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

Table of Contents

[1 Introduction 3](#_Toc425956542)

[1.1 Context 4](#_Toc425956543)

[1.2 Objective of the project 4](#_Toc425956544)

[1.3 Thesis structure 4](#_Toc425956545)

[2 Background 5](#_Toc425956546)

[2.1 Introduction to design 5](#_Toc425956547)

[2.2 User Centered Design 5](#_Toc425956548)

[2.2.1 Tools and practices 5](#_Toc425956549)

[2.3 Data-driven design 6](#_Toc425956550)

[2.3.1 Big data 6](#_Toc425956551)

[2.3.2 Business Intelligence 7](#_Toc425956552)

[2.3.3 Artificial Intelligence 7](#_Toc425956553)

[2.4 Double diamond 7](#_Toc425956554)

[3 Description of the work undertaken 9](#_Toc425956555)

[3.1 Discover 9](#_Toc425956556)

[3.1.1 Numerous meetings at the Council 9](#_Toc425956557)

[3.1.2 CRM data 9](#_Toc425956558)

[3.1.3 Mosaic data 10](#_Toc425956559)

[3.1.4 IBM Cognos 10](#_Toc425956560)

[3.2 Define 16](#_Toc425956561)

[3.2.1 Preliminary activities 16](#_Toc425956562)

[3.2.2 Designing of the solution 30](#_Toc425956563)

[3.3 Develop 30](#_Toc425956564)

[3.3.1 Analysis objectives (questions to be answered) 30](#_Toc425956565)

[3.3.2 Implementation 31](#_Toc425956566)

[3.3.3 Problems 34](#_Toc425956567)

[3.4 Deliver 35](#_Toc425956568)

[3.4.1 Presentation at the Council (evaluation) 35](#_Toc425956569)

[4 Analysis or Evaluation 36](#_Toc425956570)

[4.1 Evaluation of the tools used 36](#_Toc425956571)

[4.2 Evaluation of work undertaken 36](#_Toc425956572)

[4.2.1 Report 1 36](#_Toc425956573)

[4.2.2 Report 2 36](#_Toc425956574)

[4.2.3 Report 3 37](#_Toc425956575)

[4.3 Evaluation of methodology used 37](#_Toc425956576)

[5 Conclusion 37](#_Toc425956577)

[6 Appendix A 37](#_Toc425956578)

[7 Bibliography 40](#_Toc425956579)

# 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.

## Introduction to design

## User Centered Design

### 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

## Data-driven design

* 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

### 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.

### Business Intelligence

#### Introduction to Business Intelligence

#### Business Intelligence and User Centered Design

#### Architecture

#### Data analysis in Business Intelligence

#### Structured and Unstructured data

(Baars & Kemper 2008)

#### Organisation’s ability to learn

Materials from previous semester

### Machine learning

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

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

* DeepMind: https://www.youtube.com/watch?v=EfGD2qveGdQ

## Double diamond

http://design.gov.au/doing/projects/business-and-government-interactions-project/lost-in-translation-shared-striving-and-mutual-misunderstanding-in-business-and-government-interactions-findings-report-august-2013/

# 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.

## Discover

### Numerous meetings at the Council

a lot of meetings, interactions regarding ideas and systems at the Council

### CRM data

Its properties, etc.

- the data I'm working on is May extract, only entries with UPRN

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

22 registered in 04.2015

5224 registered in 05.2015

24671 in total in 05.2015

I am using the table “incidents”. Table “Answers” is like a “knowledge base” for consultants

* + General enquiry, no UPRN, in most cases to log activity on a channel, e.g. someone calling with a question
  + UPRN vs UCRN vs USRN

- I couldn't speak to someone knowledgable about the CRM system. As a result I couldn't determine things about the CRM dataset, e.g. difference between 'subject' and 'Product Hierarchy'

- the same thing is reported through 2 channels, but have different value in field "subject": web - "Recylcing bags or bins", phone - "Recycling bins or bags". Error in implementation? Product hierarchy value is correct - in both cases it's "Recycling Bins or Bags"

### 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/

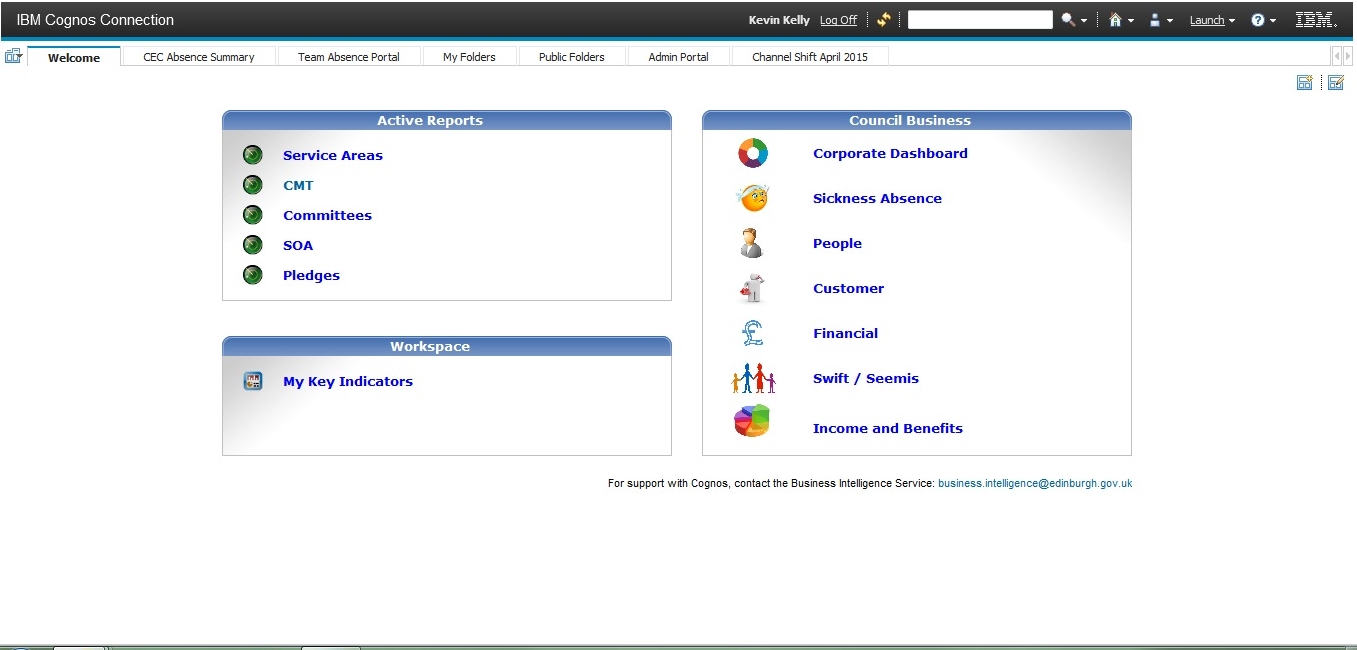


Figure 2 Welcome page of IBM Cognos Connection 10.2.1 (web interface to the entire package)

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)

Alternatives to Cognos include:

* Qlik http://global.qlik.com/uk
* CAFE – cognos analysis for Excel
* Tableu http://www.tableau.com/

Cognos seems to be dealing well with ETL things, e.g. exporting data from many varied formats, the initial processing stage of integrating data

- handy things about Cognos is that if you cahnge a field in one place it automatically gets updated elsewhere

- great source: http://kb.mit.edu/confluence/display/istcontrib/Cognos+Report+Studio+-+Filtering+a+Report

- if you remove an item from report it's permanently removed (e.g. when working on a chart). In most cases you don't want it to be removed completely from the report, just from the chart. In that case, use cut instead.

## Define

prototyping, discussing my ideas, trying to narrow down

### Preliminary activities

#### Meetings in the Council

1. I developed a list of things that I could do
2. 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

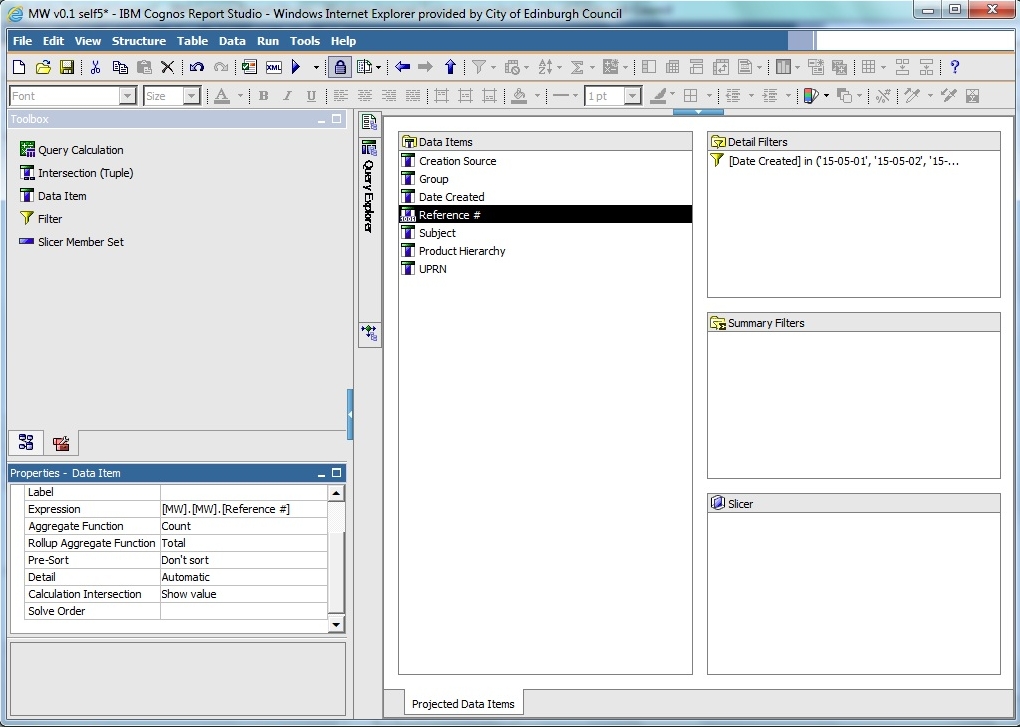


Figure 14 first chart - count total



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

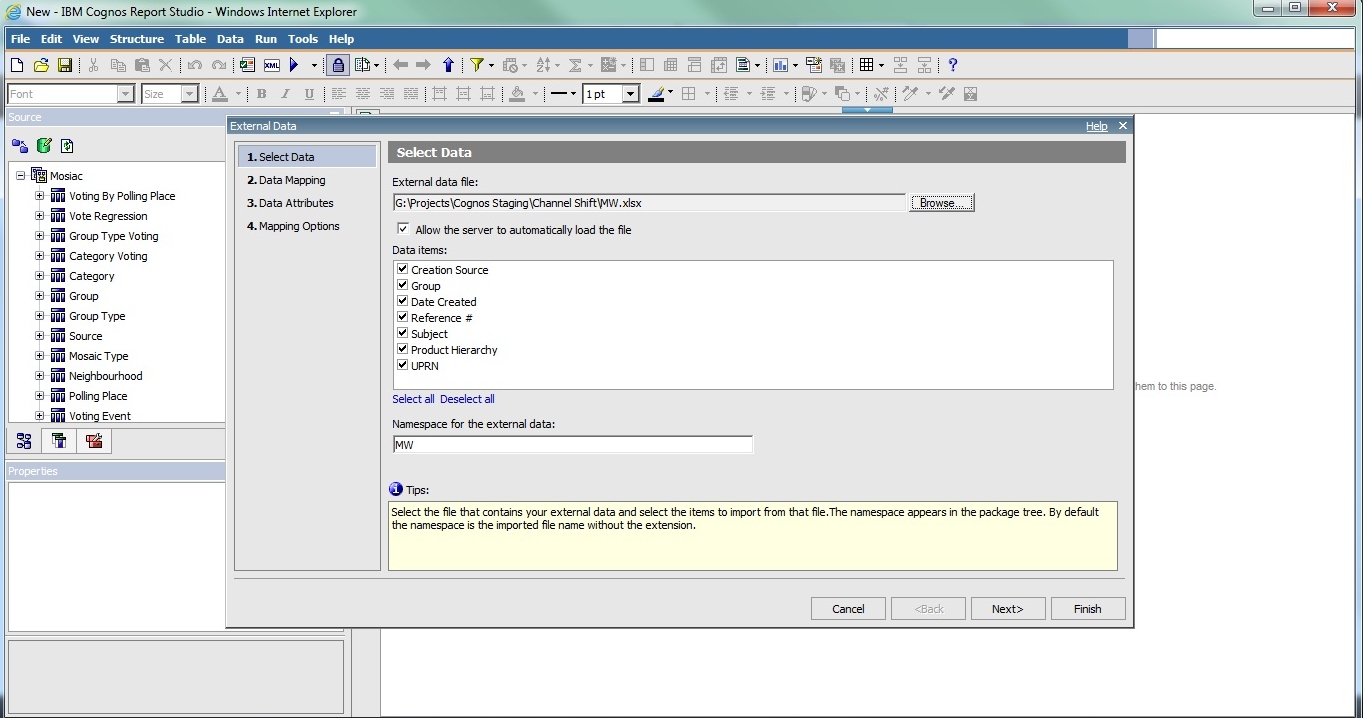


Figure 19 Importing CRM data into Report Studio with Mosaic loaded



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

1. Lack of pass to CEC, Not having a laptop from CEC with network ID
2. Lack of access to Cognos
3. Lack of access to CRM (RightNow)
4. Not knowing (and not knowing who knows) the structure of the database in CRM system
5. 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)

### Designing of the solution

## Develop

### Analysis objectives (questions to be answered)

The initial reports served as a learning experience during which I became familiar with systems available at the Council. I also got the necessary access rights and gained a better understanding of what kind of analysis is possible.

I had some ideas for the analysis that would be interesting from the perspective of a person implementing or improving transactions at the Council. After some discussions I created the following list of reports that were subsequently generated:

1. Cases of intentional use of multiple channels for the same issue
2. Patterns of behaviour across different channels
3. Who are the primary users of CEC services

The following subchapters provide more information on each of these points.

#### Report 1 - cases of intentional use of multiple channels for the same issue

This analysis is aiming at identifying cases where citizens want to report an issue, but do not trust in it being handled the same way through different channels. The reasoning behind such behaviour is that if many tickets are opened for the same problem, one of them will “get the job done”. It will be solved the quickest possible way, because if the process behind one channel has more resources available it will be handled quicker than with the process behind another channel.

The underlying assumption is that entries in the CRM system will not be identified as related to the same problem and that time of delivery differs across different channels.

The purpose of this analysis is not to provide evidence about the assumptions being right or wrong, but to verify if such behaviour exists among receivers of the Council services.

#### Report 2 - patterns of behaviour across different channels

The purpose of this analysis is to understand patterns of behaviour of citizens across many channels.

Some of the patterns that might be revealed include:

* The user initiated a service through a channel of preferred choice (e.g. web-form). However, after not hearing from the Council for some time, the user is unsure about the status of the process and uses a channel that is considered trust worthy (e.g. face-to-face) to confirm its state.
* If the above pattern occurs only for one type of transaction it might suggest a problem with a particular service. For example, if many users try to report a missed bin over a web-form, but eventually use a phone to do it (or switch after a few uses)
* An active user who uses a particular service is trying out a digital interface, but for some reason goes back to how he access it before
* Numerical evidence for how quickly people adapt new channels (e.g. how effective an information campaign was)

#### Report 3 - who are the primary users of CEC services

Designing is a task that should be conducted with the user in mind and having an understanding of who is the primary receiver of the design helps tremendously. For this reason, designers use “personas” which make it easier to know how the user thinks or behaves. The more detailed and accurate information about the user one has, the better design decisions one can make, which results in interfaces and services that better fit the needs of people.

The questions that will be answered within this part of work are trying to increase the understanding of users receiving Council’s services. In particular, 3 user groups are recognised:

* never used CEC services
* uses CEC services occasionally (defined here as having no more than 3 interactions with CEC)
* active user of CEC services (more than 3 interactions with CEC)

The analysis is trying to identify socio-economic backgrounds that users from all 3 groups have. Because CRM data contains only data about citizens who used CEC services it will not contain information about the first group. However, by determining who is interacting with CEC one can conclude who is not using those services. In other words, social groups that do not appear in the CRM dataset can be categorized into the first group.

Some of the questions that could be answered include:

* who are primary users in general
* which social group has the most interactions within a service
* which service is most popular within a social group

### Implementation

Final stage (second iteration):

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

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

3. Create queries for the data analysis (stages)

4. Create visualisations

5. Conduct user studies to confirm some of the assumptions

6. Create recommendations/suggestions

#### Report 1

It might put in the same category: people who made a call and a consultant suggested doing it on the web, who tried to fill a webform, but something went wrong and had to fill it in again (but system caught it as a seperate query) -> find cases where this behaviour occurs more than once. This was to identify malicious uses, but there weren't any, so the page 2 shows how much people "struggled" with web-forms

Report 1

Query 1: all data with an additional counter for how many issues someone filed on one day

Query 2: all data with a counter for one person, reporting on one day multiple issues on the same subject

Query 3: filter out: entries with channel not specified, with only one issue reported a day, with only one occurrence of multiple issues per day

(here you can see that 05-15 and 05-18 and 05-20 was probably the same issue and that diffferent values in subject column are, as I suspected, refering to the same thing, but just come from different channels - inconsistent implementations)

Query 4: number of issues on one day, by one person, on the same subject (last row is a problem with data, will be removed). In other words, there were 194 people in May who reported on the same subject, on the same day twice.

#### Report 2

This report is based on 2 counters. The first one provides information about total number of issues reported by a citizen (“count no of issues”). It is assumed that UPRN uniquely identifies the user. The second one counts number of separate channels used to report those issues (“channels used”). Then filters are used to remove from the report cases with number of issues below 2 (“count no of issues > 1”) and number of channels used below 2 (“channels used > 1”).

In order to make it easier to analyse the results, entries are grouped using UPRN – entries coming from one user are listed next to each other. Within this group, they are order using reference number (“Reference #”). Reference number is used instead of date because there are multiple cases where there is more than 1 entry during a day. In such cases, ordering by the date does not ensure the same sequence as the sequence of creation. Reference number on the other hand, consists of 2 parts: day of creation and a serial number assigned in an ascending manner. In the following example, both numbers were created on the same day, but the second one was created later which can be determined by the second part of the string:

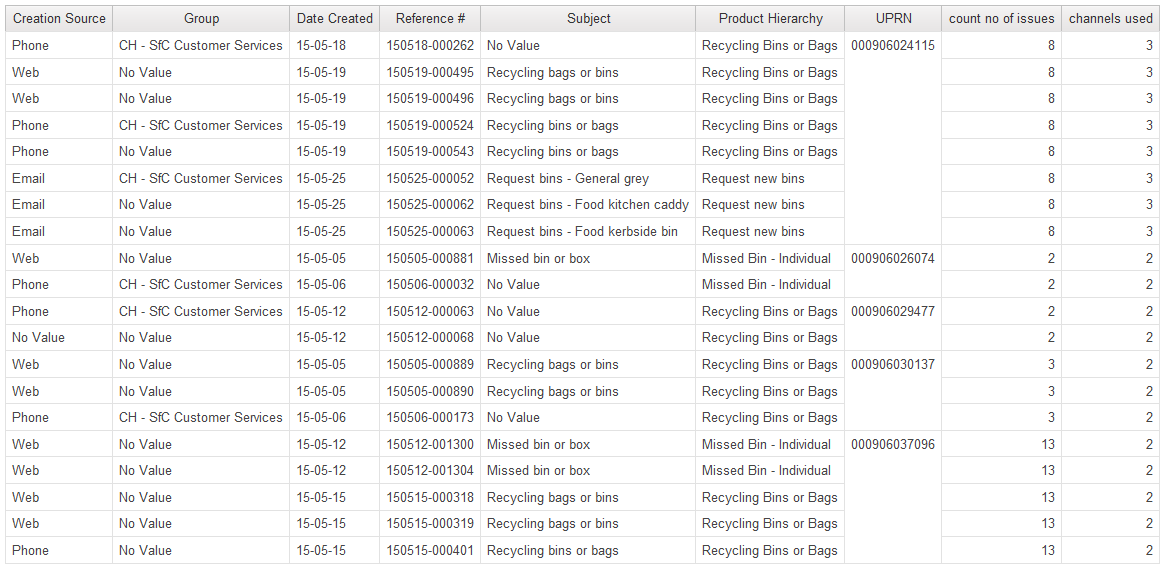
150511-000837

150511-000849

The diagram below presents part of the report that was generated.

Citizen with UPRN 000906026074 submitted a “Missed bin or box” request through the web-form on 5th May and made a phone call the next day regarding a service from the same “Product Hierarchy”. Considering this is the entire activity of this user, it can be speculated, that both interactions were about the same issue. In order to verify such claim these interactions would have to be investigated further, e.g. using comments left by consultants (unstructured data (Baars & Kemper 2008)). Such additional information would be extremely useful in determining reasons for the user following up over the phone.

Citizen with UPRN 000906037096 submitted 2 queries for the same issue on 12th of May. The next day he submitted again 2 queries over the web for a different service, but what is really interesting is that after submitting them he made a phone call to the Council. This is a great example of identifying a situation where someone was not able to achieve a desired result over the web. It might have been a simple question, but clearly the person made a phone call after going through the web-form. It might be the case that this information was missing on the web-form and other users needed it as well.



#### Report 3

Report 3

Query 1: group users into 2 categories as described above ("no interactions" will not have entries in incidents table so there will be only 2 categories: below 3 interactions, above 3 interactions)

Query 2: assign Mosaic groups to both categories (1 chart)

Query 3: services that both categories use (1 chart)

Query 4: Which group is the most active within a service (selectable field -> chart)

### Problems

- in report 2, is there a way to find all interactions regarding one issue? (e.g. someone filled a web form which created the entry, but then phoned about it - is it recorded somewhere?)

- formatting not working in outlook 365

- not enough space on network drive (I was always using somebody's else computer

- not having credentials to any of the services at the Council (I always had to use somebody else's. People were very uncomfortable with me using their credentals, different people had different access rights, I often would arrive at the Council just to find out that the person doesn't have access to what I did the day before and the person responsible for granting access is on holidays)

- not having a work station (I could only work using a spare workstation and that was when someone was present, e.g. I had to arrive in the morning when they did, leave when they did)

- difficulties creating charts using calculated fields

- when setting filter (Data Item1 > 2) I got a result (but when started the entire report from scratch it's working properly):

- self calculated values (in contrary to values from a database) do not work with automated aggregation function, you have to use none as the aggregation function

[screens]

- the source file had blank entries, this was showing up in the table as an empty category with 4k entries

[screens]

## Deliver

### Presentation at the Council (evaluation)

Give and describe a presentation in the Council

# 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.

These evaluations should be treated as part of a bigger “transformation” project. Identifying cases where users struggled with a web interface by CRM data analysis should be one of many ways of improving service used together with for example focus groups.

## Evaluation of the tools used

Describe evaluation of the tools used

## Evaluation of work undertaken

### Report 1

### Report 2

In report 2, the analysis would benefit a lot from using unstructured data. In many cases, analysis as conducted here will not provide definitive answers, but will pinpoint areas that can be improved: cases where citizens are struggling, not trusting. The information provided is very detailed, e.g. it will not only show that residents were struggling with a particular interface, it will also show whether all of them struggled or only a specific part of users. This can be then used to improve the design with very specific personas in mind.

In report 2, it is very difficult

- different values in the field subject depending on the channel (e.g. report 2)

It would be very useful to throw into Cognos more data sets:

* web analytics - It misses cases where someone wanted to report an issue, but in the end didn’t – struggled and failed. After that they went to another channel (e.g. phone), but the initial attempt was not registered in the CRM system.
* Unstructured data – probably exists in the CRM data, but in a different table. I used incidents, but there are dozens others
* Create capability to store unstructured data in the CRM system if it’s not there yet
* Stress to consultants to put as much information as possible, make a workshop for them
* Putting procedures in place that would improve the design of a web-form/service based on such case studies would increase Council’s capacity for improvement and learning.

### Report 3

## Evaluation of methodology used

# Conclusion

* you can go very deep with those analysis, e.g. I started with the goal to identify malicious uses of multiple channels, I generated 3 reports, I realized I could do a few more looking for how people learn different channels -> the rate at which people file multiple times within one channel goes down with time (they learn)
* There are many open questions and potential for further study

# Appendix A

Files:

• MW v0.1 self - file to start playing with Cognos, shows how to combine CRM data with Mosaic data

• MW v0.1 self2 - first attempt to create a chart, doesn't work

• MW v0.1 self3 - attempt to create a chart, the query itself is working (everything is correct from the sytax point of view), but data is unavailable (aggregate function? filtering error?)

• MW v0.1 self4 - debugging the self3 report, added a list to show filtered data returned from the query. Result: there are some entries after filtering, but not as expected

• MW v0.1 self5 - from the start, this time the results are as expected (type of result, number of result)

• MW v0.2, working chart, 3 days, type of report - day - first chart working as expected, it shows numbers about queries from all services grouped into Mosaic groups, counted by reference, limited to 3 days (x - mosaic group, e.g. B, G, K; y - count by reference, multiple columns - different values in 'subject')

• MW v0.2, working chart, entire May (previous was 4 days), Mosaic group - subject - second chart working as expected, more accurate; x - type code, e.g. A01, A02; y - count by reference; only 4 services - 4 columns; for the entire May

• MW v0.2, working chart, entire May, Mosaic group - subject - similar to previous one, not as detailed, x - Mosaic groups, e.g. B, G; y - count by reference; columns - different services (different values in 'subject'); entire May

Here implementation of questions started:

• one issue on multiple channels v1.0 - find cases of misuse of multiple channels

The CRM dataset I was working on was limited to May only, entries without UPRN were filtered. There were 979 entries left as a result

• MW report 3 = MW Report 3, blank chart - base report that can be used to generate different reports about number of interactions with the Council, couldn't overwrite the original one, created another one to have a more meaningful name

• MW report 3, chart 1 - both above and below 3 interactions on one chart

• MW report 3, chart 1.1 blank - template for the charts 1.1 and 1.2

• MW report 3, chart 1.1 - chart with only above 3 interactions, using filter

• MW report 3, chart 1.1 correct - chart with only above 3 iteractions, using filter and a list to show data (identical to the previous one, extended with the list, couldn't overwrite the previous one)

• MW report 3, chart 1.2 correct - chart with only below 3 interactions, with list

the last 3 charts did not have axis sorted, I figured out how to sort axis starting from chart 3.2

further parts of report 3 has 2 dimensions which have many entries. As a result it would have a lot of columns and would not be easily readible. So I decided to split it into 3 charts: groups x services; detailed groups x services; selectable service -> group

• MW report 3, chart 2.1 - Mosaic group x services

• MW report 3, chart 2.2 - Mosaic group detailed x services

• MW report 3, chart 2.3 group code - the use of the selected service across different Mosaic groups

• MW report 3, chart 2.3 group type code - the use of the selected service across different Mosaic detailed groups

• MW report 3, chart 3.1 - blank report that is used as a basis for the other 3.1 reports

• MW report 3, chart 3.1 above 3 - Mosaic groups that active users belong to

• MW report 3, chart 3.1 up to 3 - Mosaic groups that occassional users belong to

• report 3.1 - 3 charts on one page

working 27.07

• MW Report 2, 2.1 identify citizens who interacted multiple times

• MW Report 2, 2.2 multiple interactions through different channels

Recommendations for further analysis after 2.2: analyze the issues, what happened there, comments, any unstructured data around this issue that is available in the system, closed times and dates, when this person interacts with the council again ask them questions what was the problem.

The result might be that only people with multiple issues will be contacted, this does not have to be the case. It should be compared with no feedback at all and in this case you can identify people who could provide feedback with very high accurracy. As a counterbalance, the questions might be asked to people from the same social group.

Analyze a few examples:

• Someone starts on the web, the next day they call

• Someone start with face-to-face, then they open a number of tickets on the same issue through web the next day

# 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.