, setting of the work in the proper context. This should contain sufficient information to allow the reader to appreciate the contribution you have made.

## Big data

There are many definitions of what big data is and in some cases not only do they differ, but even stand in contradiction. This might be due to the fact that early cases of use of the term happened in different fields (Ward & Barker 2013) (Demchenko et al. 2014). Most commonly, big data is associated with data storage and data analysis, which in themselves are not new concepts at all. A description that is widely accepted as fundamental in coining the term big data is the “3 Vs” definition provided by Gartner in 2001 (Douglas 2001) (Ward & Barker 2013). Since then, the “Vs” description has been used and expanded (to “5 Vs”) by many (NIST 2015), (Demchenko et al. 2014), (McAfee & Brynjolfsson 2012), (Minelli et al. 2012).

The “5Vs” of big data are as follows:

* Volume – 90% of world’s data was generated over the last 2 years; by some, big data is considered when dealing with volumes over peta bytes (10^15)
* Velocity – more data being received than can be processed using “traditional” data analysis approach before a decision has to be made; processing of real-time data streams is becoming essential
* Variety – different types of data are being accessible (structured data, sensory data, social media data, voice recordings, photos, videos)
* Veracity (validity) – lack of control over quality and accuracy which leads to inconsistencies and incompleteness
* Value – how to get value out of data

It’s not just about having big data, it’s about finding patterns, getting analysis that will help make the right decision.

## Artificial Intelligence

https://www.youtube.com/watch?v=0qOf7SX2CS4

Proliferation of Big Data and AI into other fields, e.g. design.

## Design

### Data-driven design

* DeepMind: https://www.youtube.com/watch?v=EfGD2qveGdQ
* What would a cup say if it could speak? What would be the “optimal” design that we as humans cannot see?
* Genetic algorithms:
  + http://brainz.org/15-real-world-applications-genetic-algorithms/
  + http://www.geek.com/science/genetic-algorithms-create-a-better-race-car-555526/
* Amazon poster generated automatically to be personalised
* Web analytics

### Human-driven design

## User Centered Design

#### Review of recent works

#### Tools and practices

Design tools:

* Ethnographical methods
  + Interviews/surveys
  + …
* Usability inspection methods (as early as possible):
  + Heuristic evaluation
  + cognitive walkthroughs
* Usability testing methods (later on, after some ground was established):
  + Qualitative
    - In-person testing (so-called hallway testing)
    - Remote testing
  + Quantitative
    - comparison test (a-b testing)
    - Success rate
    - task scenarios
* prototypes
  + “low-fidelity” – as early as possible
  + “high-fidelity” – later on, after some ground was established
* participatory workshop (focus group)
* Think a-loud technique
* personas
* hackathon

Analytical tools:

* 5 why model
* Swot analysis
* Cost-benefit evaluation
* QUPER model

## Business Intelligence

### Introduction to Business Intelligence

### Review of recent works

### Business Intelligence and User Centered Design

### Architecture

### Data analysis in Business Intelligence

### Structured and Unstructured data

(Baars & Kemper 2008)

# Description of the work undertaken

this may be divided into chapters describing the conceptual design work and the actual implementation separately. Any problems or difficulties and the suggested solutions should be mentioned. Alternative solutions and their evaluation should also be included.

## CRM data

Its properties, etc.

For unregistered users there is an assumption about the postcode that they are close to home.

CRM documentation

- system description RightNow (Oracle CRM)

- description of the project in the Council, what stage they are at

CRM data documentation

- what information is collected

## Mosaic data

Mosaic documentation

- sources of data

- viability (accuracy)

- Structure of data

- How you can process

## IBM Cognos

### Introduction

Addressing the need of businesses for software helping to achieve a competitive advantage, IBM has a rich portfolio of analytics products. These include solutions in areas of predictive analytics, risk analytics, prescriptive analytics, enterprise performance management and business intelligence (IBM 2015). Majority of IBM products in BI belong to Cognos family and include very specialized applications like “Cognos Supply Chain Performance Procurement Analytics” as well as general purpose tools like “Cognos Business Intelligence”.

The solution used at CEC is IBM Cognos Business Intelligence 10.2.1 and it is a set of tools that significantly eases processes such as importing data from different formats (e.g. csv, xml, xlsx), combining relational and multidimensional data, generating reports (real time reports, drag-and-drop GUI, database queries in SQL and OLAP), scheduling and redistributing reports, publishing reports on multiple platforms and many more. Tools available at the CEC include: Report Studio, Query Studio,

The same results can be achieved using different tools, but each of them is better fitted for a specific purpose. Report Studio was designed with reports creation in mind, Query Studio was optimized for creating and editing complex database queries, etc. CEC has two types of instances of IBM Cognos – production and development machines, accessible under different URLs.

IBM Cognos BI is an enterprise class SOA platform (Browne et al. 2010). Its n-tiered architecture is made up of:

* + The web tier – provides user sessions connectivity to applications
  + The application tier – load balancing and processing of requests, managing storage of customer application data
  + The data tier



Figure 1 IBM Cognos BI architecture (Browne et al. 2010)

### Working with IBM Cognos BI

IBM Cognos can be accessed using either a web interface called IBM Cognos Connection or a Windows application. For the purpose of this project only web interface was used.

Web access point: IBM Cognos Connection, http://c-cog-dev-app-1/ibmcognos/

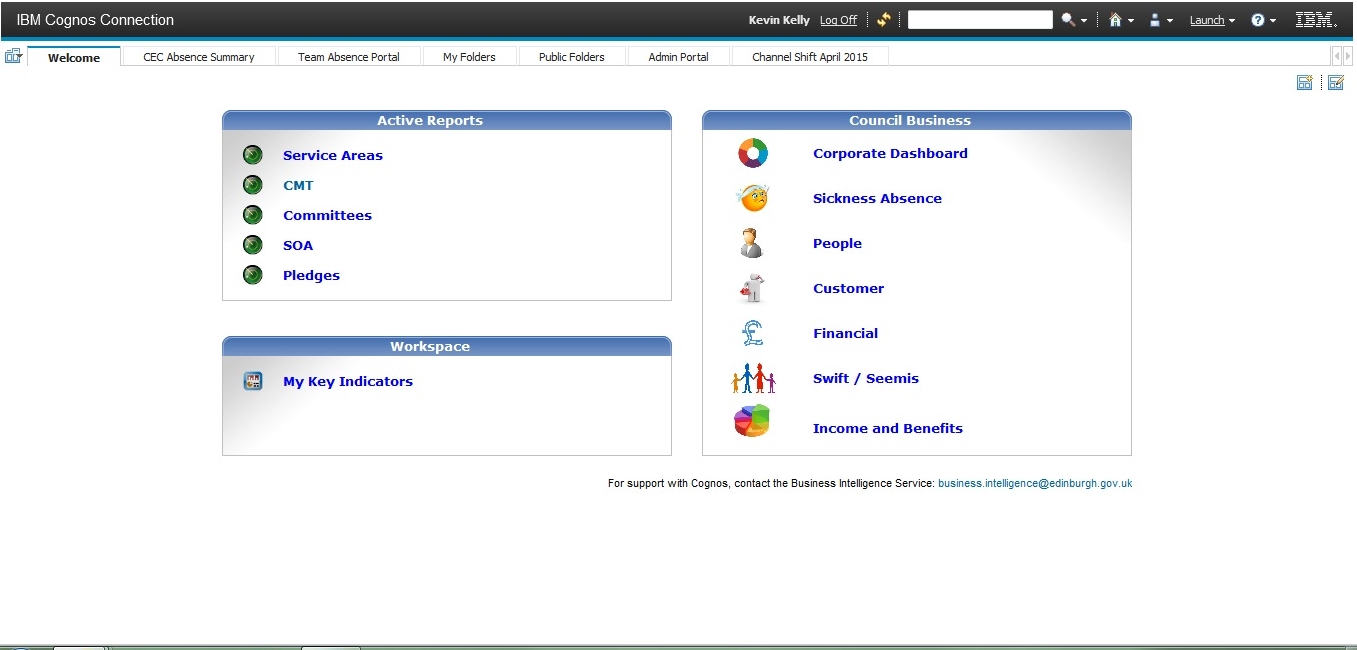


Figure 2 Welcome page of IBM Cognos Connection 10.2.1 (web interface to the entire package)

From this welcome page you can start applications available within your license, e.g. Report Studio. The first step after starting Report Studio is a selection of data package.



Figure 3 Select data package for Report Studio

After selecting the data package, one can either open an existing report or start creating a new one. In the latter case, a number of templates are available.



Figure 4 Report Studio welcome page



Figure 5 Select a template

The following figure shows Report Studio with a blank report and Mosaic data loaded.



Figure 6 Blank template, Mosaic data loaded

IBM Cognos help center: http://pic.dhe.ibm.com/infocenter/cogic/v1r0m0/index.jsp

google: ibm Business Intelligence Getting Started Guide

knowledge centre: http://www-01.ibm.com/support/knowledgecenter/#!/

workflow (cognos report studio guide, page 56):

Type of report you want to generate: Do you think about your data as a set of tables (relational) or a number of dimensions intersecting at cells (dimensional)

## Preliminary activities

### Meetings in the Council

* I developed a list of things that I could do
* Sally suggested a specific topic (personas and CRM)

### First iteration (proof-of-concept)

Proof of concept (first iteration, working with Kevin to see if it's even possible)

- we quickly exported some data from the CRM system

- tried to link it with Mosaic package

- failed

- I then got access to all the systems, I wanted to document what we tried to do and move on to other solutions (e.g. importing both datasets to MS Access, setting up a local database and populating it with the data , MS Excel, use Python scripts)

- I managed to solve the problem - create a new package

- played a bit creating reports



Figure 7 Creating query



Figure 8 First report - no analysis, plain CRM data



Figure 9 second query



Figure 10 page layout



Figure 11 first chart



Figure 12 first chart - data to confirm chart is valid



Figure 13 first chart - date filter

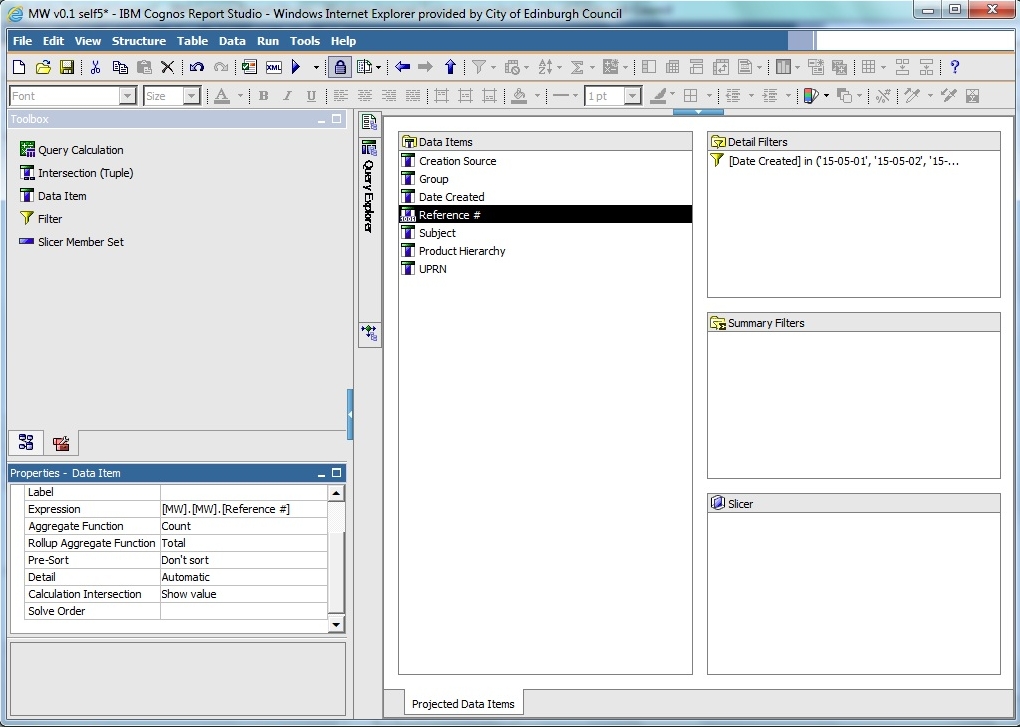


Figure 14 first chart - count total



Figure 15 first chart - other dimension



Figure 16 second chart – group category, all subjects, entire May, zoom in to legend



Figure 17 third chart - group code, all subjects



Figure 18 fourth chart - group code, missed bins, recycling bags (4 categories)

### Problems

#### Linking problem

I considered getting an extract from CRM and Mosaic, populating an SQL database (on my machine) and conducting analysis using Python libraries:

The datasets would be significantly limited, the entire project would not be as interesting to CEC.

IBM cognos link problem:

1. welcome page

2. Launch the "Report Studio" app from the list menu on the right

3. Select the Mosaic package

4. Start a new report

5. Select the type of report

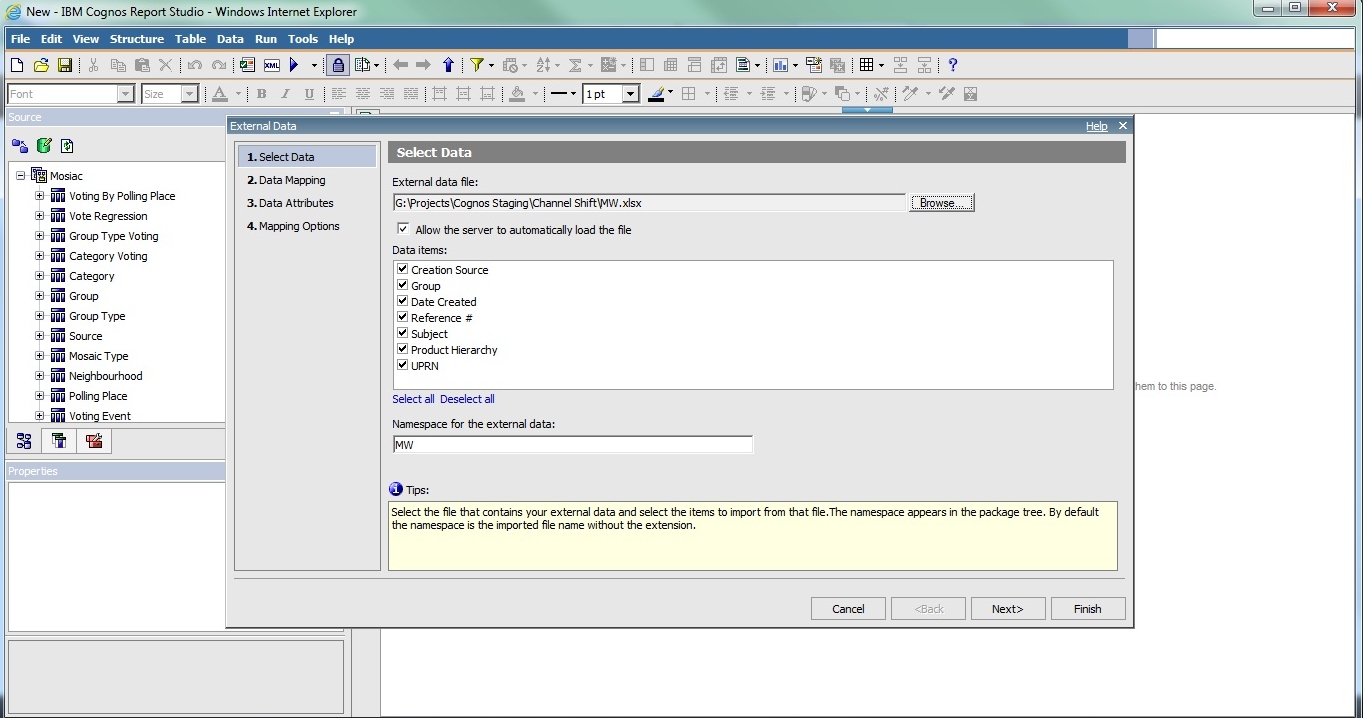


Figure 19 Importing CRM data into Report Studio with Mosaic loaded



Figure 20 Importing CRM data to Report Studio, Mosaic loaded, second step - link problem

We used IBM Cognos Report Studio. There was a package already created containing Mosaic data. There are different packages (see prtscrn). they can also be launched from the web interface (My folders, Public folders). It's just a different way of launching it.

We were trying to import data using Manage External Data and create links between the two. The package is not overwritten, there is another copy created containing Mosaic and CRM. Administrator sets the limit for the number of possible packages created (ask Kevin about the number, it was created by the 3rd party administartor).

https://www-304.ibm.com/support/knowledgecenter/#!/SSEP7J\_10.2.2/com.ibm.swg.ba.cognos.ug\_cr\_rptstd.10.2.2.doc/c\_rs\_cwa\_mng\_ext\_dta.html

the limit of 20000 entries is not Concil specific it's actually a limit on Cognos, source: https://www-304.ibm.com/support/knowledgecenter/#!/SSEP7J\_10.2.2/com.ibm.swg.ba.cognos.ug\_cr\_rptstd.10.2.2.doc/c\_prep\_ext\_data.html%23prep\_ext\_data

website with the solution:

https://www-304.ibm.com/support/knowledgecenter/#!/SSEP7J\_10.2.2/com.ibm.swg.ba.cognos.ug\_cr\_rptstd.10.2.2.doc/t\_id\_rs\_persdat\_161mapdata.html%23id\_rs\_persdat\_161MapData

#### CRM data inconsistencies

Problems to solve:

- when I work on a query and put a filter on data, generate report, save file, everything works fine. Then I use the same file to work on another report and start from removing the filter - when I run the report then the dates are still filtered.

- duplicate fields in CRM data ('UPRN', 'second UPRN', 'UPRN 2')

- what’s the difference between UPRN and UCRN

- what's the difference between 'subject' and 'product hierarchy'?

- duplicate values in field 'subject' ('Recycling bins or bags', 'Recycling bags or bins'). It's not that one value is not used and the other is. There are entries all throughout May for both.

#### Other problems

* Lack of pass to CEC, Not having a laptop from CEC with network ID
* Lack of access to Cognos
* Lack of access to CRM (RightNow)
* Not knowing (and not knowing who knows) the structure of the database in CRM system
* In many cases there was no knowledge of what was done a few months ago (e.g. “there are some profiles in the Council” – it turns out those were Mosaic profiles)

## Work undertaken

Description of what I did, code, screenshots

Final stage (second iteration):

1. Analyze the CRM data set, extract the data I'm intersted in

2. Create a package in IBM Cognos that would contain CRM and Mosaic

3. Create queries for the data analysis

4. Create visualisations

5. Conduct user studies to confirm some of the assumptions

6. Create recommendations/suggestions

### Analysis objectives (questions to be answered)

After generating simple reports which proved capabilities of systems avialable, a list of analyses to be conducted has been created:

* Identify cases where one problem has been reported through multiple channels
* Identify primary users of CEC services (e.g. above 3 issues reported, which groups report the most within a service, which service is most popular within a group, who are primary users in general)
* Provide evidence for results of learning new behaviour by citizens (information campaign)

### Implementation

### Problems

# Analysis or Evaluation

results and their critical analysis should be reported, whether the results conform to expectations or otherwise and how they compare with other related work. Where appropriate evaluation of the work against the original objectives should be presented.

# Conclusion

# Bibliography

Baars, H. & Kemper, H.-G., 2008. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. *Information Systems Management*, 25(2), pp.132–148.

Browne, D. et al., 2010. *IBM Cognos Business Intelligence V10. 1 Handbook*, IBM Redbooks.

Demchenko, Y., De Laat, C. & Membrey, P., 2014. Defining architecture components of the Big Data Ecosystem. In *Collaboration Technologies and Systems (CTS), 2014 International Conference on*. pp. 104–112.

Douglas, L., 2001. 3d data management: Controlling data volume, velocity and variety. *Gartner. Retrieved*, 6.

IBM, 2015. IBM Business analytics. Available at: http://www-03.ibm.com/software/products/en/category/business-analytics [Accessed July 9, 2015].

McAfee, A. & Brynjolfsson, E., 2012. Big Data. The management revolution. *Harvard Buiness Review*, 90(10), pp.61–68. Available at: http://www.buyukverienstitusu.com/s/1870/i/Big\_Data\_2.pdf.

Minelli, M., Chambers, M. & Dhiraj, A., 2012. *Big data, big analytics: emerging business intelligence and analytic trends for today’s businesses*, John Wiley & Sons.

NIST, 2015. NIST Big Data Working Group (NBD-WG). Available at: http://bigdatawg.nist.gov/home.php [Accessed July 10, 2015].

Ward, J.S. & Barker, A., 2013. Undefined By Data: A Survey of Big Data Definitions. *arXiv.org*. Available at: http://arxiv.org/abs/1309.5821\npapers3://publication/uuid/63831F5F-B214-46D5-8A86-671042BE993F.