Emerging business intelligence at JGZ

Prins, W.

Award date:
2016

Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science
Architecture of Information Systems Research Group

Emerging Business Intelligence at JGZ

W. Prins BSc

Supervisors:
dr. ir. B.F. van Dongen
J. Murk MSc

Eindhoven, Friday 13th November, 2015

Abstract

The aim of this thesis is to identify the opportunities and possibilities for growth in the Business Intelligence domain for GGD Brabant-Zuidoost. To achieve this the current level of Business Intelligence maturity at GGD Brabant-Zuidoost is determined and a case study is performed. This case study uses currently available data to answer questions that local management of the Youth Health Care department has. Lastly the information gained from these two approaches is combined with literature on growing and improving structural Business Intelligence usage within an organisation. We conclude that GGD Brabant-Zuidoost is still in the starting stages of its Business Intelligence endeavour. Still questions could be answered for the local management which helped streamline parts of the process and gave more quantitative grounds for decision making. To further develop Business Intelligence within the organisation a change in culture will be needed to make sure employees throughout the company are on board, ready to provide a sense of direction and realistic expectations.

Preface

This thesis "Emerging Business Intelligence at JGZ" is the culmination of my time at the TU/e. It details my work at the JGZ department of GGDBZO focusing on the application of process mining techniques and Maturity Models. With this work my studies in the Master of Science program in Business Information Systems at Eindhoven University of Technology is concluded.

The work was undertaken under the supervision of Boudewijn van Dongen and Judith Murk who helped guide me through the process of writing a thesis and formulating my own approach to the subject and helping me stay on track. I am very grateful for their guidance during this time. I would also like to thank my colleagues at JGZ who were always open minded towards my research and I very much enjoyed working with them.

During my time at TU/e I learned many things and have grown as a person. Although it was tough at times and not all fun and games, I can say I enjoyed my time here. I would like to thank "Secret society the garden of knowledge" for making the thesis writing experience bearable. My friends with whom I shared many a barbecue, my fellow board members for all the unique experiences we shared. My family who was always there for me throughout my studies. Lastly my girlfriend Надя for helping me stay motivated throughout my graduation project and her constant support and enthusiasm.

Alea Iacta Est

Wessel Prins

Eindhoven, Friday 13th November, 2015

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The term Business Intelligence(BI) refers to the usage of models or data to assist with business decision making. The field took off in a big way with the emergence of computers in the business environment, making the collection of data easy. Companies collect humongous amounts of data but at the same time giving meaning to these mountains of data has become a challenge. New techniques are constantly being developed to try and find the useful information, the needle in the haystack, that is hidden within the data. Data mining techniques are prevalent throughout the field, but one of the more recent techniques to achieve this goal is process mining. Process mining does not only look at the available data but also takes the underlying business process into account. It is however very dependent on the quality of the data that is available at the organisation. To this end an analysis of the organisation can be made to determine the quality of data and the way its usage is incorporated in the organisation.

1.1 Company background

The GGD has been charged with the duty to perform a number of healthcare tasks for the municipality. These tasks all pertain to the field of public health care, creating awareness among the public, signalling of social-medical problems and determining whether or not a group or individual needs preventive health care. They are tasked with ensuring that individuals receive Health Care from the correct professional, in this have a guiding function. One of the major areas where it performs this function is the area of Youth Health Care, "Jeugdgezondheidszorg"(JGZ) in Dutch. For the purpose of Public Health Care, the Netherlands has been divided into 25 safety regions, with local departments in every region. To help coordinate between these 25 GGD departments there is the nationwide association GGD GHOR NL which is responsible for representing the interests of its members. It is also responsible for undertaking projects in the GGD's domain that require a more centralised approach.

Recently there have been some changes in the responsibilities for the domain of public health care. Where before the budget and responsibilities laid with central government, this has now been moved to the municipalities. The landscape of healthcare is rapidly changing, and organisations need to change as well to fulfil the new demands that rise from this. Furthermore to make sure healthcare stays affordable in the near and distant future processes must be optimised. In this research we provide information to this end by harnessing the power of the data.

For this thesis a case study was performed within the department of Youth Health Care at GGD Brabant-Zuidoost(GGDBZO), which has its headquarters in Eindhoven. It is responsible for the aforementioned task of facilitating preventive care and referring clients to the correct professionals. This task is for a large part performed in conjunction with schools, where children spend a large portion of their time. JGZ is responsible for children aged 4-17 and routinely invites all children multiple times starting in the second grade of primary school. To track the health

of the students a dossier is kept which contains all relevant data that the professional performing the consultation wrote down. This dossier is kept in a digital fashion within the system called KIDOS. The dossier contains both the diagnosis made, any appointments made with the client, whether or not the scheduled appointment took place and a reason why if it did not. Furthermore it tracks which account was used to input the information. Unfortunately a large part of the data is only available in plain text making it very difficult to analyse this information. Because of this, and the confidential nature of the information, we only analyse the appointments, their date and time and, who performed the data-entry for the appointment. Every year JGZ has a target for the amount of children that should have a routine check-up. This target is based on data supplied by schools in the region of the previous year, e.g. all children in 3rd grade should be seen for a routine check-up, schools make a list of all children in 2nd grade and it is assumed that the vast majority of these children move to 3rd grade at the same school. Adding up these numbers for all schools gives us the target for next year. JGZ makes a year planning based on this information, checks a few times a year whether or not the targets are met and at the end of the year a definitive answer to this question is given. This approach has some disadvantages, as there is no insight into the current state of affairs at all points in time, only at set intervals. Problems will therefore only become apparent at a later date instead of immediately, this make active steering and control of the process difficult. For this project an attempt was made to provide more insight into the current situation and ideally develop a system that can support this on-line insight into the state of the process.

1.2 Organisational Structure

The GGD organisational structure of the GGD shown in Figure 1.1. With the Executive board consisting of City Councillors from all municipalities that the GGD does work for. The organisation is divided into multiple pillars with a Sector Head overseeing the different teams within that pillar. Every team itself also has a manager that oversees the team. For the rest of this thesis the pillar JGZ is central to the research done.

1.3 Tasks of GGD

The GGD is a mostly government funded organisation, although it does perform some additional services the main part of its income comes from the municipality. GGD is hired by the municipality to perform its obligation with respect to public health and in this way the municipality is a client of the GGD specifying both the parameters of its task and holding the budget. On the other hand the citizens of the municipality are also the GGD's clients except now for the actual service it provides. The GGD is often seen as a semi-governmental organisation, as mentioned in Section 1.2 City Councillors are part of the organisational structure of the GGD. The GGD is however a semi public company and also does other contract work besides it mandatory tasks for which it is payed on a contract basis. This is different from the way funding is handled for its tasks appointed by law where funding is based on the number of citizens it has to serve.

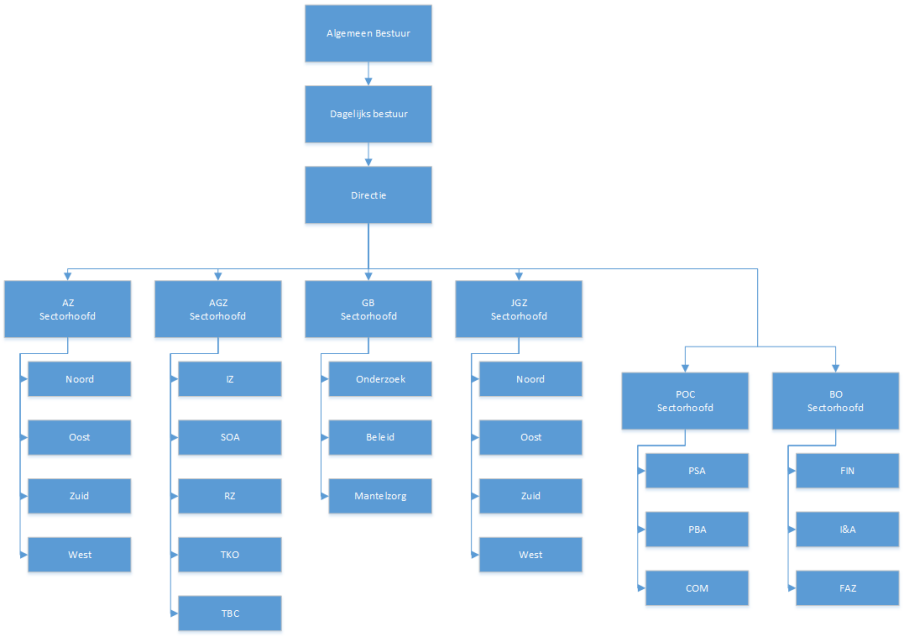

Figure 1.1: Organisational structure of GGD

1.4 Research goal

The goal of this research is to gain insight into the current state of affairs with regards to Business Intelligence at JGZ and, to give recommendations for future improvements. This is done by assessing the BI maturity at JGZ and applying selected process mining techniques to the available data with the goal of providing the business decision makers with useful information. Using the Maturity assessment and the results gained from applying process mining techniques recommendation are given and frequent pitfalls mentioned.
Problem statement: What are the opportunities and challenges for BI usage and growth at the JGZ department of GGDBZO, an organisation with emerging BI awareness?

To solve this problem the following research questions have been defined;

- What is the maturity level of the GGDBZO with respects to BI and data maturity?
- How can the available data be used to answer the questions that management has about the state of the process?
- What steps can GGDBZO take to advance its maturity in Business Intelligence?

Chapter 2

Related Work

2.1 Process Mining

Process mining is a technique designed to bring the two sides of Business Process Management together, namely the model and data.[16] The technique attempts to align the behaviour that is modelled in a type of Business Process Model and the data that is recorded by the organisations systems. Process mining focusses on the data pertaining to events that took place and its input data set is also know as an event log. An event log contains activities (i.e. real world activities that took place) which are related to cases, i.e. specific instances of a certain process. For example an on-line shop has an ordering process, an activity in this process could be "Check payment". While in this example a case could be a customer ordering a specific product. Activities must be distinguishable and unique so that the recorded data can be mapped correctly. More often than not additional information is used, e.g. the person or device, performing the action and, the time and date it took place. Process mining can be divided into three main topics: *discovery*, *conformance*, *enhancement*.

Discovery deals with discovering and generating a process model from the available data, this is done without prior knowledge of the actual process.

Conformance is the act of checking whether the process observed in the data conforms with the process such as it is detailed in the model or as specified.

Enhancement takes the results of conformance and attempts to improve the process or its model with the results gained.

Process mining is a fairly new technique but it can already be applied to provide organisations with operational support.

2.2 Maturity Models

Maturity models are a means of determining the level of structured approach to a specific topic within a company. We can look at maturity in a number of dimensions, but for this work the focus is on two types of maturity, namely data and process maturity. This is done to get a better grip on the current situation at GGDBZO and to help determine what possibilities are present. Later in Chapter 3 the level of maturity is of importance since it is a good way of determining whether or not a question can actually be answered at the current level of maturity.

Maturity models are a good way to gain an insight into the current state of affairs at a company for a specific subject. It is based on the principle that organisations "grow" from a starting state to some defined goal. During this growth the company goes trough many phases or levels that are pre-defined. Each of these levels has a number of metrics associated with it all of which have a baseline defined as part of a certain maturity level. This allows an organisation to look at itself critically and rate itself on each of these metrics, after which the corresponding maturity level can be ascertained. This is part of the descriptive nature of maturity models, it describes

a company's current characteristics. Besides this descriptive aspect, many maturity models are also of a prescriptive nature, prescribing what steps need to be taken to elevate the organisation to the next phase of maturity. This assumes that all organisations must go through a similar process in order to reach a set level. In practice this is often difficult, while the descriptive nature of maturity models is well explored, the prescriptive nature is less so. This is because change processes in organisations are much more complex than an evaluation of a status quo. Another limitation of general Maturity Models is that they do not take domain specific challenges into account since most of them focus on "the general company".[1]

The rest of this chapter is structured as follows, the two different types of maturity, BI and data maturity, are both explained separately following the same structure. A brief overview is given of what the type of maturity means and what it indicates.

2.3 BI maturity

Before going into the concept of BI maturity the definition of Business Intelligence must first be established. The field of BI is about 50 years old starting with decision support systems in the '60s which used computer assisted quantitative modelling to assist with decision making. But it wasn't until 1989 that the term Business Intelligence was used in its current meaning.[10] There are many definitions for the concept of BI some broad and others more narrow in their concept. For the rest of this work when referring to BI the following definition is employed, "Broad category of technologies, applications and processes for gathering, accessing and analysing data to help its users make better decisions".[17] Data quality is an important prerequisite for a well functioning Business Intelligence environment. It is important to note that implementing BI within the health care domain brings with it extra challenges.[6] Most research into the subject has been directed towards commercial companies but healthcare has its own unique challenges since it deals with many actors and has patient sensitive information.[7] During this thesis we briefly touch upon health care specific challenge but the main focus is on delivering relevant management information.

There are many BI maturity models.[11] We discuss Gartner's maturity model as shown in Figure 2.1. This is the first maturity model that is considered, as it focuses on building towards a BI and Analytics platform which will support enterprise-wide decision making and is constantly evolving together with the needs of the company. It details both IT changes that must be made as well as changes the users have to go through. The second model that we discuss is the TDWI Maturity Model by TDWI Research as shown in Figure 2.2. This model focusses on the technical aspects of BI maturity, while it is not restricted to this view. Together these two nicely complement each other and both are fairly well documented, which can be a problem for BI maturity models.[11] The goal of the Gartner BI maturity model is to assess the current state of the organisation and to provide handholds to talk to management with the intent of explaining the value of increasing BI maturity. The model scores the organisation along six dimensions; Business Drivers, Performance Measures, People and Culture, Processes and, Platforms and Tools. Table 2.1 shows the characteristics of the different levels for each dimension. All these dimensions are used to determine a current situation, it is important for all these dimensions to be at a similar level in order for the organisation to take full advantage of the BI environment.

The TDWI model offers interesting insights into the problems that most organisations encounter when first attempting to implement BI into their daily business. Since this situation matches the current one at JGZ it is interesting to look at these challenges from this perspective. The main point of the TDWI model is the idea that two big leaps need to be made by the organisation in order to get to a higher level. It details small evolutions for the organisation to get up to the correct level which enables the revolution that needs to happen to evolve to the next level. For JGZ "The Gulf" is the most relevant at the moment. The gulf is the result of the disparity between ambition and reality that causes a lot of initial BI projects to fail. Large BI projects undertaken by IT take many months to complete after which the results are already out of date and do not live up the managers expectations. This is caused by the fact that the business

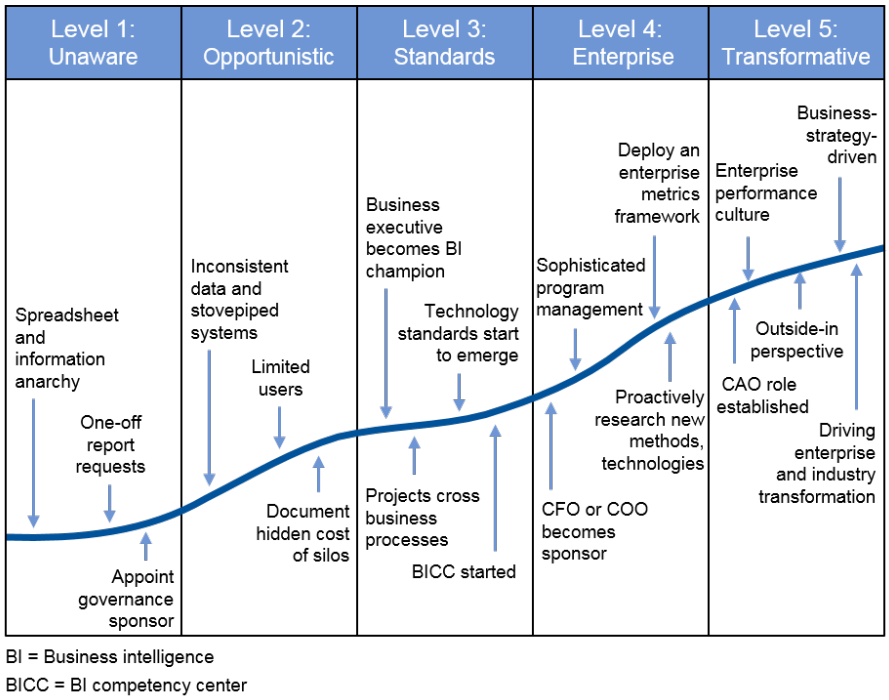

Figure 2.1: Gartner BI Maturity [5]

does not stand still while the BI environment is being developed and, perhaps more importantly, because it is difficult for the business drivers to exactly define what kind of data and information they need until they see it.

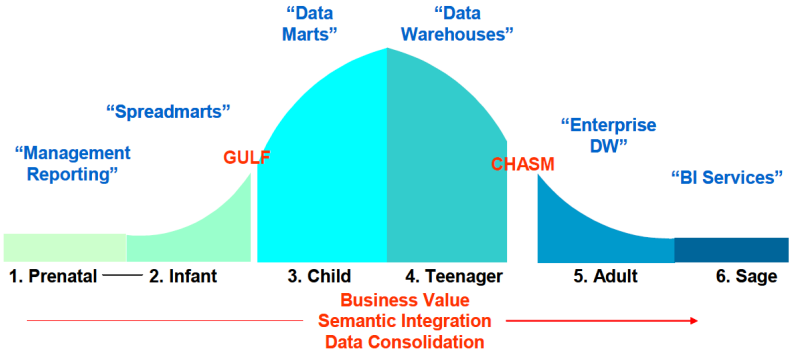

Figure 2.2: TDWI BI Maturity [3, 4]

Table 2.1: Characteristics of each Gartner BI maturity level [5]

Level	Unaware 1	Opportunistic 2	Standards 3	Enterprise 4	Transformative 5
Business Drivers	"CxO wants to know"	Single business function	Multiple business functions	Enterprise	Cross-enterprise and cross-industry
Performance Measures	Undefined	Single function Process optimization	Cross-functional Process analysis and management	Enterprise performance metrics and management framework	Leading enterprise and industry performance measures
People and Culture	Data analysts, Excel power users Gut-feel decision making	Functional managers and analysts Knowledge in the hands of a few	Executives, functional managers, analysts Fact-based decision making valued	Executives, functional managers, analysts, front-line workers, all decision makers Fact-based decision making encouraged	Customers, suppliers, partners Fact-based decision making assumed, negative results accepted rather than hidden
Processes	Undefined and uncontrolled No formal BI training	Silos Backlog of reporting requests, reactive Limited classroom training on tools	Shared, integrated Governance Primarily waterfall development, some agile Training on tools and data	Well defined, architected and governed, adaptive Balance risk and opportunity Agile Myriad training approaches	Inside-out and outside-in focus The digital enterprise
Platforms and Tools	Disparate ERP systems, spreadsheets, canned reports in transactional systems	Packaged or embedded, silos Custom ETL Reporting oriented, ad hoc query, tools, OLAP, limited	Standards, multiple applications, tools Robust data integration, MDM, data quality Dashboards Data discovery complements reporting tools	Consistent, integrated ERP system Multi-tier, multigeography hub-and-spoke architectures Logical data warehouse Advanced analytics	Interconnected information, process and analytic services across the value chain Pervasive data discovery and data preparation Innovation lab Use of cloud Analytic ecosystem
Program Management	None	Application- and project-focused	BICC	BICC Enterprise architecture Hybrid centralized, decentralized delivery	Strategic business initiative Value-based prioritization

2.4 Data Quality

Since most the case study research makes use of the data and information available at the organisation, it is important to look at the quality of the data. Data quality is a much discussed topic in research but the concept of data quality is a difficult one to capture since it is dependent on the organisation and the goal it has in mind for the data. When the business stakeholders are only interested in simple questions like "When are our employees logged into the system" the data only needs to contain timestamps and employee id's for it to be of sufficient quality. When this question gets more complex to, for example, also include "What do our employees do when they're logged into the system" the data is suddenly insufficient. For this research a more specific case of data quality is of particular importance, namely the quality of event logs.

In order to make use of process mining techniques it is important that a quality event log is available to apply the techniques to. To this end a brief analysis of the available event log is made to better determine what information could possibly be gained from the event log and what is at the moment still out of reach.

Event log maturity is described in Table 2.2. It details five levels of event log maturity, starting at logs constructed from a paper trail up to automatically generated event logs which are reliable and systemic. The model describes the current levels and their characteristics but does not explicitly describe ways of improving event log maturity, this is still one of the challenges in the field.

Table 2.2: Event log Maturity [2]

Level	Characterisation	Examples
⋆⋆⋆⋆⋆	Highest level: the event log is of excellent quality (i.e., trustworthy and complete) and events are well-defined. Events are recorded in an automatic, systematic, reliable, and safe manner. Privacy and security considerations are addressed adequately. Moreover, the events recorded (and all of their attributes) have clear semantics. This implies the existence of one or more ontologies. Events and their attributes point to this ontology.	Semantically annotated logs of BPM systems.
⋆⋆⋆★	Events are recorded automatically and in a systematic and reliable manner, i.e., logs are trustworthy and complete. Unlike the systems operating at level ⋆⋆⋆ , notions such as process instance (case) and activity are supported in an explicit manner.	Events logs of traditional BPM/ workflow systems.
⋆⋆⋆	Events are recorded automatically, but no systematic approach is followed to record events. However, unlike logs at level ⋆⋆ , there is some level of guarantee that the events recorded match reality (i.e., the event log is trustworthy but not necessarily complete). Consider, for example, the events recorded by an ERP system. Although events need to be extracted from a variety of tables, the information can be assumed to be correct (e.g., it is safe to assume that a payment recorded by the ERP actually exists and vice versa).	Tables in ERP systems, event logs of CRM systems, transaction logs of messaging systems, event logs of high-tech systems, etc.
⋆⋆	Events are recorded automatically, i.e., as a by-product of some information system. Coverage varies, i.e., no systematic approach is followed to decide which events are recorded. Moreover, it is possible to bypass the information system. Hence, events may be missing or not recorded properly.	Event logs of document and product management systems, error logs of embedded systems, worksheets of service engineers, etc.
⋆	Lowest level: event logs are of poor quality. Recorded events may not correspond to reality and events may be missing. Event logs for which events are recorded by hand typically have such characteristics.	Trails left in paper documents routed through the organization ("yellow notes"), paper-based medical records, etc.

Chapter 3

Maturity of JGZ

For a few years now JGZ has been experimenting with the use of Excel to gain a greater degree of insight into their processes. These spreadsheets are mostly based on planning information from the employees and the targets associated with this planning. They give insight into the yearly targets and the current standing towards these targets for groups of employees. These Excel files, containing information pulled from a multitude of different systems, are characterised by untrustworthy data and only give a very vague overview of the current situation with few ways of scoring the performance and generally lacking in-depth actionable information.

The efforts towards BI are still largely seen as belonging to IT with input being provided by JGZ management. The development of the system is likely to come from IT budget while managers are attempting to define what information it is they are looking for from this BI system.

As mentioned before there is no single system that contains all the information available to JGZ. Data is stored in different databases and tools which are bought from and run by an outside vendor. These vendors maintain control over the information and the tool. This leads to data specialists extracting the data they need from these databases and adopting spreadsheets containing this information in an effort to make better use of the available data. These spreadsheets are usually stored on network drives and in this way provided to the managers or sent to them via e-mail.

3.1 BI maturity at JGZ

The organisation is not yet used to working with BI and the advantages it offers. The division responsible for delivering reports is small and sits together with the IT division. Users are not motivated to make use of the systems and relevance of correct data entry is seen as low. The data analysts have attempted to make some proofs of concepts to demonstrate the possibilities it can offer. This approach however had the unforeseen result that management considers the created proof of concept to be sufficient for their endeavours and does not see a need for investments towards a more structural approach at this point in time. So while the relevance and some of the advantages that can be gained are seen, the willingness to allocate budget is very limited. At the same time what background systems are used in the organisation is in flux, many projects are ongoing to replace existing systems for CRM and patient files. These new systems are supposed to alleviate headaches of the current systems with regards to access and the ability to interface with each other. Concurrently to this process, a new division is becoming part of the GGD for which the IT department is scrambling in order to get everything in working order, this to ensure the required baseline for work is met. The BI and data-analyst projects and the organisations goals with regards to this are not yet well embedded within the organisation. Projects and developments depend largely on one key figure to lead the way. JGZ has two types of data, medical data and process data. For the purpose of process mining the process data was the most useful. This could be augmented by the medical data, this is however difficult and is a possibility that can be explored at a later stage.

3.1.1 Applying the Gartner model

Looking at the metrics as shown in Table 2.1 for each of the levels we position JGZ on the Gartner BI maturity model. The challenge in doing this is that the questionnaires used to perform such studies are not publicly available and can only be used if bought. The available information from the maturity models does however provide enough handholds to perform an analysis and give a reliable verdict on this matter. To further improve the reliability of the outcome we also apply the TDWI maturity model to the data. It is expected that both models give a similar result and that combining these two gives a better insight into the current state of affairs as well as a better path for improvement.

Business Drivers

The drivers for BI within JGZ are the managers who want to know more about the current state of affairs and get a better grip on what is happening behind the scenes.

Performance Measures

Performance measures are about how the organisation measures its business performance. At the moment JGZ only uses the simple measure, "is our quotum being met", no real optimization is being done.

People and Culture

The analysis is done by the data-experts and data-responsible, the analysis consists of spreadsheets containing information extracted from the databases. These Excel files are shared with the managers to help them make decisions. They have gut feelings about what the data should say and this does correspond with reality but managers are looking for more confirmation.

Process

An analysis of the process is more difficult for this instance since it's a very knowledge intensive process that can not easily be captured in process models.

Platforms and Tools

JGZ uses many disparate systems and spreadsheets at the moment to manage all their data, while the organisation is looking at new tools that allow for innate coupling of data, at the moment this is not yet possible.

Program and Management

Decision on what tools and programs to use are made on a project basis by central management. A small group of people is the driving force behind these projects.

3.1.2 Classification

All in all we determine that the maturity level of JGZ and GGDBZO as a whole is somewhere in between the unaware and opportunistic state with regards to BI maturity. There is no central system containing all the information and spreadsheet use is prolific. An overall BI plan is being designed but as of yet there is no centralised policy in place. Data experts are being asked to create custom reports via the use of spreadsheets based on unreliable data extracted from vendor bought databases. These reports are time and labour intensive while at the same time lacking the actionable data managers really need. Instead they give a current state of affairs overview of the available information.

Looking at the descriptions for the BI maturity levels for the Gartner model, similarities are seen with both the level 1 and level 2 maturities. The platform seems to be at a higher level than the users are at the moment, a distributed platform is available from which data can be extracted using SQL leading to Excel files consisting of aggregated data. The databases do not allow for in-depth analysis since they deal solely in live data. This leads to a very spreadsheet heavy approach. Meanwhile most of the organisation is unaware of the opportunities the data could provide, although awareness is increasing. Data reports are solely used by managers in an attempt to get better foundations for their decisions and are created every quarter using spreadsheets. This is done using a very labour and time intensive approach since all reports are hand-crafted using Excel. Data is collected as a by-product of users using the dossier application, but it is not well defined what data is and is not logged, at the same time the system does not appear to be designed with BI in mind. Important relations between appointments are missing and access-logging does not detail the actions performed, limiting the usefulness of the log.

3.1.3 Applying the TDWI model

For the TDWI model we have less exact quality dimensions to work with but we can go off the characteristics for each level as shown in Figure 2.2. From this we determine that JGZ shows characteristics of both Spreadmarts and Data marts. Data is stored in multiple purpose specific databases as described for Data marts. On the other hand, analysis of this data is for a large part done using spreadsheets. Data definitions and identifiers are not consistent across these different systems and a large portion of data is entered manually by users. This causes inconsistency in the data and overall low data quality which is analysed more in depth in Section 3.2. Primarily we conclude that JGZ is in the Spreadsmarts stage since one of the primary characteristics for Data marts, empowering knowledge workers and not just management has not yet occurred.

3.2 Event log Maturity

To assess the event log maturity we look at the available data which is extracted into two separate tables. Table 3.1 details the different actions that where scheduled to be undertaken in relation to a child. It does however not contain any access records corresponding to this planned action, these access records are stored in Table 3.2. A merging of these two tables is achieved based on “Gba Nummer”, “Start datum”, “Open Datum Tijdstip”, “Stam Nr”. This merging of data is very imprecise since the time stamps in “Start Datum” and “Open Datum Tijdstip” do not exactly match, instead an approximation is used in order to merge logging and action data. This approach yields a very rudimentary event log. This is further refined by merging log data points performed by the same actor for the same case within a time window into a single action. The first of these is considered the start and the last action that conforms to these constraints is considered the end event if it exists. Using this approach an event log containing, start & end events is created which have resources and a corresponding action. Later more information is coupled to this event log, this process is detailed in Section 4.2.

Table 3.1: Fields in the Kidos action table

BSN	↑ Gba Nummer	Personnr	Voornaam
Achternaam	Geboortedatum	Schooljaar	Actietype
Actietype Code	Datum	Streefdatum	↑ Startdatum
Status Label	Status Code	↑ Stam Nr	

Table 3.2: Fields in the Kidos logging table

BSN	↑ Gba Nummer	Personnr	Voornaam
Achternaam	Geboortedatum	Document Audit Id	Sluit Datum Tijdstip
↑ Open Datum Tijdstip	Tijd Date Opening Dossier	Tijd Opening Dossier	Uur Opening Dossier
Minuut Opening Dossier	Am Pm Opening Dossier	Dagdeel Opening Dossier	Dagdeel Volgorde Opening Dossier
Aantal Dossier Openingen	Seconden Geopened	↑ Stam Nr	

3.2.1 Applying the event log maturity model to the data

Using Table 2.2 to position the Event log Maturity of JGZ it is easy to determine that the logs are at least of level ⋆⋆ since the events are automatically generated as a by-product of using the information system. For the logging data it could be said that it is of level ⋆⋆⋆ since it is an access log it can be trusted that all accesses actually happened and the events match reality. There is however no systematic approach to which events do and do not get logged, sometimes an access has an end time but most of the time this information is omitted. Also not all actions performed find their way into the access log but some do, it is unclear how this selection is made.

3.2.2 Classification

Overall we classify the quality of the merged to be level ⋆⋆ since there is no guarantee that the information from the Kidos Action table is actually correct, this data is manually input by humans and could therefore easily contain errors which would in turn also appear in the merged dataset. In the end there is a certainty that events happened but it uncertain what those events were, this has to be guessed based on the time relation to actions which are uncertain.

3.3 Conclusion

In Section 3.1 and 3.2 we concluded that JGZ is still in the beginning stages of maturity. The organisation is not used to working with data in a BI fashion and systems are lacking. The existence of multiple systems and datasets, that cannot be directly related is an obstacle that must be overcome. We consider the organisation to be in between the unaware and opportunistic level of the Gartner BI maturity model and at Spreadmart stage in the TDWI maturity model. For data maturity we focus on event log maturity and consider this to be at level ⋆⋆. In Chapter 4 we look at all available data and construct an event log that is usable for process mining. This process will however be labour intensive because we lack instruments and high quality data. This means that it is very difficult to perform process mining and results are less trustworthy. Still we can apply the techniques and results are usable with reservations.

Chapter 4

Data Exploration

The dataset that was used for the study consists of planning data from the school years 2012-2013 and 2013-2014. From these school years only the data from primary schools was looked at, during which the two routine check-up in 2nd and 7th grade occur. The focus of this work is on the cases that are in 2nd grade, since this is the first time JGZ comes in contact with these children, there are more cases in which a follow-up appointment is deemed necessary. When the children are older the routine check-up has less follow-up and is therefore less interesting for the purpose of process mining, since only the activities are considered and not the content of those activities.

4.1 Exploring the available data

The investigation started by looking at what data was available and what part of this was relevant for the questions that needed to be answered. A first attempt at this was made by making a database dump of all the data concerning patient appointments in the areas of Eindhoven and Helmond. Using this small dataset an exploratory search was done to see what could be learned from the data. The data however, was not immediately usable in its raw form. There were separate files for the data-points that were logged by the system indicating when patient records were accessed by whom and the records of what actions were performed for the patient. For the data to be useful a relation needed to be made between the actions and the timestamps recorded by the access records. To this end a python script was created which related these two on the assumption that patient records are being accessed with the goal of performing some action for an appointment. The other assumption that was made, is that if the accessing of a patients record is close in the time dimension to an appointment with that patient, that the record was accessed for that appointment. Using these assumptions the python script was created which merged the two datasets based on this premises. At first only the datapoints that were within the timeframe of an appointment were related to the appointment, this however led to a very small percentage of the datapoints actually being related to actions. In an effort to try and add an activity to as many datapoints as possible, lax bounds were assigned to determine when a datapoint is related to an appointment. After some experimentation these bounds were set such that all datapoints 48 hours before and 48 hours after the appointment took place belong to the appointment. This approach gave better results but still a large amount of datapoints were left without an action tied to it. To remedy this the choice was made to introduce two artificial types of actions “Preparation” and “After care”, which both belong to a certain action. For example the action “BOGr2” has a corresponding “Preparation BOGr2” and “After care BOGr2” which concern datapoints that have no nearby appointment of which the action is relate-able to it. The creation of these datapoints is as follows; for each datapoint the appointment that is closest to it in the time dimension is found. If the timestamp of the appointment is more than 48 hours after the timestamp of the datapoint, the value “Preparation action” is assigned to the datapoint. If the timestamp of the appointment is more than 48 hours before the timestamp of the datapoint, the value “After care

action" is assigned. In this way all datapoints are assigned an action and it is expected that this grants a better insight into the goings on of the process.

4.2 Data Coupling

As mentioned in Section 4.1 the available data is spread over different files. In order to make use of all the data a coupling must be made between different data sources. This rest of this Section describes the different available data sources and how their coupling was handled.

4.2.1 Merging data points

At the end of Section 4.1 it is described how datapoints are assigned to actions. This gives a list of all datapoints with their corresponding action. It is however also useful to relate the datapoints that are about the same action and from this create a timeframe during which patient records were accessed for a specific action. To this end datapoints are merged into a single timeframe when they have the same "action id" as well as the same "Employee ID". From these the first and last timestamp are determined and the total amount of dossier accesses for this action is calculated. The timestamp of the first datapoint for this action and the timestamp for the last datapoint of the action are recorded and labelled "Start time" and "End time" respectively. This decreases the size of the dataset and results in log data that is usable for processmining.

4.2.2 Appointment data

Besides the data that was available from the patient records concerning appointments, there was separate data detailing the appointment records of employees. An attempt was made to relate the employee appointment records to appointments in the patient records. The goal of this attempt was to easily relate employees to appointments. This however ran into some problems, most notably that there was a disparity between the appointment times of the patient records and the appointments in employee records. This disparity could be anything from a couple of minutes to multiple days, a symptom of level ⋆⋆ event logs. This was a problem, especially since appointments can be closely followed by each-other in the instance that an appointment is first had with a doctors assistant before the doctor is seen. Because of the impreciseness of the data it was very difficult to make the correct coupling. Furthermore after one of the first feedback rounds, the possibility to add an employee id to the records from the patient records was discovered decreasing the usefulness of the coupling with appointment data. In light of this new discovery and the earlier mentioned difficulties of coupling the data, this approach was abandoned at this time.

4.2.3 Employee Coupling

After the attempt to couple the patient records to the appointment records failed a new way to add employee information to the dataset was needed. This was done by adding the account that was used to access patient records was added to its dataset. This account id could be coupled to the list of accounts with their corresponding account number "Stam nummer". This data is added to the patient record through an SQL query.

4.2.4 School Coupling

To compare the different schools and regions, data about the school needed to be added to the dataset. To achieve this, data from the "middoffice" dataset is used. This dataset contains additional information about the patient, which we easily relate to the patient records using the "BSN" available in both datasets. From this dataset the postcode, schoolcode and the city in which the school is located are extracted which can be used to divide the patients according to location and the school they go to. This is especially useful for the purpose of determining differences based on location.

4.2.5 Employee Team Coupling

Employees in the JGZ department are divided into smaller teams based on the region most of their work takes place in. These regions are “Noord”, “Oost”, “Zuid”, “West” as shown in Figure 1.1. To add this data to the patient record dataset a coupling must be made between the employee ID used in the patient record database and the employee id used for the work planning, which are unfortunately different, a clear symptom of low data quality. Since no corresponding id’s were available this required some manual steps in order to couple these two datasets. After close inspection of the datasets a correlation was found between the abbreviation used for an employee and an employee’s login name. An employee’s abbreviation is the first letter of its first name followed by the first two letters of it last name, if that is unique otherwise the first letter of the first name is used followed by the first and last letter of the employee’s last name. An employee’s login name follows a similar pattern, namely the first letter of an employee’s first name followed by its last name. So if an employee’s standard abbreviation is unique, it is identical to the first three letters of its login name. Using this a large amount of records is already be matched, while the non-unique ones require additional manual input to relate login name and employee abbreviation. Using this coupling, an employee’s team, manager and the amount of hours (s)he is scheduled to work can be determined.

4.2.6 Additional datasets

Besides the main dataset, some additional datasets are used to answer specific questions that the management has. The dataset “Triage + Uitvoering” contains for every patient the type of employee that should see him or her, i.e. whether a consultation should be performed by a doctor, doctor’s assistant or nurse. The dataset also contains what type of employee performed the actual consultation. From this it is extracted whether or not such guidelines are actually followed. Before this dataset could be used its data quality needed to be improved by sanitising it, especially the “performing professional” field. Different abbreviations were used for the Triage and the actual execution, these needed to be homogenised and some typos had to be corrected. For metrics that are concerned with the actual dossier openings, the patient record dataset without merged time-frames is used, since this gives a more precise overview of when exactly records were being accessed by employees. Which is used to gain a better insight into the working hours of employees.

4.3 Resulting dataset

The dataset that is used for analysis, unless stated otherwise, is the dataset that results from coupling all aforementioned datasets. This means that it consists of the access-log data coupled with appointment data as the base. Time-frames are merged to limit the size of the dataset and in an attempt to create activities that have a more meaningful duration. This dataset is extended with school and employee information that is linked to the events. The links between datasets contain an uncertainty since identifiers were not consistent among datasets and some approximations and manual is work was needed to complete the coupling. A anonymised snapshot of the data is shown in Figure 4.1 and 4.2.

Case ID	Activity	Resource	Start Timestamp	Complete Timestamp	Geboortedatum	Document Audit Id	Tijd Date Opening Dossier	Dagdeel Volgorde Opening Dossier	Aantal Dossier Openingen	Seconden Geopend	.....
Case 1	Voorbereiding OO	Value 21	53:29.0	30:27.0	Value 1	Value 49	Value 48	Value 1	Value 1	Value 3	.....
Case 1	Voorbereiding OO	Value 1	41:40.0	41:40.0	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	.....
Case 1	Voorbereiding OO	Value 3	18:48.0	25:59.0	Value 1	Value 50	Value 49	Value 1	Value 1	Value 3	.....
Case 1	Voorbereiding OO	Value 2	03:32.0	03:32.0	Value 1	Value 2	Value 2	Value 1	Value 1	Value 2	.....
Case 1	Voorbereiding OO	Value 1	25:52.0	25:52.0	Value 1	Value 51	Value 50	Value 2	Value 1	Value 14	.....
Case 1	Voorbereiding OO	Value 2	13:57.0	34:54.0	Value 1	Value 52	Value 51	Value 2	Value 1	Value 3	.....
Case 1	Voorbereiding OO	Value 3	01:45.0	01:45.0	Value 1	Value 3	Value 3	Value 1	Value 1	Value 3	.....
Case 1	Voorbereiding OO	Value 1	16:34.0	17:33.0	Value 1	Value 4	Value 4	Value 1	Value 1	Value 4	.....
Case 1	OOI voll BO bb	Value 2	37:47.0	29:43.0	Value 1	Value 54	Value 53	Value 1	Value 1	Value 15	.....
Case 1	OOI voll BO bb	Value 4	44:58.0	33:39.0	Value 1	Value 53	Value 52	Value 1	Value 1	Value 3	.....
Case 1	Nazorg OOI voll BC	Value 4	35:40.0	36:22.0	Value 1	Value 5	Value 5	Value 2	Value 1	Value 3	.....
Case 1	Nazorg OOI voll BC	Value 1	55:48.0	55:48.0	Value 1	Value 6	Value 6	Value 2	Value 1	Value 5	.....
Case 1	Voorbereiding Nal	Value 2	52:32.0	52:32.0	Value 1	Value 7	Value 7	Value 1	Value 1	Value 3	.....
Case 1	Voorbereiding Nal	Value 4	37:28.0	37:28.0	Value 1	Value 8	Value 8	Value 2	Value 1	Value 3	.....
Case 1	Nabespreken schc	Value 22	44:37.0	47:26.0	Value 1	Value 55	Value 54	Value 1	Value 1	Value 3	.....
Case 1	Nabespreken schc	Value 4	49:23.0	49:23.0	Value 1	Value 56	Value 55	Value 2	Value 1	Value 3	.....

Figure 4.1: A snapshot of the resulting dataset

.....	Stam Nr	actie Datum	actie Streefdatum	actie Startdatum	actie Status Label	cumul openingen	Functie	Locatie	Schoolcode	Soort	Postcode cijfers	Vestigings plaats school	verschilStartStreef	Postcode gebied
.....	Value 22	Value 1	Value 1	Value 1	Value 1	Value 3	Value 3	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 3	Value 1	Value 1	Value 1	Value 1	Value 4	Value 3	Value 3	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 2	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 1	Value 1	Value 1	Value 1	Value 1	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 2	Value 1	Value 1	Value 1	Value 1	Value 3	Value 2	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 3	Value 1	Value 1	Value 1	Value 1	Value 2	Value 3	Value 3	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 2	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 4	Value 1	Value 1	Value 1	Value 1	Value 3	Value 4	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 4	Value 1	Value 1	Value 1	Value 1	Value 1	Value 4	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 1	Value 1	Value 1	Value 1	Value 1	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 2	Value 2	Value 2	Value 2	Value 1	Value 2	Value 2	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 4	Value 2	Value 2	Value 2	Value 1	Value 2	Value 4	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 23	Value 2	Value 2	Value 2	Value 1	Value 3	Value 4	Value 5	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1
.....	Value 4	Value 2	Value 2	Value 2	Value 1	Value 2	Value 4	Value 2	Value 1	Value 1	Value 1	Value 1	Value 1	Value 1

Figure 4.2: A snapshot of the resulting dataset continued

Chapter 5

Process Mining

After the dataset was established work could start on the application of process mining techniques to answer the relevant business questions. To this end we mainly make use of two tools.

The process mining tool Disco developed by Fluxicon[1] with the goal of gaining fast insights into the supplied data. It automatically creates a process map from the data input and allows for straightforward filtering of the dataset using a set of default filter types. It gives detailed statistics of the entire process or the filtered result set that we want to focus on. It has build-in functionality for anonymizing the dataset and outputting results is straightforward. During this chapter it is frequently used to explore the process and expect areas of interest.

ProM, short for Process Mining framework[2], is an open source framework for process mining algorithms. It has a large number of available plug-ins that focus on different algorithms and techniques. This academic tool makes available many cutting edge techniques in the domain of process mining and during this chapter it is used for the application of more specific techniques.

5.1 Process Discovery

In order to gain a feeling for the data and to see what is happening behind the scenes according to the data, it was imported into the process mining tool Disco to get a better look at the process. To do this the data was formatted such that the data adhered to the basics of an event log; events, cases, timestamps. Furthermore start and end events were created for every action in an attempt to create a more useful log file. For this purpose a child is considered a case, an appointment an event and the user that in-putted the data the resource.

The first thing that springs to mind when seeing the process in Figure A.1 is that it is a huge spaghetti of events and transitions. Finding a clear flow or path from beginning to end is non-trivial. We are however able to use some of the knowledge that we have about the process to filter for parts that are deemed interesting. After consulting with the business stakeholders it was decided that the focus would be on the routine consult and the resulting follow-up appointments. This specific question gave us cause to filter the dataset for this smaller more defined part of the process enabling a better analysis. The rest of this Section therefore focuses on this part of the process as shown in Figure A.2.

5.1.1 How is escalation of the process handled?

The majority of the process is just routine where the consulting professional sees no need for any follow up appointments. For a small percentage however additional consultations are considered necessary by the professional. It would be interesting to know how the different types of professionals escalate the process and who performs these additional consultations.

[1] www.fluxicon.com/disco
[2] www.processmining.org/prom/start

Using Disco a filter is created which displays only cases that have a follow-up appointment after the routine consult, the result of this is shown in Figure A.2. These are considered to be escalated cases and are defined as cases in which a routine consult BOGr2 is eventually followed by any of the Investigation on Indication(OOI) appointments. For this metric only the 10 033 cases in which a BOGr2 routine consultation is performed are considered, which is 22% of the total 44 901 cases. Next filtering for all cases that are followed by an OOI, only 1930 of the cases remain, i.e. $1928 / 10\,033 * 100\% = 19.22\%$ of all cases that have a routine BOGr2 consultation are followed by an escalation. By using the combination of Action name and Employee function as the activity name it is straightforward to see by what type of professional an event is performed.

Table 5.1 shows the amount of routing consults which are being performed and by whom. Furthermore it shows how many of these consults actually get completed and how many cases were planned in total. From this it we see that in 98% of the cases when a routine consult results in escalation, this consult actually took place, which is to be expected. In the remaining less than 2% escalation was deemed necessary even though the consult did not take place. The reason for this needs to be further investigated, it could be a problem with the data or have another underlying reason.

In Table 5.2 the different combinations of professionals that performed the routine consult leading to escalation is shown. From this it is clear that, if an Arts performs the consultation alone escalation is rarely deemed necessary, while the Assistent and Verpleegkundige most often find cause for escalation. The amount of completed events here is higher than the total amount of events for certain combinations, this is explained by the combination changing if events that did not reach a completed stated are omitted.

Table 5.3 shows what type of professional performs the investigation after a case was escalated. For this additional consultation 90% of patients actually show up. It does not seem that there is a correlation between the type of professional and the patient not showing up. One noticeable point from Table 5.4 is that cases with more than two additional appointments are much less likely to reach completion, only 83% versus the 94% of cases that have only one additional appointment.

Table 5.1: BOGr2 before escalation

Performer	Total	Completed	Difference	Precentage Completed
Arts	626	607	19	96.96%
Assistent	1384	1321	63	95.45%
Verpleegkundige	750	733	17	97.73%
Cumulative	1928	1895	33	98.29%

Table 5.2: BOGr2 Variants by performers

Performer	Total	Completed
Assistent	682	693
Verpleegkundige	407	427
Assistent ->Arts	336	347
Assistent ->Verpleegkundige	103	105
Verpleegkundige ->Arts	52	53
Arts ->Assistent	48	50
Arts	40	39
Rest	260	181
Cumulative	1928	1895

Table 5.3: OOI performers after BOGr2 escalation

Performer	Total	Completed	Difference	Percentage Completed
Arts	1274	1145	129	89.87%
Assistent	703	630	73	89.62%
Verpleegkundige	757	680	77	89.83%
Cumulative	2734	2455	279	89.80%

Table 5.4: Amount of OOI after BOGr2 escalation

Amount of OOI	Total	Completed	Difference	Percentage Completed
Cases with 2+ OOI	786	657	129	83.59%
Cases with 1 OOI	1142	1082	60	94.75%
Cumulative	1928	1739	189	90.20%

Feedback

The results of this report were first and foremost an increased understanding of the process. While the managers have a lot of gut feelings about the process these results confirm some suspicion that they already had. Furthermore the insight that Assistenten perform a great deal of OOI cases was unexpected since these should mostly be performed by the Verpleegkundige and Arts.

5.1.2 What are the most common escalated activities?

The analysis is performed using Disco and applying a follower filter such that only cases that have "BOGr2" event which is eventually followed by some "OOI" are displayed. Using this filter it is easy to see which of these is the most common.

Looking at the resulting process map shows that "OOI15 Gehoor" is by far the most common escalated activity at 660 occurrences followed by "OOI15 Groei" which has almost half that many with 383 occurrences. A full overview of the escalated activities and their frequencies is shown in Table 5.5.

Feedback

It is interesting to note that OOI 15 Gehoor has the most occurences which was unexpected for the managers. The expectation was that the smaller or more broad OOI cases would have a higher occurrence since they apply to more general observations. As a result of this report a discussion has started with the goal of moving the BOGr2 out of the fall/winter season since the risk of

Table 5.5: Escalated activity frequency

Activity	Frequency
OOI 15 Controle	54
OOI 15 Gehoor	660
OOI 15 Groei	383
OOI 15 Houding	71
OOI 15 Motoriek	55
OOI 15 Visus	168
OOI 30 Controle	334
OOI 30 Gesprek	218
OOI 45 Gesprek	243

children having a cold or ear infection during this time is much higher, while not indicating a direct problem. Furthermore since the amount of small OOI occurences is pretty low, a general monthly consultation day for these OOI cases is considered to make planning easier and allow for a better fit with the customers.

5.2 Social Network Analysis

Social network analysis is the practice of analysing social structures using graphs. In such a graph the nodes represent a person or resource, while an edge between two nodes signifies that there is a relation between two resources. The nature of the relation depends on the network, a very simple way of thinking of a social network is by using the example of Facebook. Here a node is a person while a connection exists if two users are friends with each other. This same concept can be used to examine event logs with the goal of finding cliques that work together or a hierarchical structure of who hands-off work to whom.

5.2.1 What do the teams look like according to the data based on the locations they work at?

To answer this question the community detection plug-in for ProM is used.[15] Furthermore a modified version of the merged patient records dataset is used as input for the plug-in. The dataset is modified such that the action corresponds with the postcode of the patient, since the community detection tool that was used relates similar tasks, in this case it relates similar postcodes resulting in clusters according to this field.

The basic idea was to use social network analysis to detect the different teams that work for GGDBZO. For this the simple approach of relating employees that perform the same tasks does not work since all teams do roughly the same work. The differentiating factor is the locations that they work at. These areas are defined by the numbers of the postcode where the work is performed. Using this information the social network algorithm “Similar task” with the option “Correlation Coefficient” is used to extract the teams from the dataset. Where the “Task” that they perform is an action at a specific postcode, so Similar Task is defined as working at the same location because of our changes to the dataset. One of the challenges in finding the correct clusters is the fact that the degree between nodes greatly varies, causing clusters to be situated at very different levels in the set. To combat this the dataset is progressively filtered, first the most distinct clusters are identified and then removed from the dataset. This process is done multiple times until all clusters have been identified in a satisfactory manner. This an anonymised version of this process shown in Figures 5.1 until 5.6. Where Figures 5.1 and 5.2 use the same dataset, after this the detected smaller clusters are filtered and the remaining data is used to create Figures 5.3, 5.4 and 5.5. After this a last filter step is made and the remaining data is clustered as seen in Figure 5.6. A larger version of these images is shown in Appendix B.

Feedback

The feedback from this question led to a better understanding of the dataset, and showed that the communities that were found corresponded nicely with what the managers already knew about teams. An outlier, which almost exclusively worked alone, was also detected by the algorithm and this also was recognizable by the manager. This gives a clear indication that the communities that were detected in this way are meaningful.

5.2.2 What are the characteristics of the different teams?

Using the teams discovered in Section 5.2.1 and assigning a letter to each team it is possible to couple this data to the already existing employee information and use this to deduce additional information about the teams. Using this extra data the same analysis techniques as described in

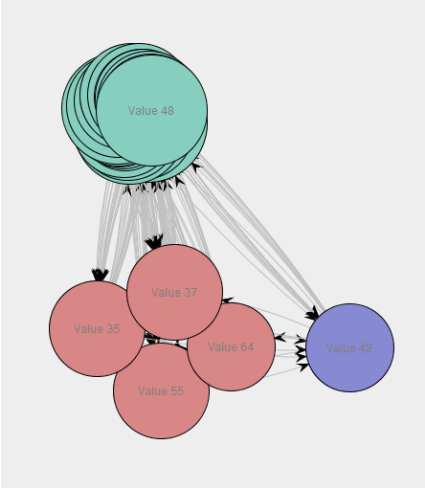

Figure 5.1: Cluster step 1

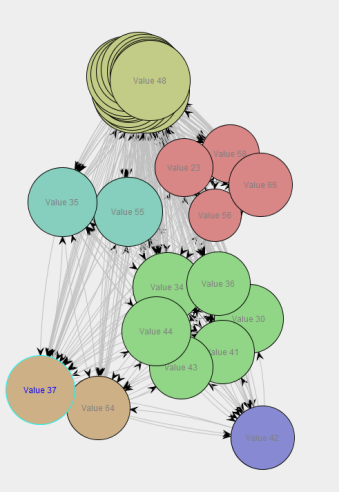

Figure 5.2: Cluster step 2

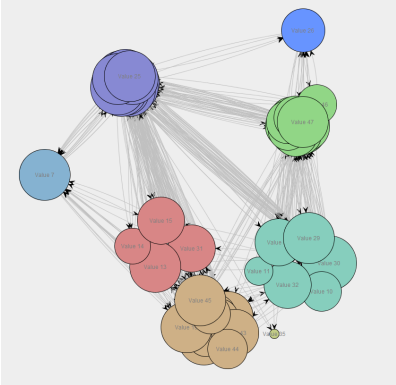

Figure 5.3: Cluster step 3

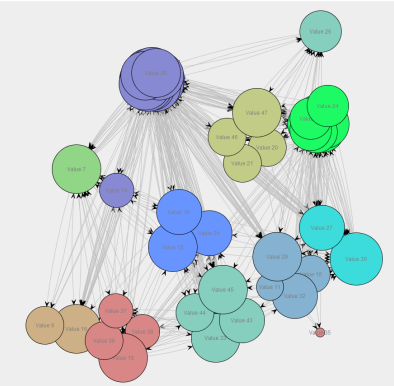

Figure 5.4: Cluster step 4

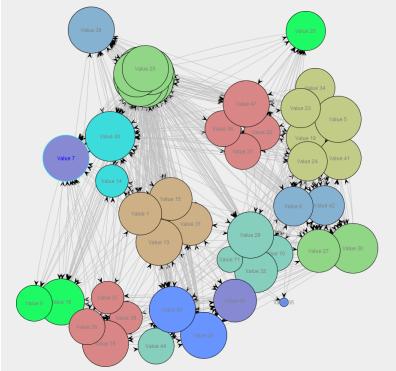

Figure 5.5: Cluster step 5

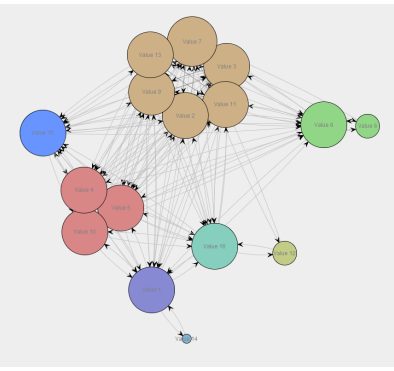

Figure 5.6: Cluster step 6

Table 5.6: Escalation statistics

Action	A	B	C	D	E	F	G	H	I	J	K	L	M
BOGr2	56	583	817	321	1197	633	1186	834	484	740	646	668	14
BOGr7	1496	577	661	399	1535	792	1244	562	344	416	815	783	890
BOGr2 esc	2	82	134	24	183	82	152	110	97	126	111	89	1
BOGr7 esc	118	30	51	14	75	86	87	28	44	14	71	65	31

Action	N	O	P	Total
BOGr2	430	2031	2	10642
BOGr7	459	1834	351	13158
BOGr2 esc	89	341	0	1623
BOGr7 esc	53	157	50	974

Table 5.7: Escalation percentage

Action	A	B	C	D	E	F	G	H	I
BOGr2	3.60%	14.10%	16.40%	7.50%	15.30%	13.00%	12.80%	13.20%	20.00%
BOGr7	7.90%	5.20%	7.70%	3.50%	4.90%	10.90%	7.00%	5.00%	12.80%

Action	J	K	L	M	N	O	P	Avg
BOGr2	17.00%	17.20%	13.30%	7.10%	20.70%	16.80%	0.00%	15.30%
BOGr7	3.40%	8.70%	8.30%	3.50%	11.50%	8.60%	14.20%	7.40%

Section 5.1.1 and 5.1.2 are used to first find how often an escalation happens and next to see look more in-depth in outlying teams to see what these escalated activities are exactly. The previous analysis is extended by filtering the dataset for a specific team each time the analysis is performed.

Table 5.6 shows the amount of routine appointments for both 2nd and 7th grade of primary school as well as the amount of cases that resulted in a follow up appointment i.e. escalation. Table 5.7 shows the relative amount of cases that were escalated by each team, from this we see that team P has almost twice as many escalations for BOGr7 as the average. Team I and N have a higher than average amount of escalations for both BOGr2 and BOGr7. It was of interest to know what activity was frequently followed for these teams, the results of this is shown in Table 5.8 for team P, Tables 5.9 and 5.10 for team I and, Tables 5.11 and 5.12 for team N.

Table 5.8: Escalation outliers P BOGr2

Escalation	Frequency
OOI 15 Groei	24
OOI 15 Houding	13

Table 5.9: Escalation outliers I BOGr2

Escalation	Frequency
OOI 15 Gehoor	44
OOI 45 Gesprek	11

Table 5.10: Escalation outliers I BOGr7

Escalation	Frequency
OOI 15 Groei	18
OOI 15 Houding	17

Table 5.11: Escalation outliers N BOGr2

Escalation	Frequency
OOI 15 Groei	29
OOI 30 Gesprek	18

Table 5.12: Escalation outliers N BOGr7

Escalation	Frequency
OOI 15 Groei	25
OOI 30 Gesprek	15

Feedback

This report gives multiple interesting insights, first of all it shows that BOGr2 results in many more escalations than BOGr7, this confirms some expectations but real hard numbers on this were not known before hand. Now that this information is known it is interesting to investigate why BOGr2 has more escalations, this could be because experts make great efforts to be on the safe side with young children, but the root cause is, at this moment, not known. Next it is interesting to note that there are still differences between teams and how they handle escalation. This gives cause for further investigation with the goal of equalizing the quality of service being offered. This is however not trivial since the work is knowledge intensive and it is important that a good balance is struck between allowing the professional to make his/her own decisions while also maintaining a certain level of standardization.

5.3 Performance Analysis

We now have a feeling for what the process and its resources look like, in this section we take a more in depth look at a few performance measures.

5.3.1 Are the recommendations from the Triage followed?

For this analysis the "Triage + Uitvoering" data-set is used which contains the needed data. The calculations could easily be performed by using Excel in the following way, the three different roles are all assigned a value; 0 for "Assisstent", 1 for "Verpleegkundige", 3 for "Arts". This is done for both the Triage result and the actual execution. Next Triage and execution for the same case are matched and the difference between the Triage value and execution value is calculated giving one results as shown in Table 5.13. A negative value indicates that the Triage considered that a higher educated professional would be suitable for the consult than the one that actually performed the action. Conversely a positive value indicates that the opposite is true. From a medical standpoint negative values are something that should be prevented. While from a financial point of view the positive values might be equally bad.

Table 5.13: Triage levels

Level	Label
-3	Assistent instead of Arts
-2	Verpleegkundige instead of Arts
-1	Assistent instead of Verpleegkundige
0	Correct
1	Verpleegkundige instead of Assistent
2	Arts instead of Verpleegkundige
3	Arts instead of Assistent

In Table 5.14 it is shown that the majority of routine appointments follow the indication made by the Triage. In 2012-2013 a large amount of appointments with a value of 3 and -3 are seen

Table 5.14: Triage results compared to execution

Level	Label	2012-2013	2013-2014	Grand Total
-3	Assistent instead of Arts	5.64%	0.01%	2.82%
-2	Verpleegkundige instead of Arts	0.53%	4.83%	2.68%
-1	Assistent instead of Verpleegkundige	4.85%	4.77%	4.81%
0	Correct	76.31%	76.47%	76.39%
1	Verpleegkundige instead of Assistent	3.21%	7.47%	5.34%
2	Arts instead of Verpleegkundige	0.84%	6.44%	3.65%
3	Arts instead of Assistent	8.62%	0.00%	4.30%

Table 5.15: FTE per team

Eindhoven	Noord	Oost	Zuid	Grand Total
17.86	16.08	16.93	8.8	59.67

while this has disappeared in in 2013-2014, where the majority of mismatches is from the other four levels.

Feedback

The results of this are acceptable, the large majority of appointments follow triage. The process of triage is however very labour intensive and there is an ongoing discussion whether or not the current benefits of triage outweigh the costs, this report can be used as input for the discussion. The big mismatch seen in 2012-2013 on the levels 3 and -3 results from a change in policy which was not yet correctly processed in the data resulting in these values.

5.3.2 How does productivity differ between teams?

For this metric a comparison is made between the 4 main teams that exist within JGZ are rated on the amount of appointments they have per week per FTE. This analysis was done using Excel by extracting the week number for each appointment date and tallying these up for each team. This total amount was divided by the amount of FTE for each team as shown in Table 5.15. The resulting cases per week per FTE are shown in Table 5.16.

Table 5.16: Cases per timeframe per FTE

Week Nr.	Eindhoven	Noord	Oost	Zuid	Average
1	8.119	22.015	11.692	25.227	15.569
2	28.275	99.751	74.005	128.295	79.537
3	31.131	99.689	119.838	118.523	90.716
4	34.602	79.229	91.231	110.114	79.336
5	32.531	131.841	105.721	142.955	99.162
6	47.704	118.097	113.184	120.341	97.788
7	25.700	77.674	93.408	78.409	68.795
8	31.747	103.047	104.042	134.545	90.263
9	40.314	102.736	114.925	117.727	95.408
10	34.938	83.333	63.868	75.455	65.443
11	40.426	115.112	114.179	140.114	97.486

Continued on next page

Table 5.16 – continued from previous page

Week Nr.	Eindhoven	Noord	Oost	Zuid	Average
12	34.658	109.888	114.925	116.591	93.682
13	34.938	109.080	131.343	131.364	98.190
14	28.555	92.289	101.057	111.477	81.180
15	31.019	121.393	109.204	118.409	94.771
16	26.708	98.881	83.520	127.159	79.504
17	29.675	89.863	90.299	116.023	79.068
18	36.226	16.791	16.853	42.273	29.898
19	26.540	60.261	40.609	31.591	52.154
20	30.347	103.856	114.801	98.068	89.459
21	31.355	101.368	90.112	94.886	80.225
22	21.781	96.206	85.199	115.227	76.923
23	38.522	124.565	120.647	123.864	105.128
24	44.569	116.915	98.570	121.364	97.185
25	33.707	128.607	135.697	125.341	109.837
26	37.010	130.970	154.602	183.523	123.948
27	51.960	106.281	130.784	131.932	110.960
28	50.840	82.836	97.886	89.318	86.140
29	41.153	24.689	47.512	67.614	45.768
30	11.422	10.261	4.291	8.523	10.927
31	4.199	24.067	0.560	10.227	23.831
32	5.207	30.846	19.279	30.909	39.132
33	19.373	61.194	54.478	84.091	66.801
34	29.675	95.647	98.445	136.250	85.386
35	47.368	141.231	141.356	138.977	116.323
36	43.505	168.719	154.851	150.455	127.015
37	41.545	132.774	157.711	177.273	119.323
38	46.473	143.284	164.490	144.545	121.200
39	44.177	155.846	147.139	182.500	125.256
40	42.217	150.187	161.629	161.818	123.546
41	35.554	85.075	123.570	101.364	83.358
42	23.404	67.102	62.562	116.818	60.600
43	34.770	68.905	67.910	98.068	63.315
44	28.219	78.607	79.726	92.727	67.052
45	32.475	81.405	96.704	138.295	79.873
46	35.498	90.299	104.726	93.750	78.917
47	27.940	113.868	106.965	117.841	88.302
48	31.355	108.396	95.585	107.500	86.073
49	30.739	112.002	96.455	107.841	86.878
50	36.618	101.555	106.468	115.000	89.023
51	36.730	99.316	109.453	128.523	90.783
52	27.436	53.047	50.871	46.932	44.545
53	1.120	3.918	1.244	1.591	2.162
54	0.112	0.000	0.000	0.000	0.034
Grand Total	1702.184	4924.813	4976.182	5629.545	4293.179

Feedback

This process went through two phases, we started by creating this same overview according to the teams that the dataset designated. This however led to some problems since the different

source data sets had different teams. Some sets spoke of team "Helmond" while this team did not formally exist. This problem was spotted during the first feedback round. After this was corrected the teams showed a much more comparable productivity with the exception of team Eindhoven.

When looking at these numbers it is important to note that they only correspond to the work done with the school children limited to dossier access and entry and not the other duties that JGZ has. To get a complete picture more information is needed. The report does however provide an instrument for achieving a homogenization between teams. When looking at Table 5.16 there are some obvious high and low points in productivity. One of these, at the start of the school year was most unexpected and, has led to better instructions and adjustments to the working process to ensure a better handover of work and planning for the next school year. Since the amount of available work is high at this time and many planning opportunities are still available it is expected that this will increase productivity at this stage of the year.

5.3.3 What are the working hours of employees?

This question is easily answered by using a pivot table in Excel and counting the amount of times a specific hour is seen in the timestamps.

The results as shown in Table 5.17 are clustered according to the hour in which the work was done, i.e. 0 contains all work between 0:00 and 0:59.

As shown in Table 5.17 productivity is highest before noon, after which it starts steadily dropping after 18:00 most employees appear to have gone home. Between 20:00-21:59 there is a small spike in the amount of dossier openings which could indicate that people are preparing for the next day. It is notable that the amount of dossier openings does not get back to the same level as it was in the morning after lunch, only a steadily decreasing line is seen. This does however assume that productivity is only linked to the opening of patient files. This does not consider that in the afternoon other activities may take place that do not involve the dossier which might skew the results.

Feedback

From the numbers seen in Table 5.17 most work in the dossier is being done during the morning. This Table however lacks information about other types of work that the employee is doing or it could also be possible that mornings are used to catch up on data entry and consultations with clients do take place in the afternoon. The results gave cause for the managers to look at the planning of professionals from which it could be seen that indeed the majority of appointments takes place during the morning. This is however undesirable since many clients prefer to have their appointments at the end of the day. One of the actions that was taken to better facilitate this was moving team-meetings, which were often planned in the afternoon, to other times of the day.

5.3.4 What is the turnaround time of the process?

To answer this question we first need to define the turnaround time of the process since in this case it is not straightforward. Normally we would look at the time between the first and last event in the process, but since the first event in the process is the making of the schedule this does not work. Furthermore the process is recurring every few year with additional check-ups. To give this question some meaning, we define the process starts at the time of routine check-up for a given year, and ends after all follow-up appointments are completed.

With the current level of data this analysis is not possible to do, since it is not clear in the data if an appointment has a follow-up, we can never be sure that the process for that year has ended. Furthermore it is very common that the dossier is opened at a later date without an actual appointment being present. This noise in the data combined with the fact that it is uncertain whether or not the process has ended makes it so that almost all cases take similar time.

Table 5.17: Dossier accesses per team per hour, normalised by FTE

Hour	Eindhoven	Noord	Oost	Zuid
0	8.286	10.883	17.424	33.522
1	1.679	1.803	2.244	3.522
2	0.503	0	0.177	0
5	0	0.062	0.354	0
6	0	0	3.662	0
7	2.015	53.669	20.318	2.840
8	125.531	653.669	499.822	581.136
9	373.180	1277.674	1133.845	1263.864
10	459.574	1517.164	1353.396	1483.750
11	504.871	1539.801	1359.008	1620.682
12	449.216	1379.789	1087.596	1425.227
13	438.017	1241.791	1086.887	1313.068
14	416.349	1073.197	1029.829	1281.818
15	368.645	922.077	924.276	1005.795
16	326.259	725.684	875.310	808.977
17	118.645	266.355	349.320	384.431
18	17.357	92.599	139.397	81.136
19	25.083	95.584	124.158	51.704
20	35.106	146.766	137.625	76.818
21	51.735	170.024	144.595	80.227
22	61.814	73.880	126.521	64.204
23	29.059	35.323	60.189	42.500
Grand Total	3812.934	11 277.799	10 475.960	11 576.250

This question cannot be answered with the current level of data, and is also not well defined for this case since the process does not really end.

Feedback

The logging of data needs to be improved in order for this answer to have a satisfactory answer.

5.3.5 How long does it take for a cancelled appointment to be followed by a successful appointment?

The analysis is done using Disco, the results of this question are especially interesting for the BOGr2 case since a planning for this activity is made automatically without input by the parents who have to attend the meeting. To get the desired information a combination is made between activity and activity status, this displays the connections between different statuses.

In total there are 864 cases of a person not showing up for the consult, 436 of these were eventually followed by a completed consult and the average time between these two is 69.8 days while the median is 49.9 days. In 123 of the 436 cases where a client did not show up for his or her consult, the client did not show up multiple times. 47 of these 123 instances in the end still led to a completed consult, the time between the median duration between the first time the client does not show up and the completed consult is 83.9 days.

Feedback

The reason 428 clients never have a successful consult after missing their first consult is most likely because the consult is not mandatory, although a consult is always planned the parents can always decide that no consult is necessary. On a whole of the ten thousand consults only 4% of clients do not show up. The time between a meeting where the client did not show up and a successful meeting is, ignoring outliers, 49.9 days which was expected by the managers since a planning is made for 7 weeks in advance. The question is however whether this is desirable and good enough for a customer centric organisation.

5.4 Conclusion

Using the dataset compiled in Section 4.2 it was possible to answer a number of questions that were of interest to management. The results of these led to a better understanding of the underlying process and a validation of the gut feeling for these managers. So far the analysis has only scratched the surface of the underlying process focussing on statistics that describe the current situation with its metrics. It is important to note the quality of the data, as assessed in Chapter 3, was not of a high enough level to make absolute claims on the current findings. The dataset in its present form contains unreliable events, as a result only general trends should be taken from the analysis results.

These results however are better than information that was available before the project started. And they are of relevance to the managers, it is important to learn form the short comings of the data and to take this into account for future improvements. Furthermore it is important that the users of the statistics generate in this chapter are aware of the short comings of the data when applying its results.

The generated reports and information were an eye-opener to the managers who had little experience working in this way and therefore did not know what to expect. The reports have given cause to further investigation and in some cases have led to changes being made to the process or extra instructions for the employees. The information was useful and there is a desire to further utilize the available tools and techniques to streamline the process. At the same time the whole way that service is being provided by JGZ is being discussed at this moment. Which leads to the question of how useful it is to optimize the current process while it might be completely revamped in the near future. Moreover the organisations BI structure is also in flux at the moment and a

vision is still being developed, it was however very good to see how much further the organisation has come since introducing the electronic patient files only 7 years ago.

While the generated process-models were difficult to read for the managers who have no experience with these kinds of models, it did open their eyes to some of the problems with it as well. As mentioned the process is not clear indicating a lot of variation in practice which has given them cause to rethink the way they work at the moment and increase their efforts into homogenizing the way their service is delivered.

Chapter 6

Recommendations

When looking at the BI infrastructure of GGDBZO it is important to also take the umbrella organisation GGD GHOR Nederland into account. As mentioned in Chapter 1 GGD GHOR NL takes care of projects that require a more centralised approach. At the moment it is working on a nationwide data warehouse with the goal of facilitating information exchange between different GGD locations and as a platform which enables the disclosure of data to outside parties. The vision is that a platform for open data is created with which market parties and citizens can interface to either enrich the data, e.g. with locally obtained measurements on air quality, or to use the available data and capitalize on this. This platform of course works with varying levels of trust and access, for example a physician performing research could be able to get more low level information than a company which would only get aggregated highlights. The data this platform would provide could also be a great source of knowledge for GGDBZO and JGZ, e.g. by providing detailed statistics on neighbourhoods and spotting trends that must be acted upon.

While GGD GHOR NL is working on this nationwide data store, the choice for analytics and BI front-ends is left to the independent division. This has the advantage that a fitting application can be chosen for each discipline, it is however very important that guidance is provided to make sure all 25 GGD locations and their different departments embrace this new capability. This does however mean that interfacing with this national system should be taken into account when designing local systems at GGDBZO.

BI in the medical domain has some extra challenges that most organisations do not have. There are many parties involved in the process such as the government, patients and doctors. Since the GGD also has a guiding function with regards to referring patients to the correct professional it is important that close collaboration is achieved with these parties. JGZ is also dependent on for example the municipality to supply it with data on who its clients are. Lastly there is an increasing trend of patients demanding a greater degree of ownership over their medical data. It is expected that a nationwide solution like the EPD will be fully implemented in the future, it is therefore important to design the system in such a way that information can be disclosed in a safe manner.

6.1 Organisational Challenges

One of the main challenges for organisations starting with their BI endeavours is setting realistic expectations and making sure that employees see the value of a centralised BI approach.[8] This misconception often originates from the different perspectives that IT and Business users have on what a BI platform should entail.[14] IT more often than not wants a monolithic approach with all data stored in one place regardless of use and is hesitant to allow a large group of people access to this one system containing all the information. At the same time the business is eager for more access and better information, but lacks the training to use the often very specialistic tools that IT uses. Change is necessary to bring together the Business and IT users, to make them see each

others point of view. It is important these two sides are brought together and strategic planning is made that is realistic for both IT and the Business users. One of the common problems in all major change projects which is also applicable to BI is high expectations and reality failing to live up to this.[3] One way of bringing people together and managing BI expectations, is by creating a specific business unit to do just that. This unit is commonly referred to as the Business Intelligence Competency Centre, which has the goal of connecting solution delivery to decisions through analytics.[9] A very important aspect for the success of this process is the willingness of the people participating in it, it is therefore important to find business sponsors that are willing to put time and effort into this process. Usually there are already some people working on ad-hoc Business Analytics through spreadsheet usage, we refer to these people as Spreadsheet Jockeys. It is important to recognise these people as they have already seen the advantages that BI can offer for the decision making process. They have an idea of the possibilities it can offer and what information is relevant to them. The downside is that their work is done in a very much ad-hoc manner, reports are manually generated as the need for them arises and more often than not work is done multiple times by different people that are unaware of each other. Bringing them together and exchanging information and best practices is a first step towards identifying what information is relevant to the organisation. At this point it is important to communicate the success stories to other business stakeholders. A feeling of relevance for BI must be ignited across the organisation. In this process we are able to learn from other organisations that have already gone through this process.

Many organisations had a very monolithic approach to BI, by making the BI team sole responsible for the analytics. The advantage of this approach is everyone has access to the same information and data-consistency is easier to maintain. The downside of this approach is that it does not do justice to the diverse demands that users have. Moreover this approach forgets the entrepreneurs that started doing small scale BI using spreadsheets. Often these Spreadsheet Jockeys are not part of the BI team and are taken out of the loop, while formerly they were the biggest champions for the organisation's BI approach and their engagement is of great importance for the acceptance of BI throughout the organisation. It is therefore important that an approach is chosen that both centralizes part of the BI infrastructure without forgetting its users. Since the past has proved that simply making information available through IT systems does not mean that it will be utilized by the end users.[13] While not all business users have the time or expertise to write custom reports that extract the data they need from the BI system it is important that those that are able to, have the ability to do so.[12] Users should however be aware of the shortcomings of these ad-hoc reports and a balance should be struck between the usage of ad-hoc reports and further verification of these reports to allow informed decision making. We have identified multiple types of information and tool needs. We identify the following roles: Information Consumers, the Spreadsheet Jockeys, BI experts and, IT experts. The information consumers are typically managers that use created reports to support their decision making and find areas of interest. The Spreadsheet Jockeys can be managers or local data experts, they are situated close to the business and have taken it upon themselves to create reports to suit their needs, they usually have some BI training but it is not their main task. The reports they create are good to gain insights but some errors should be expected. The BI experts almost solely concern themselves with BI task, they drill the data to find interesting facts and know the pitfalls of BI reports. Reports they create are considered trustworthy. The IT experts usually don't make use of the data themselves, instead they facilitate the usage of BI systems and tools. They are responsible for keeping the databases consistent and the tools running.

6.2 Technological Challenges

In Section 6.1 we have identified four different types of users for the BI system. The old approach of having one system to service all these different users is unsuitable for these diverse needs. A solution to this problem is proposed by [14], here a tiered BI solution is proposed, with each tier

having different access rights and tooling to better target the needs from a diverse set of users.

It is tempting to start by setting up a large scale project that identifies all needs and goes to work building an encompassing data suite. The problem with this approach is the fact that projects such as this are very time consuming and by the time they are done the needs of business consumers may have changed such that the resulting system is no longer useful.[3] It is therefore important to build a system and define a vision that incorporates this concept drift that is inherent to BI, the business will change and this means that the questions that need to be answered by BI also change. A problem with many guidelines for BI maturity is that it takes a long time for the organisation to reap the benefits of BI. The TDWI BI maturity model for example takes until the final stage before the monolithic approach is abandoned to allow for more close to business application that are blended with the daily workings of the staff.[4] This seems counter productive to the very important cultural changes that the organisation has to go through. The longer BI remains a separate activity removed from the daily work of business stakeholder, for whom the systems are build, the less inclined they will be to support and use the system. This does not mean that the model is no longer useful, with regard to the challenges to building an organisation wide data warehouse it is very useful. We however propose to make a distinction between the process of integrating data and establishing an organisation wide set of reliable up to date information and development of the analytics and reporting tools. It is important to realise at every step that the building of a BI system is not a goal of itself but rather a means of supplying accurate management information. For the business stakeholders to be on board for the project new systems should not limit their capabilities but rather empower them. This adds extra requirements for the systems that are to be designed, a new system should never be a downgrade for the end business users. Applications that the business uses should therefore be able to cope with a changing background data warehouse. As data quality improves and definitions get streamlined the end user applications should be able to handle this change.

6.3 Event log improvements

As described in Section 3.2 the maturity of the combined event log used for the analysis in Chapter 3 is of level ⋆⋆. To improve the quality of the event log, it is important that we identify what type of events are relevant to the process. Events should be recorded in a reliable manner and a clear relation between events should be present. Furthermore instances of "gaming the system" should be removed, at the moment it is for example very easy to set such generous deadlines for appointments that in practice they will always be met. This gives a skewed view of reality and in the end does not help the organisation advance and improve the service it provides to its customers. Such data should either be enforced to have a close relation to reality if its purpose is to provide management information. Or the organisation has to let go of this formal definition, allowing the experts to all use it in their individual way, and accept that this bit of information can not be used for analysis. This is just an example but it is an overarching problem with user provided information. The organisation needs to either enforce a common definition and usage, or accept that the data is for private use by the expert only. As mentioned before the recording of events and data related to these events needs to become systematic and well defined. The applications used to record the data need to accommodate this by providing clear information on what is and is not logged. At the moment the systems used by JGZ already record to which patients performed actions belong, meaning there is some definition of a case already. What the systems are lacking at the moment is to what task events belong. An educated guess is made by relating logged events and planned appointment, working on the assumption that work is done to complete appointments. This approach is however lacking, and leads to untrustworthy event logs. It is therefore important that the access log does not only detail when a patient's records were accessed but also why. This is however non trivial since if the general dossier is always opened it is difficult to determine to what purpose it was opened. This is however not impossible, example solutions are to divide the dossier in sections, so it is at least know what section the user was interested in. And to allow users to add to an appointment, this would still enter the information

in the general dossier, but with an added connection to the corresponding appointment. This way users cannot only see the general dossier but also see what information relates to which appointment. The key objective is to improve the quality of the logged data without annoying the user.

6.4 Application

For this project we have used BI and event log maturity models to assess the current state of the BI approach at the organisation. We have run ad-hoc pilot projects to gain a better insight into the needs of the business. This was however a very isolated endeavour and further projects and collaborations with key players within the organisation are needed to fully bring the described approach into practice. The pilots detailed in Chapter 3 do however succeed in offering the involved managers a glimpse of what is possible should the organisation realise BI centric decision making. Using the maturity models we define what steps should be taken to increase the BI maturity level of the organisation. It is important to take into account that change can not be forced onto the users, this will only cause them to undermine the project, therefore the most important development that the organisation must go through is a change in culture. A few managers have now seen what opportunities BI can give, it is imperative that this initial project is followed up by a systematic approach involving all key employees.

Start by recognizing who the Spreadsheet Jockeys are within the organisation and bring them together to help identify their needs and objectives with regards to BI. Use this to streamline and homogenise some of their needs, make sure that when talking about a KPI they are talking about the same definition. Bring in a BI specialist to lead these sessions and make sure clear goals are defined. These results can be used as valuable input when determining where and how to invest money into the company's BI endeavour. Next it is important to offer training to this group focussing on practical BI applications, remember that BI is not their main job but a way of supporting their primary function. Their training should keep this in mind and focus on practical applications of BI while also educating them on the pitfalls and limitations of their ad-hoc reporting. Allow them to and facilitate them in putting their newly learned skills into practice and give them a platform where they can share the results with the rest of the organisation and more importantly the department for which they mainly do work. This will help information consumers see the new opportunities that the BI approach can offer and spur them to define new questions and requirements. At some point the Spreadsheet Jockeys will need further support from BI specialists to supply them with more advanced combined datasets and their expertise in making sure that a report is trustworthy and ready to be used for decision making instead of only a guideline. Up to this point the BI approach can be fairly low budget and effort, existing structures can be used to create more custom reports and the organisation will come to realise the opportunities that a more systematic BI approach can offer. This is an important step the organisation must make, if envisioned end users do not see a need for this change, they will not be committed to the project and its success. Use these pilot projects to create a vision for what BI should entail within the organisation. The pilot projects are also a quick way of identifying what does and does not work for the organisation, which metrics are important to the decision makers and what is needed for that. The outcome of this should give enough handholds to develop a preliminary BI strategy, identify the organisations weak points and strengths.

The BI experts have two primary functions in this approach. The first is to provide BI tools to the local team experts, formerly known as Spreadsheet Jockeys, such that they can use more high level information to satisfy their specific needs and allow them to easily create new types of reports. When a new report created by the domain expert seems promising the BI expert can check the techniques used and make sure that the results are reliable before making it available company wide. The second role the BI expert has is more data science centric, looking for new patterns in the data that could be promising for the business and making this data available to the domain expert such that they get more tools to satisfy their information need. It is up to the BI experts to ensure that a sense of direction is maintained and an overall BI strategy is enforced.

The Business users supply a bottoms up request for information that the organisation must satisfy. At the other end the BI expert delivers a top down suite of capabilities to satisfy this thirst for information. The domain experts are in between these two supplying the business users with understandable information by making use of tools offered by the BI expert. IT is responsible for maintaining these systems and making sure that the systems and applications align with business needs. In this approach the business defines the needs of the BI system while the BI experts supplies the means by which to satisfy those needs. The domain expert uses the tools provided by the BI expert to answer the business questions posed. Make sure the BI expert has the time and freedom to explore new BI techniques and experiment with them. Allow for room to inspire the business users with new types of information that can be relevant to the business end users. Accept that it is not the managers task to create new and innovative reports but that (s)he simply has a request for actionable information, whatever that information might be. It is up to the expert to supply different types of information and in this way, together with the business users, find out what information is relevant.

Chapter 7

Conclusions

In this thesis we have assessed that the BI maturity of GGDBZO is still in an early stage. Both from a technical and a cultural point of view. There are many disparate systems and their data is not easily related. The organisation is not used to working with BI tools and decision making is mostly based on experience and gut feeling. We consider GGDBZO to be in between the unaware and opportunistic level of the Gartner BI maturity model and at Spreadmart stage in the TDWI maturity model. From a data maturity perspective the multiple systems and manually in-putted data is a detriment to the overall data quality. We consider event log maturity to be level ⋆⋆ with a major problem being the lack of a clear relation between events and tasks. Using a labour intensive approach we were however able to construct a dataset that was usable for process mining.

During our research we looked at multiple avenues of approach for using the data with the goal of process mining in an organisation that is at the start of its journey in the world of Business Intelligence. We have assessed that while JGZ has a lot of technological challenges ahead of it, the currently available data is indeed usable to answer questions about the current goings on at JGZ. One of the big challenges that had to be tackled before this management information could be delivered was moulding the available data in a consistent dataset which was usable for process mining. However after this was done managers that we worked with did manage to use the information provided by the case study to start changes in the working process aided by the quantitative data. So while it is difficult to get a starting dataset there are still many easy victories to be achieved by an organisation that just starts with process mining.

Looking at future prospects and further integrating BI into the GGD organisation the most difficult hurdle that must now be overcome is a change in organisational culture. Without accomplishing this advancements on other levels have a high chance of failing. While at the moment there is most certainly a group that is enthusiastic about the possible benefits of BI and Process Mining this group is still a minority. Beyond this it is also important to improve the systems being used since the current ones are not ideal for extracting and coupling data. While new systems are being worked on and the promise they hold is big. For now the investments made towards better IT systems focus on the day to day production work that JGZ must complete, there is still less of a focus on employing this same approach to the generation of management information. Current low level spreadsheet solutions are still considered proficient by general management and while new initiatives are deemed interesting budgetary commitment is still lacking. If the current status quo is maintained it is still possible to undertake BI endeavours as proved in Chapter 5 but the work will be very labour intensive and the results not as reliable as a more structured approach. It is imperative that time is made available for the exploration of the available data in order for the organisation to grow in its BI maturity, having just enough manpower to keep up with business demand will not suffice. A large investment in Business Intelligence is necessary for the endeavour to succeed, just like GGD GHOR NL is making a large investment towards a centralised system, GGDBZO should invest in its future.

Overall it can be said that while BI and data maturity is still low, JGZ can at present gain

some quick insights into its process by applying process mining on the available data. The work is however very labour intensive and options will most likely be quickly exhausted. To continue gaining benefits from BI and process mining, the whole organisation needs to see the benefits that such an approach can offer and structural investments and leadership towards this goal are needed.

7.1 Future Work

During this research it was shown that while some process mining techniques can already be applied to the data, the level of detail is still limited with the current data quality. GGDBZO is however in the process of transitioning to a new CRM which should fix a lot of data consistency issues and a new patient dossier system which will allow for better access to the data. After this effort to improve the data it can be worthwhile to re-examine the possibilities with regards to process mining. During the writing of this thesis we also got a better insight into the inner workings of the GGD and its organisational structure. The interesting interactions between local GGD branches, the national association, different municipalities that GGD services and its client patients. A full assessment of the relationships between all these parties can be made to identify the data needs of all parties and determine the fitness of the current systems. An analysis can be made of the interconnectivity between the different departments within GGDBZO, the tools and best practices that are used and how governance is situated to manage all this. The expectation is that most departments are struggling with the same challenges.

Another noteworthy development is the nationwide system that is soon being launched by the national GGD association GGD GHOR NL. This is an interesting data integration problem where at first a stable data basis is being created. This data warehouse must however be kept consistent while receiving input from the twenty five different GGD departments. Keeping the data consistent is a challenge in and of itself since not all GGD departments use the same local systems. Furthermore the data in this data warehouse can be used as an input for data mining endeavours. An overview can be created of all the different tools being used by GGD locations throughout the Netherlands, in addition to this the maturity level can be evaluated for each location. This can be used as input to compare and homogenize the different locations and a baseline for talking about best practices.

Bibliography

[1] P. Brooks, O. El-Gayar, and S. Sarnika. A framework for developing a domain specific business intelligence maturity model: Application to healthcare. *International Journal of Information Management*, 2015.

[2] F. Daniel et al. The Process Mining Manifesto. In *BPM 2011 Workshops, Part I*, 2011.

[3] W. W. Eckerson. Gauge Your Data Warehouse Maturity. *TDWI Research*, 2004.

[4] W. W. Eckerson. Beyond the Basics: Accelerating BI Maturity. *TDWI Research*, 2007.

[5] C. Howson. ITScore for BI and Analytics. *Gartner Research*, 2015.

[6] R. L. Leitheiser. Data Quality in Health Care Data Warehouse Environments. In *Proceedings of the 34th Hawaii International Conference on System Sciences*, 2001.

[7] T. Mettler and V. Vimarlund. Understanding business intelligence in the context of healthcare. *SAGE Publications*, 2009.

[8] Chandler Neil. The Business Intelligence and Analytics Leader's First 100 Days. *Gartner Research*, 2015.

[9] J. Parenteau and E. Zaidi. Redefine, Reorganize, Revamp and Rebrand Your BICC to Shift Focus to Analytics. *Gartner Research*, 2014.

[10] D.J. Power. A brief History of Decision Support Systems. *DSSResource.COM*, 2007.

[11] I. H. Rajteric. Overview of Business Intelligence Maturity Models. *Journal of Contemporary Management Issues*, 2010.

[12] K. Schlegel. Deliver Business Intelligence With a 'Think Global Act Local' Organizational Model. *Gartner Research*, 2010.

[13] K. Schlegel. How to Deliver Self-Service Business Intelligence. *Gartner Research*, 2012.

[14] J. Tapadinhas. How to Architect the BI and Analytics Platform. *Gartner Research*, 2014.

[15] W. M. P. van der Aalst, H. A. Reijers, and M. Song. Discovering Social Networks from Event Logs. *Computer Supported Cooperative Work (CSCW)*, 2005.

[16] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, chapter Process Mining, pages 7–11. Springer Verlag, 2011.

[17] B. Wixom and H. Watson. The BI-Based Organization. *International Journal of Business Intelligence Research*, 2010.

Appendix A

Process Exploration

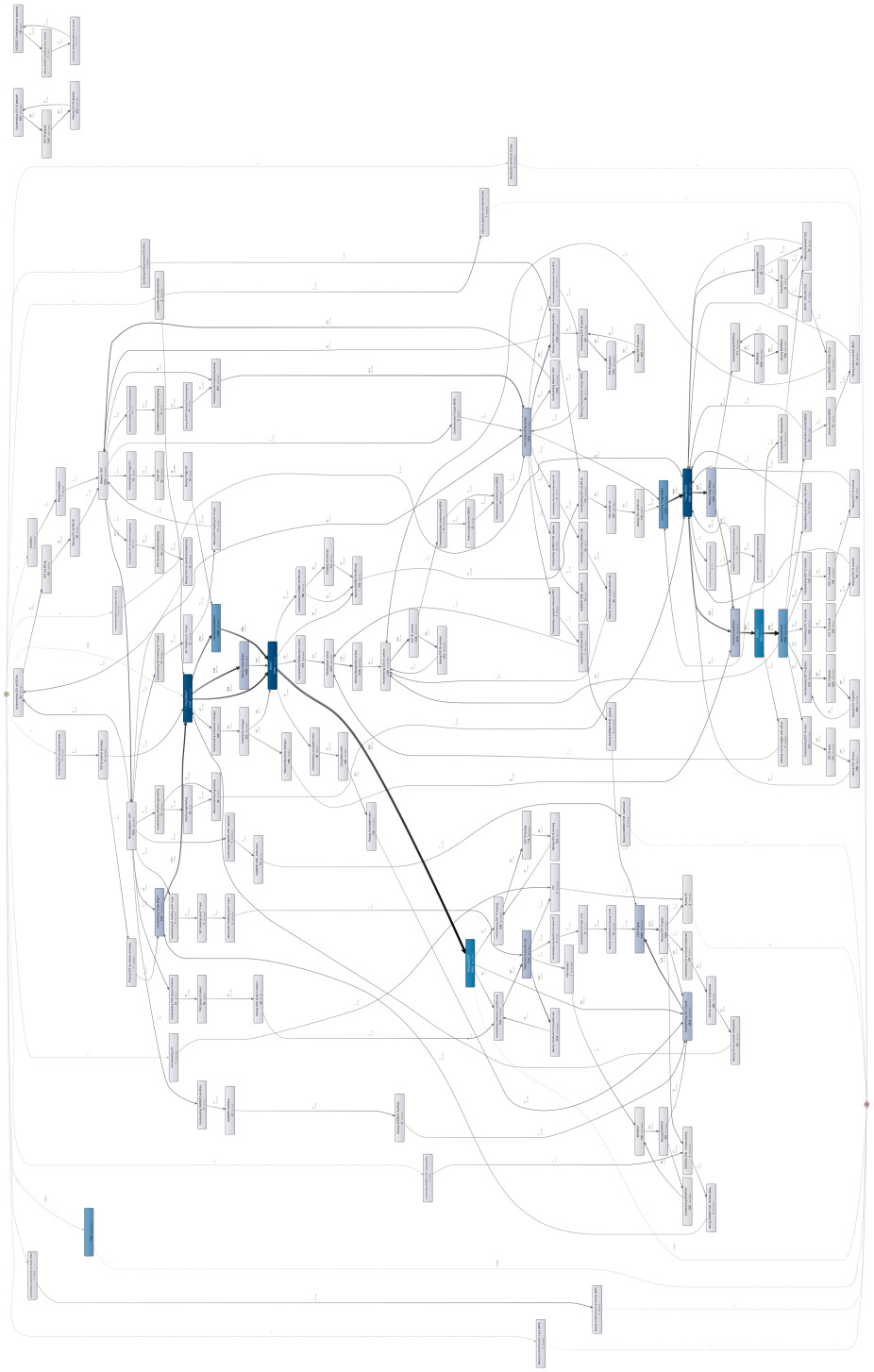

Figure A.1: Process map of the complete process

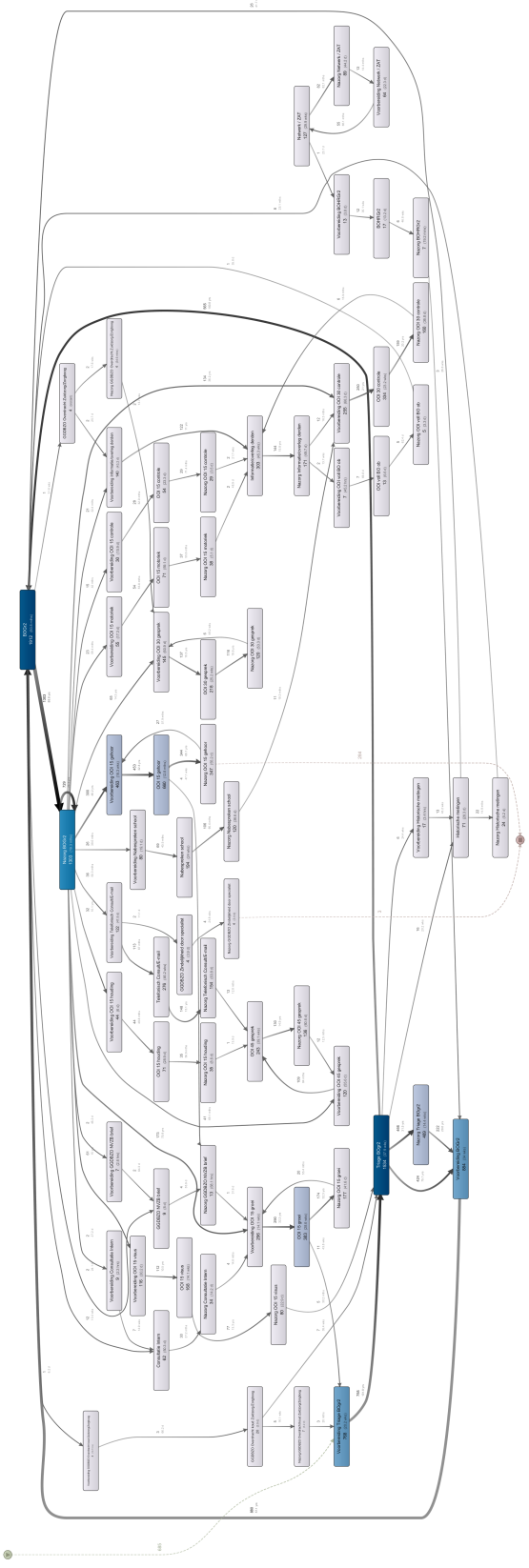

Figure A.2: Process map showing escalation of BOGr2

Appendix B

Social Network Analysis Clusters

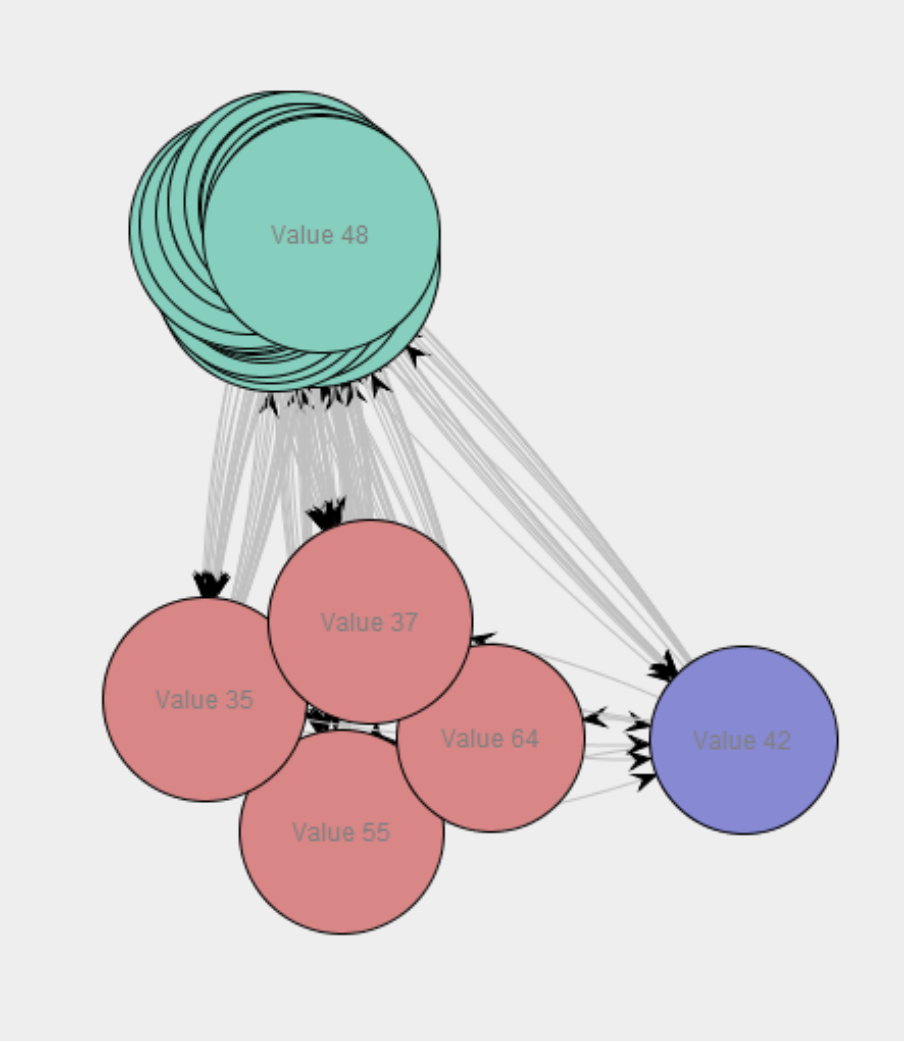

Figure B.1: Cluster step 1 large

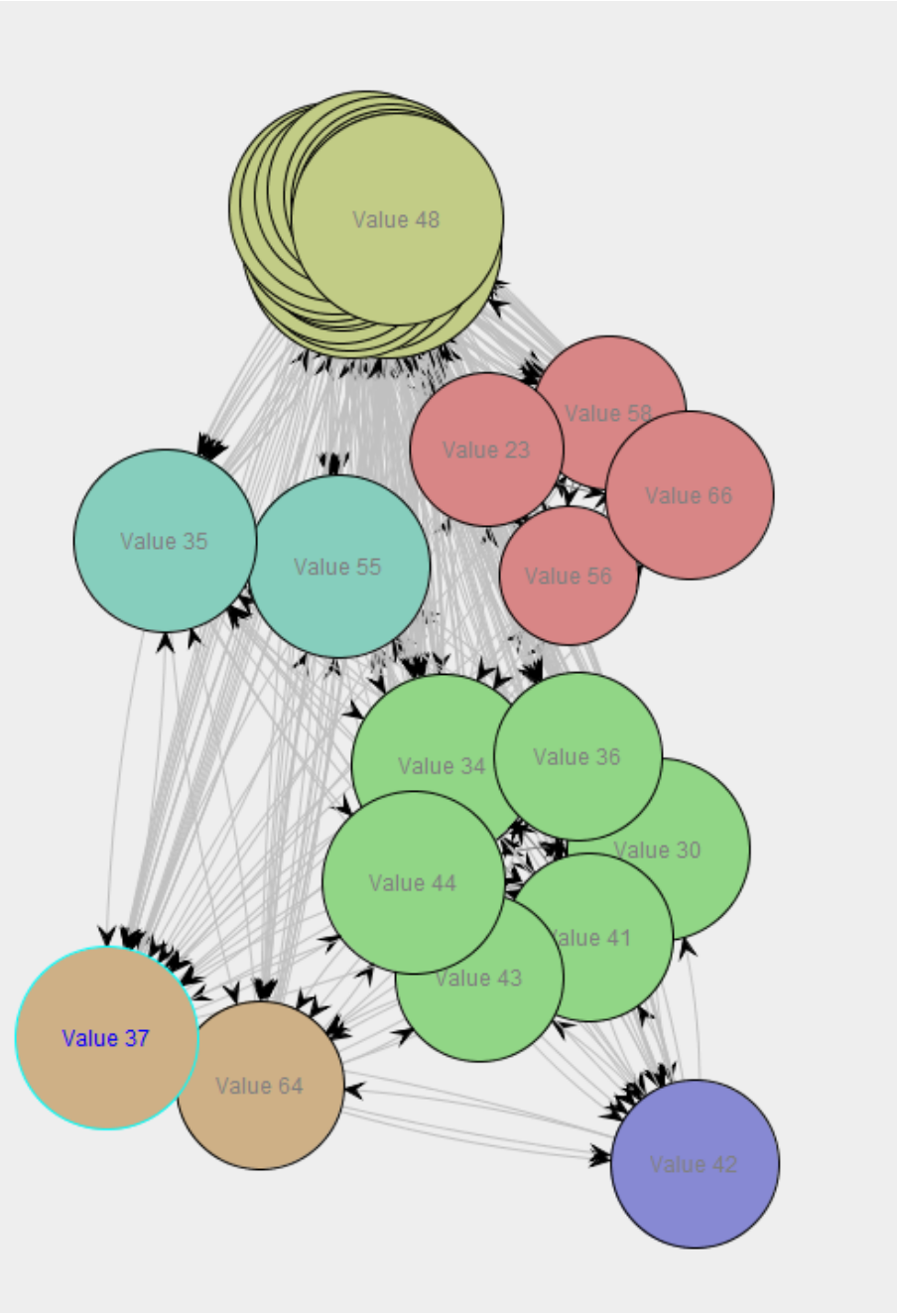

Figure B.2: Cluster step 2 large

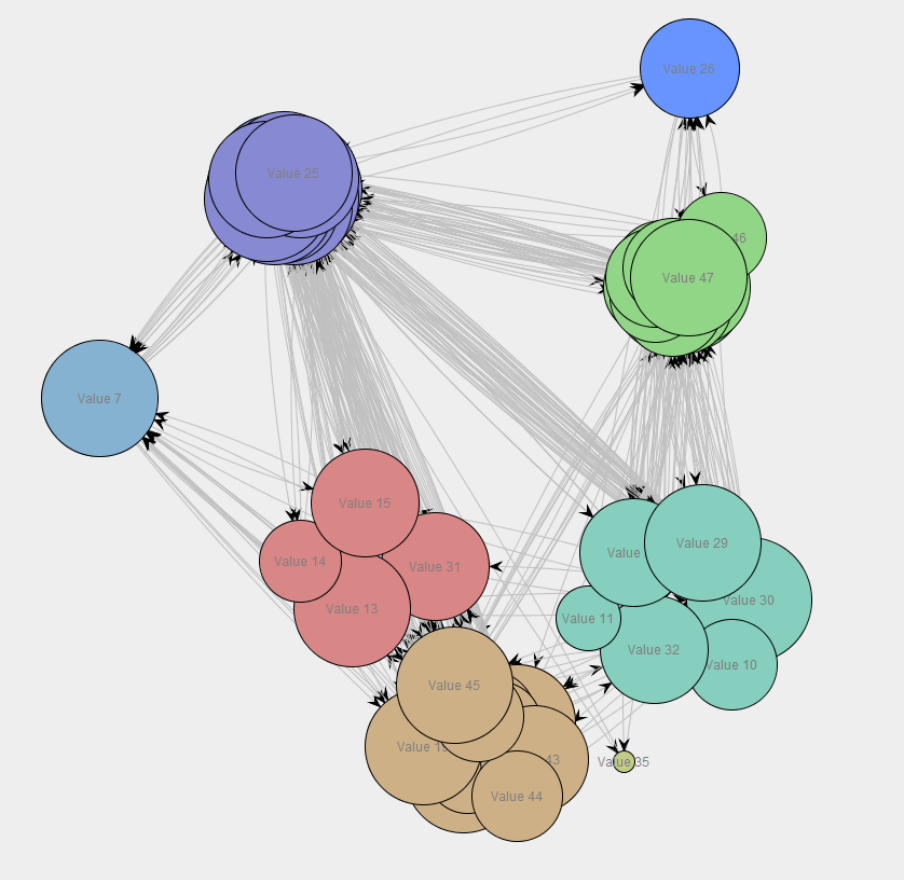

Figure B.3: Cluster step 3 large

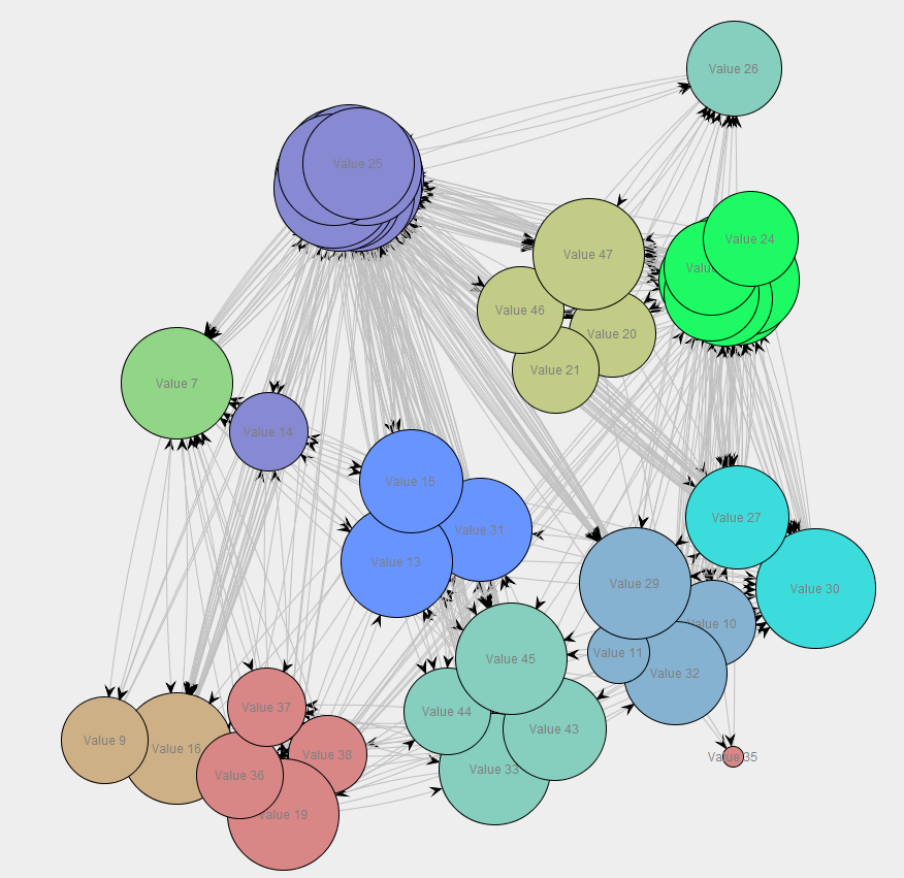

Figure B.4: Cluster step 4 large

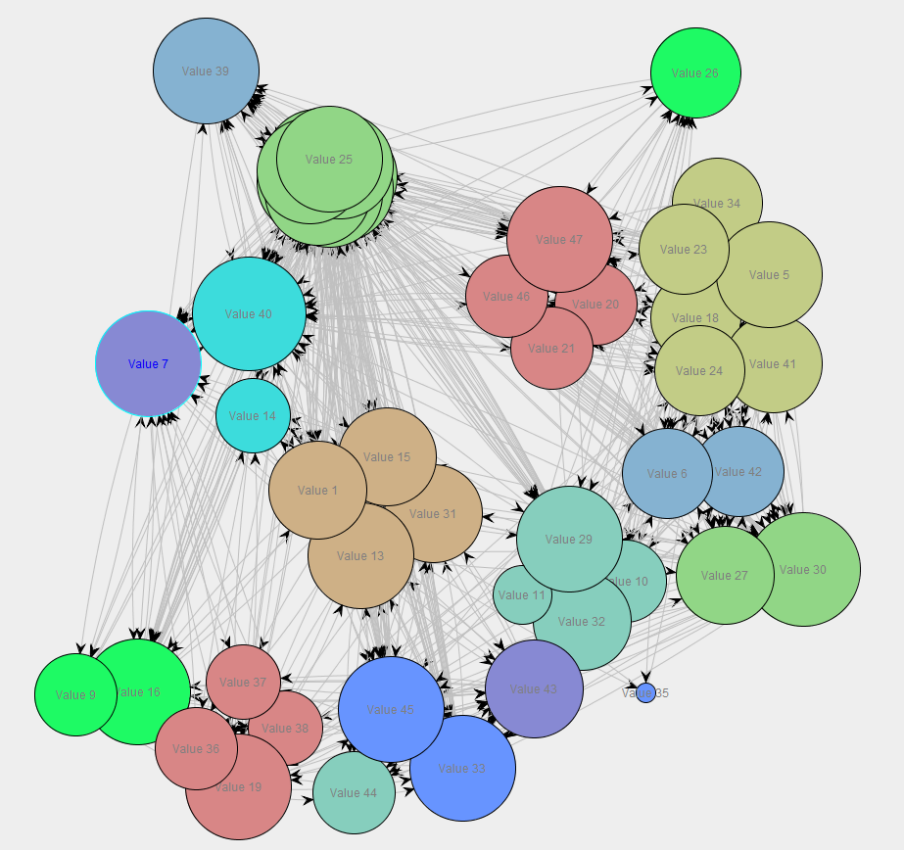

Figure B.5: Cluster step 5 large

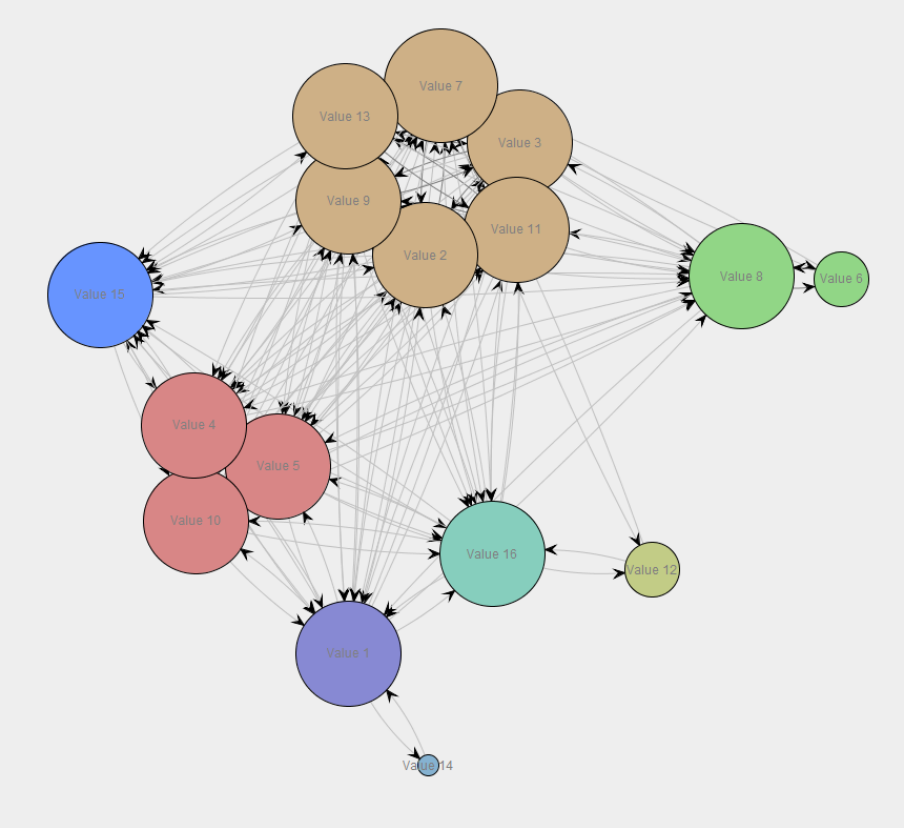

Figure B.6: Cluster step 6 large