Social Performance Review of Five Dutch Municipalities: Future Fit Cases for Outsourcing?

Josef K. J. Martens, Paul Verheul

Capgemini, BTECH Division, Reykjavikplein 1,
3543 KA Utrecht, The Netherlands
{jef.martens, paul.verheul}@capgemini.com
j.k.j.martens@alumnus.utwente.nl

Abstract. Dutch municipalities are under government pressure to spend less and are subject to municipality amalgamation. As previous research shows, amalgamation is applied to attempt to decrease spending, aligning with budgets, to no avail and no significant increase of service levels. The Capgemini Information Managers Survey results find that performing under budget pressure is topic #1 of participants' concern for public and private organizations. The BPI Challenge 2015 dataset underlines the complexities of permit applications which indicate case management behavior. Across five municipalities, process uniformity, employee behavior and leadership styles are compared. In a subset of permit applications, big differences in processing and assessment of permits are found. The results indicate the need for BPM in combination with CMMN to cover the complexities in case driven processes for further assessment for harmonization. Additionally, a classification of entries for the past BPI Challenges is provided to improve process mining reporting efforts.

Keywords: Process Mining, Business Analysis, Social Network Analysis, BPMN, CMMN, Municipality Amalgamation.

1 Introduction

In the context of the Business Process Intelligence Challenge 2015 the authors would like to take the opportunity to provide their insights in the dataset by providing answers to the questions posed. The dataset for 2014 was quite challenging and leads to participating with more drive in 2015 and a different research team setup. Our reasons to participate in the challenge are: the challenge provides a good exercise for applying process mining skills, it provides an opportunity to connect business & IT viewpoints and it provides a great opportunity to connect science & practice viewpoints.

The structure of the document is as follows: we describe the business need for the questions asked, the scientific need for the research as presented in this paper and the research scope. Chapter 2 is the literature review, chapter 3 discusses the research

method, chapter 4 provides the result overview of the research, chapter 5 are the conclusions regarding the research outcomes and chapter 6 are the recommendations.

1.1 Business Need

In the current market conditions where the authors are working, most projects require a strong business case, which needs to be validated by multiple stakeholders. Business responsible are challenged by the market in providing more differentiated products and services, where typically the budget driven IT departments are challenged with declining budgets. Therefore, pressure on cost reduction and leveraging economies of scale are required by both; the demand and supply side of system integration.

Improvement of operations leads to cost reductions, which underlines the value of process research. The classification of the area of research is found in both Operations Management (OM) and the Business Process Management (BPM) domains.

The OM and BPM domains have great solutions for the straightforward process definitions, which are known as Straight Through Processing and 'Happy Flows' which tend to be applicable in Six Sigma and Lean calculations. The establishment of the Business Process Model and Notation (BPMN) as an ISO standard in 2013 (ISO 19510:2013) [01] shows that the BPM is reaching a mature stage.

There is a current trend in BPM which tries to tackle the non-straight through processes and case based process chains with the establishment of the Case Management Model and Notation (CMMN) standard [02]. There is a need to get insights in the performance and formalization of the complex part of business processes.

Capgemini provides multiple solutions that fit to this trend: Value & Performance Management allows for rules based solutions which can be tailored to non-standard processes. Recognizing patterns and differences is part of the Insights & Data solutions, performed by the Business Intelligence and Business Analysis roles in conjunction. All these solutions are part of or liaise to the Operational Excellence proposition

1.2 Scientific Need

Coming from the practice of Business Analysis (BA), it is important to make a link between daily operations and the backbone of our research area, BPM. This paper is thought of a link between science and professional activities, to highlight for undergraduates how scientific developments can be applied in a practical way. Furthermore, the area of Process Mining in the BPM expertise is heavily reliant on business operational data to create new insights based on more and more complex actual questions coming from business owners. A company like Capgemini provides a framework for professional development and fortifies this in the curriculum for Business Analysis [03].

Contributions to the BPI challenge can be considered as explorations of the scientific domain as multiple professionals provide diverging methods and tools and

provide a converging path to where process mining can be applied. By choosing real-world datasets for the challenge, a true connection between theory and practical application is established.

1.3 Research Scope

With this paper the authors provide a deeper insight into profession of Business Analysis in a client engagement context using day-to-day operational skills. Furthermore, the setting is framed in a business context to validate business drivers for a research project like process assessment using process mining. The authors provide (limited) answers to questions which are similar to customer situations, in a time constrained matter. There is no attempt to create an addition to the process mining algorithms or software. By providing a limited literature review, an overview of professional and scientific use of process mining operational activities and tools is presented.

2 Literature Review

As stated in the introduction, Business and IT decision makers are under pressure which is represented in research as provided by Jerry Luftman for the past decade. An overview of the applicable content for the BPI Challenge 2015 is provided in the next paragraph.

To support the development of the Process Mining community and the Business Analysis community, this paper provides a comparison and classification of the submissions of the past 4 BPI challenges to assess the used methodologies (partially), the tools and the functionalities used in case of process mining software.

Alongside the presented methods for Process Mining analysis in the Massive Open Online Course at Coursera by Prof. Dr. Ir. W. van der Aalst [04] and book [05], an exploration is done on the structural approach on performing research in a best practice way.

2.1 Business Issues

The main issues ‘that keeps CIO’s awake at night’ have not changed much over the past few years. From this years’ research by Luftman [06], a new topic has risen, but overall, the past five years consists of the same top 3 topics as concluded by Luftman & Derksen [07]. Table 1 presents an overview of the most important topics as mentioned by CIO’s worldwide. From the top five topics, two are selected to be of direct relationship with the BPI Challenge 2015. In the following paragraphs the topics are discussed and related to the challenge case [08].

Table 1: Key management concerns for 2010-2014, highlighted in bold are the topics relevant to the BPI Challenge 2015.

Topic *Year*	*2014*	*2013*	*2012*	*2011*	*2010*
Alignment of IT & Business	1	1	2	1	3
Business Agility	2	2	3	2	2
IT Time to Market	3	-	-	-	-
Business Cost Reduction / Controls	4	3	1	1	1
Business Productivity	5	4	1	1	1

2.1.1 Business Cost Reduction / Controls

"Cost reduction is considered the foundation of long-term sustainable competitive advantage, especially during economic stagnation and therefore this will continue to be highly ranked in all geographies" according to Luftman [06, p5]. Given the explanation of the business need in paragraph 1.1, this is fully in line with our opening statements. In the BPI Challenge 2015 case description [08], outsourcing is mentioned regarding process steps in the process flow targeted. Outsourcing is a method often applied to achieve cost reduction. The topic of 'cost reduction / controls' is a split from the previous more generic statement of 'Business Productivity' from 2014 in the CIO survey [07].

2.1.2 Business Productivity

Productivity is "the rate at which goods are produced or work is completed" [09]. In the context of municipalities for the BPI Challenge, the work completed is the approval or rejection of a permit application [08]. Business productivity for municipalities is one of the topics of the current administration in the House of Representatives in The Netherlands, which resulted in an effort for reduction of government spending by amalgamation of municipalities [10]. In the article, extensive research by Allers & Geertsema [11] is mentioned which concludes that: (1) amalgamation does not affect local government spending in negative or positive ways, (2) economies of scale may exist, but are not measurable as lower spending, (3) service levels are not increased as measured by house prices and expenses in non-nationally regulated topics such as culture & recreation, (4) long term administration spending goes down. It may be that the critique is applicable towards Business School curricula as stated by Ghoshal's article from 2005 [12] have influenced the values and behavior of management executives in an negative way, therefore overacting and applying theories to everything based on learning experience rather than applying research to validate current experiences.
Capgemini continuously researches topics that are important to Information Managers (IM), a role which is becoming of increased importance and awareness in public and private organizations. The number one topic for IMs [13] is the amount and variety of work to handle with less budget, for two consecutive reports.

2.2 BPI Challenge revisited

In 2014, we the authors, started but did not finish research on the dataset provided by the BPI Challenge. The reason was that we virtually drowned in the data and could not retrieve results as requested in the limited timeframe as set for the challenge by ourselves. Therefore, this year's entry starts with revisiting all previous 28 entries by professionals for the BPI Challenge from 2011 through 2014 [14-42]. The authors aim to provide insights in what is successful research by comparing participants, methods, tools and filters. For this research paper, the outcomes are used to define how to approach BPI Challenge 2015.

2.2.1 Participants

In Table 2 an overview of all authors is provided and a classification whether the author(s) are a science or business representative.

Table 2: Classification of authors to be Researcher(s) and/or Professional(s) participating in the BPI Challenge

Author(s)	*Researcher(s)*	*Professional(s)*
Bose & van der Aalst	Y	
Caron et al.	Y	
Varvaressos	Y	
Adriansyah & Buijs	Y	
Bautista, Wangikar & Kumail Akbar		Y
Bose & van der Aalst	Y	
Kang et al.	Y	
Molka, Gilani & Zeng	Y	Y
Verbeek	Y	
Arias & Rojas	Y	
Bautista et al.		Y
vanden Broecke et al.	Y	
Dudok & van den Brand		Y
van Geffen & Niks		Y
Hansen		Y
Hevia & Saint-Pierre	Y	
Kang et al	Y	
Martens		Y
Paszkiewicz & Picard	Y	
Radhakrishnan & Anantha		Y
Van den Spiegel, Dieltjens & Blevi		Y
Hansen		Y
Thaler et al.	Y	
Buhler et al.		Y
Dees & van den End		Y
Suchy & Suchy		Y
Buffett, Emond & Goutte		Y
Van den Spiegel et al.		Y

2.2.2 Methods

In Table 3, an overview of the methods used in an operational way to perform Process Mining analysis is provided for the professional participant entries for the BPI Challenges from 2011 through 2014 [14-42]. In the following subparagraphs a short description is provided per method applied.

Table 3: Overview of the methods used in performing BPI Challenge entries from 2011 through 2014

Author(s) *Methods applied*	*Describe dataset*	*Perform Pre-Analysis*	*Combine data files*	*Preserve data uniqueness*	*Classify cases*	*Create subset by filtering*	*Create new (ProM) Filter*	*Use multiple tools*	*Case performance analysis*	*Process performance analysis*	*Compare Control flows*
Bose & van der Aalst	Y	Y				Y	Y			Y	
Caron et al.	Y	Y				Y					
Varvaressos	Y					Y		Y			
Adriansyah & Buijs	Y	Y				Y		Y			
Bautista, Wangikar & Kumail Akbar	Y	Y				Y		Y	Y		
Bose & van der Aalst		Y			Y	Y	Y			Y	
Kang et al.					Y	Y		Y		Y	Y
Molka, Gilani & Zeng	Y					Y	Y			Y	Y
Verbeek						Y	Y			Y	
Arias & Rojas					Y	Y		Y	Y	Y	Y
Bautista et al.	Y	Y		Y	Y	Y		Y			Y
vanden Broecke et al.	Y	Y				Y		Y			Y
Dudok & van den Brand					Y	Y			Y	Y	Y
van Geffen & Niks									Y	Y	
Hansen						Y				Y	Y
Hevia & Saint-Pierre						Y		Y			Y
Kang et al		Y				Y					Y
Martens	Y	Y				Y					Y
Paszkiewicz & Picard						Y		Y			
Radhakrishnan & Anantha	Y					Y					Y
Van den Spiegel, Dieltjens & Blevi		Y				Y				Y	
Hansen			Y			Y					
Thaler et al.		Y			Y	Y		Y			
Buhler et al.	Y		Y			Y					Y
Dees & van den End			Y		Y	Y				Y	
Suchy & Suchy	Y	Y	Y			Y		Y			
Buffett, Emond & Goutte		Y	Y		Y	Y		Y			
Van den Spiegel et al.	Y	Y	Y		Y	Y		?			

2.2.2.1Describe dataset

The dataset as provided is described in a qualitative and quantitative way, mostly describing the number of cases, the number of events and the context of the dataset.

2.2.2.2 Perform Pre-analysis
A review of the dataset is performed to assess the relationship between activities, events, cases and other attributes of the dataset. This step involves the usage of process mining or other tools to determine the scope of the further actions.

2.2.2.3 Combine Data files
In some challenges, multiple files have been provided to perform analysis on. During pre-analysis, relationships between these files have been confirmed and a combination between the files is created by use of a data model or ontology. Most authors that combine data files use a Data Base Management Suite (DBMS) which supports Structured Query Language (SQL).

2.2.2.4 Preserve data uniqueness
Data is in some cases combined, but poses a possible error by duplicate content, where there is actually no duplication. Data uniqueness is preserved by performing a transformation to some content, for example by concatenation of multiple fields in one or more data sources.

2.2.2.5 Classify cases
Some challenge content allows for classification of content in some form, such as ordinal ranking or grouping subcategories into lesser detailed classes, creating a more generic picture of the dataset.

2.2.2.6 Create subset by filtering
Nearly all entries for the challenge filter the dataset to create a unique view or specific focus for their research. The result is a filtered log to continue the research with.

2.2.2.7 Create new (ProM) Filter
Mainly scientific researchers provide a new way to address the data by providing code or a complete package that can be implemented in the ProM software.

2.2.2.8 Use multiple tools
Most submissions provide a rounded analysis by using multiple tools to analyze the dataset. This allows for insights from multiple disciplines. When the number of tools is limited, a flat but often straightforward analysis regarding the challenge question is provided.

2.2.2.9 Perform Case Performance Analysis
Analysis regarding one or multiple different case types is performed. Variables as throughput times, process flow, resources or alike attributes are reviewed and discussed.

2.2.2.10 Perform Process Performance Analysis
Analysis on the process itself is provided regarding bottlenecks, alternative flows and process improvements. It differentiates from the topic mentioned in 2.2.2.9 because of the collection of cases versus categories of cases.

2.2.2.11 Compare Control Flows
Last but not least, control flows are compared. The sequential result of the control flow, comparisons between algorithms or tools are presented and elaborated. In some cases, visual support is provided by comparing benchmarks with the actual process through highlighting and mapping actual versus theoretical flows.

2.2.3 Tools

In Table 4 an overview of the tools used per entry [12-42] is provided. For brevity reasons, the full cross table is not included. Per class a short description is provided in the subsequent subparagraphs. Please review the articles using the tool to obtain more knowledge about the output and the original references to the applications, as it goes beyond our scope to fully document these applications.

Table 4: Classification of tools used per BPI Challenge entry

Author \Tool classification used	*Process Mining Software*	*Data Pre-Analysis Software*	*Statistical / Spreadsheet Software*	*Programming Language(s)*	*Database Engines (SQL)*	*Process Flow / BPMN*	*Visualization Tools*
Bose & van der Aalst	Y			Y			
Caron et al.	Y		Y				
Varvaressos	Y						Y
Adriansyah & Buijs	Y						
Bautista, Wangikar & Kumail Akbar	Y		Y				
Bose & van der Aalst	Y						
Kang et al.	Y				Y		
Molka, Gilani & Zeng	Y						
Verbeek	Y						
Arias & Rojas	Y						
Bautista et al.	Y		Y				
vanden Broecke et al.	Y		Y		Y	Y	
Dudok & van den Brand	Y						
van Geffen & Niks	Y		Y				
Hansen	Y		Y			Y	
Hevia & Saint-Pierre	Y						
Kang et al	Y		Y	Y	Y		

Martens	Y						
Paszkiewicz & Picard	Y			Y			Y
Radhakrishnan & Anantha	Y		Y				
Van den Spiegel, Dieltjens & Blevi	Y		Y		Y		
Hansen	Y		Y				
Thaler et al.	Y	Y	Y		Y	Y	
Buhler et al.	Y		Y				
Dees & van den End			Y				
Suchy & Suchy	Y	Y	Y	Y			
Buffett, Emond & Goutte**							
Van den Spiegel et al.	Y	Y	Y		Y		

2.2.3.1 Process Mining Software

In the classification of Process Mining software, the following tools are used: *ProM* versions 5 and 6, *Fluxicon Disco* and *Perceptive ProcessMining*. The way Disco can be used by providing a dataset to use the full feature set for analysis is supporting obtaining experience with the software. Although for 2014 *Celonis* was a sponsor of the BPI Challenge, no participant used the software, which might be because of availability for the participants.

2.2.3.2 Data Pre-Analysis Software

In the classification of Data Pre-Analysis Software, multiple tools have been used to get a view on the data structure or prepare data for analysis in any form or subsequent tool. The following tools are used: *Fluxicon Nitro*, *Ataccama DQAnalyzer*, *CA ERwin Data Modeler*, *Weka*, *Jasper ETL*, *pgAdmin*.

2.2.3.3 Statistical / Spreadsheet software

In the classification of statistical or spreadsheet software, the following tools are used: *Minitab*, *R*, *RapidMiner* and *Microsoft Excel*. *RapidMiner* in this case could also be classified in the Data Pre-Analysis software class, because of the ETL and OLAP functionalities. Core statistical features are available in *Excel*; however more complex statistical activities are used in Minitab and R by participants. Excel is often used to provide the supporting graphs and diagrams.

2.2.3.4 Programming Language(s)

In the classification of Programming languages, we also included the development studios where mentioned: Microsoft Visual Studio and Eclipse. The programming languages used by some professionals and academia are: *.Net*, *C#*, *Java* and *Python.*

2.2.3.5 Database Engines (SQL)

In the classification of Structured Query Language Database Engines nearly the full spectrum of the top 5 as monitored by DB-Engines [44] is present: Microsoft SQL Server, MySQL, Oracle PL-SQL and PostgreSQL. Most participants used the database engine to handle the data for other tools and to easily create subsets for the data.

2.2.3.6 Process flow / BPMN
For Process flow / BPMN tool classification, the following tools are used: *Sparx Enterprise Architect*, *RefMod-Miner* and *Atris Panda* in order to visualize the process (control) flow, either automated or manually.

2.2.3.7 Visualization Tools
The last class of software tools used is the Visualization tool. These tools are used to provide clear visual representations of the data in some form or structure to support high resolution images representing big volumes of data: *D3.js*, *Gephi*, *Tableau*, *wsmostudio* and the *Classification & Regression Tree*.

2.2.4 Filters / Functions

More than 27 filters and functions for ProM 5 and ProM 6 are used for the BPI Challenge entries for 2011 through 2014. An overview of the 27 most relevant filters and functions can be found in appendix B. It goes beyond the scope of this research paper to address all the individual filters for ProM 5 and ProM 6. Further information on the filters and functionalities can be found at the website concerned with ProM and development of the open source tool [43].

2.3 Research Method design
As described by Martens [31], the SEMBA method is a highly industrialized approach to performing analysis. SEMBA is created in this industrial way to have a uniform approach across the company and provide a complete as possible view on the required scenarios, current (As-Is) and future (To-Be). In the sub-discipline of the As-Is situation, Process Mining can be applied perfectly to objectively create a view on the current process flows and process performance.

For the BPI Challenge 2015, there are two questions regarding the To-Be scenario and subsequently, some elements from SEMBA will be applied. Using the best practices from the previous challenge submissions, reproducibility is the most important factor for the research method.

3 Research Method

In this chapter the research method per challenge question will be described. First, a set of lessons learned from the literature review provided in chapter two will be applied. Then, question by question, a stepwise approach is described to reach to the answer in the subsequent subparagraphs.

3.1 Applying BPI Challenge best practices

As can be concluded from the research on BPI Challenge entries as presented in chapter two, there are a few activities to be performed. In the following sub-paragraphs an overview of each step is provided.

3.1.1 First round of analysis

In a first round of analysis, multiple approaches are used to get a basic understanding of the dataset. This allows for further definition of scoping the area of research. Confirming that the datasets are equal by comparing the datasets as provided for the tool Disco [45] and the XES file format [46,47,48,49,50] is one of the activities. Then a visual representation of a process flow is generated using Disco to get some understanding of the process. Using a process flow, one can quickly assess whether the process is straightforward or complex, whether many steps are involved, the number of variants and the timeline of the data applicable. The first round of analysis is conducted until all described elements of the dataset [08] are understood by the authors. Depending on the outcomes, further scoping steps can be applied on the analysis, the data and the time to be applied.

3.1.2 Define scope of analysis

For the challenge of 2015 the scope for the research is limited to answering the questions in a straightforward way, with no further searching for evidence or patterns beyond the question as asked, different from some authors in the previous challenges.

3.1.3 Define scope of data

The datasets [45-50] provided for this challenge are all considered for the research provided in this paper. Based on the complexity or the number of results, filters may be applied to limit the scope of the dataset. Filters can be applied to various attributes of the data such as the process instance name(s), resources, product(s), etc. Each inclusion, exclusion or enrichment is mentioned in the research outcomes.

3.1.4 Define scope of time

The authors perform this research in their spare time and knowing that this type of data is highly interesting for the profession of a business analyst, one can easily lose track of time. Therefore, a limitation in time to spend on the full execution of the research, including writing and reviewing of the content of this paper is limited to 20 hours per author. A breakdown of activities per author is approximately 6 hours of literature review, 5 hours analysis, 9 hours of writing and content review.

3.1.5 Define tools to use

From the overview as provided in Table 4, only a few tools are going to be applied for this research, based on two factors: (1) expertise of the software packages is not obtainable in the provided time scope and (2) availability of the software due to no current license. Statistical and programming software is eliminated because of time constraints. Ontology and data modeling software is excluded for expertise and

applicability for the challenge. Various process mining software is excluded based on licenses not available to the authors. Database software is considered not to be needed, as the selected scope results in a limited set of rows in the datasets to consider. Other, custom created software is not used because of unavailability to the authors, as it is custom software. For a description of the tools we refer to the vendors' product pages on vendor websites.

3.1.5.1 Disco

Discovery is one of the ten approaches to process mining in the refined Process Mining Framework, classified as "Post Mortem", or historical data, trying to apply cartography and obtaining a de facto model for a control flow [04] and therefore perfectly supports in the initial assessment and overview of the dataset. In multiple situations where process mining was applied in client projects, Disco allowed for quick understanding of the situation and allowed for a deep discussion on the matter with customer experts of the process.

The fact that the software is made available for the contest is also beneficial, as well as that it is part of the Massive Open Online Course on Process Mining as presented at Coursera [04], which allows for professionals to gain experience in the applied field.

3.1.5.2 ProM 6

ProM is used because of the platitude of analysis possibilities using plug-ins for process mining. Many of the functionalities of these plug-ins are not implemented in any other software package. Furthermore, it allows for comparing the results of various viewpoints towards the same dataset. To date, only a few applications have a feature for social network analysis such as the 'Handover of Work' analysis or the 'Dotted Chart' analysis which allows for multiple levels of insights in one graphical overview. The fact that the software is freely available is also beneficial, as well as that it is part of the Massive Open Online Course on Process Mining as presented at Coursera [04].

3.1.5.3 Excel

The well-known and widespread spreadsheet program created by Microsoft™, Microsoft Excel® is used because of its ubiquitous usage for assessing data and presenting data in a structured way. Because of a wide range of functions and features, this is the tool of choice for our activities when it comes to generating overviews of subsets of data using the 'Pivot Chart' function and standard data visualization functions like a histogram as part of the 'Charts' feature.

3.2 Addressing the customer

In SEMBA, the customer requirements are important. For the challenge the customer is addressed through by answering the questions posed. No interaction with the end user is possible as the end user is not available, impeding further research iterations. The questions 1 through 5 are treated in the subsequent subparagraphs.

3.2.1 What are the roles of the people involved in the various stages of the process and how do these roles differ across municipalities?

To find the answer to this question, we propose three activities to perform to generate the answer to question 1:

1. Identify the roles in the dataset
2. Identify stages
3. Compare stages and roles across municipalities

Roles and responsibilities are often addressed in organizations using a RASCI table, where the acronym describes the Responsible, Accountable, Supporting, Consulting and the Informed role. Applying RASCI implies a strict segregation of duties between roles. If this applicable in a dataset, this behavior is identified based on shifting of activities to distinct people fulfilling a role in the organization. People are identifiable using the resource ID.

To identify stages in the process, the descriptions of the process steps provide an insight. The naming convention as provided in the BPI Challenge description for the main process steps "XX_HOOFD_YZZ" help in defining the main process, where XX stands for the number of the main process steps, Y denotes a phase and ZZ hints towards a sequential order of sub steps.

After the roles and stages are recognized, a comparison between the five municipalities can be done by sequencing the stages and the roles performing the stages and compare this between municipalities in a descriptive manner.

3.2.2 What are the possible points for improvement on the organizational structure for each of the municipalities?

To determine the organizational structure per municipality, three activities are suggested to be performed to obtain the answer to question 2:

1. Assess resource allocation based on phase, type of activity, product mix or any other set determining an organizational structure.
2. Compare organizational structures
3. Suggest improvements could be found in the segregation of duties, performance or throughput definitions.

Due to limitations in time, the third activity for suggestions will be limited to the segregation of duties.

3.2.3 The employees of two of the five municipalities have physically moved into the same location recently. Did this lead to a change in the processes and if so, what is different?

To find the answer to this question, there is one assumption that is made: the work can also be performed by resources for other municipalities. We propose three activities to perform to question 3:

1. Identify employees working for more than 1 municipality
2. Assess moment in time of the removal and introduction date(s) of the employees
3. Compare municipality control flow before and after the earliest date of changed locations.

To identify the employees the resource numbers are listed in sequential order and compared between the five municipalities. Identical numbers are assumed to be the same employee. The remaining municipalities are used where there is an overlap between employee numbers.
A filter is applied to a municipality dataset to use only the employee numbers' activities. A time frame assessment for the starting date(s) and ending date(s) is used to assess whether the pair of employees is removed from the source municipality and the pair of employees is introduced in the target municipality.
Using a dotted chart, the activity frequency and activity type can be assessed to confirm the point in time where the pair of employees may have introduced a deviation of the pattern of activities. This exercise is applied to both municipalities.

3.2.4 Some of the procedures will be outsourced from 2018, i.e. they will be removed from the process and the applicant needs to have these activities performed by an external party before submitting the application. What will be the effect of this on the organizational structures in the five municipalities?
To find the answer to this question, we propose three activities to perform to generate the answer to question 4:

1. Identify 'some of the procedures', related to applicant and 3rd party.
2. Describe the effect that can be related to less work to be done by employees.
3. Elaborate on the effect for the municipality.

3.2.5 Where are differences in throughput times between the municipalities and how can these be explained?
To find the answer to this question, we propose four activities to perform to generate the answer to question 5:

1. Compare throughput times,
2. Identify throughput time differences between the municipalities
3. Identify performance issues locations in the control flow
4. Discuss on the possible sources of the performance issues

3.2.6 What are the differences in control flow between the municipalities?
To find the answer to this question, we propose three activities to perform to generate the answer to question 6:

1. Describe control flow per municipality
2. Show process steps sequentially in table
3. Compare all process steps

4 Analysis Results

4.1 Introduction to the results

In this section our findings after applying the outlined stepwise approach as described in chapter three for each BPI Challenge question are provided. In using the four scoping limitations as defined in paragraph 3.1.2 through 3.1.4 and the toolset as proposed in paragraph 3.1.5, the following research design elements as defined from the overview in paragraph 2.2.2 have been applied as represented in Table 5.

Table 5: Overview of applied research design elements per question and topic. Y resembles the applied design, D as used in Disco.

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
Describe dataset	-	-	-	-	Y	-
Perform Pre-Analysis	-	-	-	-	-	-
Combine data files	D	D	D	D	D	D
Preserve data uniqueness	-	-	-	-	-	-
Classify cases	Y	Y	Y	Y	Y	Y
Create subset by filtering	Y	Y	Y	Y	Y	Y
Create new (ProM) Filter	-	-	-	-	-	-
Use multiple tools	Y	-	Y	Y	Y	Y
Case performance analysis	-	-	-	-	-	-
Process performance analysis	-	-	-	-	Y	-
Compare Control flows	-	-	-	-	-	Y

For each of the results and answers, a naming convention for each municipality {M1,…, M5} is applied to represent in figures and tables. Cases have been classified based on a specific attribute “(case) parts” with a value “Kap”. This subset has been identified as being one of the top 5 types for all municipalities, as a reference to be found in Table 6. In the pre-analysis activities, the “Bouw” permits were a big “Spaghetti process” which were difficult to analyze and forced the further scoping of the dataset. An indicator for the “Spaghetti processes” is the ratio of variants versus cases being close to 1, which was found in all municipalities.

Table 6: Top ranked permit per municipality. *aggregated result, "Handelen in strijd met regels RO" is the actual value for municipalities with this marking

Name Permit \ Municipality	*M1*	*M2*	*M3*	*M4*	*M5*
Bouw	1	1	1	1	1
Kap	**2**	**2**	**2**	**2**	**3**
Milieu (vergunning)	3	3	5	6	6
Bouw, Handelen in strijd met regels RO	4	4	3	4*	2*
Sloop	5	6	4	5	5

4.2 Answer to Question 1

First the described path of actions from paragraph 3.2.1 has been applied. Actors are identified by the resource numbers and provided in the municipality data files. Considering phases, there were few particular activity segregations found other than communicative such as sending confirmations which might be associated with supporting staff. An overview of all phases and frequencies are found in appendix D. For Municipality 4, resource 1550894 performed some decisions.

In Municipality 5 we see some differences in activities across the resources, where some more supporting actions are taken by a limited set of employees.

The overview can be replicated using pivot tables in Excel by cross examining the "Resource" as columns and the "(case) Responsible_Actor" and "Activity" as rows where the "Activity" serves as a clarification of what is happening in the "concept:name" and providing the "Count of Activities" as the Sigma values.

Following RASCI, a resource is the Responsible person, whereas the Responsible Actor is the Accountable person. A comparison of the Actors and Responsible Actors per municipality for the "Kap" permit handling is provided in Table 7. From the results we see that there are clear differences between number of actors per municipality and the work that is performed by each actor, dubbed the "Main Actors" for producing more than 20% of the work or the top 3 employees. This is supported by the histograms Figure 1 showing the distribution of productivity per resource.

One particular point is important to mention is that in Municipalities 1, 3 and 5, the segregation of duties is unclear, where the Responsible Actor is also performing the most activities in the permit process as Main Actor.

Table 7: Comparison of Actors and Responsible Actors between the five municipalities

Municipality	*M1*	*M2*	*M3*	*M4*	*M5*
Actors	7	6	9	5	14
Main Actors	1	1	2	1	3
(% work) Resource ID	(86%) 560872	(62%) 560530	(40%) 560673 (35%) 560749	(65%) 560781 (21%) 560752	(38%) 560596 (18%) 560602 (12%) 1254625
Activities	3345	4053	9131	5335	2597
Responsible Actors	7	4	6	3	5
Responsible ID **(Responsible & Accountable in bold)**	(53%)4901428 **(28%) 560872**	**(40%) 560530** (34%) 560458	(47%) 560696 (26%) 3069866 (25%) 3443734	(84%) 560812 (15%) 560431	**(73%) 560596** (9%) 560604 (9%) 560429

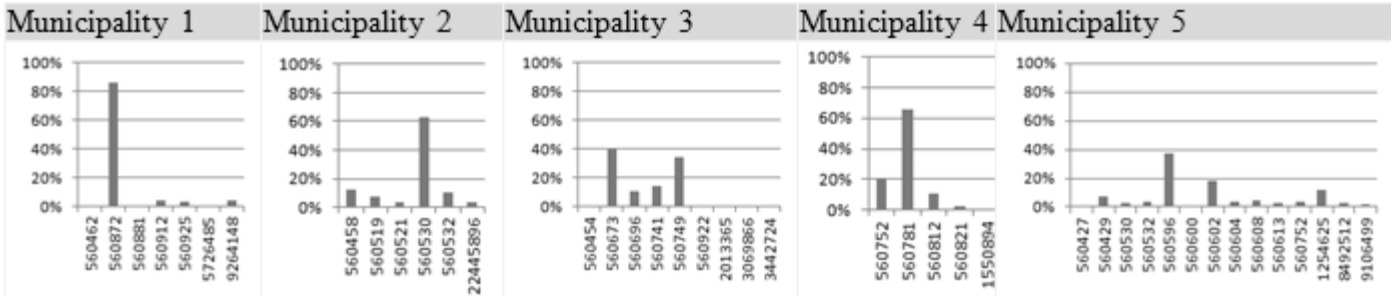


Figure 1: Overview of the productivity per Actor per Municipality

4.3 Answer to Question 2

The answer to Question 1 allows for some additional input for Question 2. Based on the segregation of duties, there are three municipalities that have unclear situations, M1, M2 and M5. By reviewing the resources and the types of activities performed for the full dataset, there are indications that there is specialization on environmental permits for M2 and M5. Results from the answer to question 4 indicate a possible improvement in the area of structuring permit Procedure classification.
Based on the research approach from 3.2.3, the improvement points are provided in Table 8.

Table 8: Application of improvement suggestions

Improvement \ Municipality	*M1*	*M2*	*M3*	*M4*	*M5*
Assess Segregation of Duties	Yes	Yes			Yes
Specialize work packages	Yes		Yes	Yes	
Clearly structure Procedure	Yes	Yes	Yes	Yes	Yes

4.4 Answer to Question 3

As a preparation step, data is exported from *Disco* as CSV file and loaded into *Excel*. Results are obtained using the PivotTable functionality in *Excel*. The dataset used is the full set for all five municipalities, no filters applied. The stepwise approach as mentioned in paragraph 3.2.3 is applied. In Appendix C a full overview of all resources per municipality is provided. In Table 9 a limited overview is provided of the selection of resources working for more than one municipality.

Table 9: Overview of resources working for more than one municipality

Resource Number	*M1*	*M2*	*M3*	*M4*	*M5*
6	Y		Y	Y	
560429		Y			Y
560530		Y			Y
560532		Y			Y
560598		Y			Y
560752				Y	Y
560849				Y	Y

From the results of the resources working for more than one municipality an overview is created for the active periods per resource as shown in Table 10. The activities for some resources are limited, i.e. resource 6, which incidentally performs actions. Resource number “6” is low, not comparable with all other resource number patterns and infrequent related to activities that it seems that this is a system or temporary employee.

Table 10: Overview of all resources working in multiple municipalities indicated by "X" in which quarter of a year a resource was active. An "I" indicates an incidental activity record. Color coding of the background indicates the municipality. Underlined resource numbers are the resources in scope.

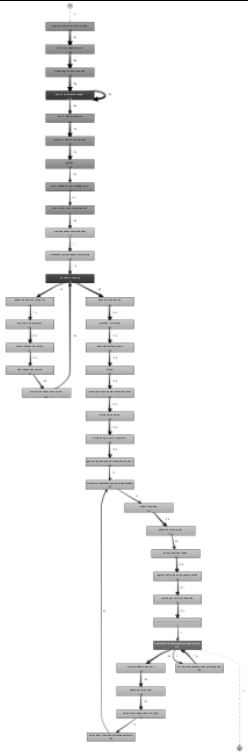
	Cases	*Municipality*	*2010*				*2011*				*2012*				*2013*				*2014*				*2015*
Resource \ Quarter			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1
6	1	1				*I*																	
	1	3				*I*																	
	1	4																	*I*				
560429	2	2																*I*					*I*
	332	5				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
560530	467	2				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
	45	5																			X	X	X	X
560632	534	2		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	93	5																			X	X	X	X
560598	10	2											X	X	X	X								
	48	5				X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X
560752	926	4				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	107	5									X		X	X	X									
560849	61	4				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	9	5																	*I*	*I*	*I*	X	X	X

With the results on further scoping to the resources to focus on, the combination of municipality to review and resources' related time frame is municipality 5 and resources {560530, 560632, 560849}. We obtain the earliest date of action for the new municipality for full-time work, being 12/05/2014.

Disco is used to further analyze the control flow, by applying a filter to the dataset to create a 'before' and 'after' plot of the control flow. The "Started in Timeframe" setting is applied, where the starting date is the first date in the dataset for M5, the last date is 12/05/2014 for the 'Before' control flow. The 'After' control flow is created in an equal manner, setting the starting date as 12/05/2014 and the end date as the last date in the dataset for M5. Table 11 shows the resulting control flows.

The analysis on a visual basis shows that there are few alternative routes no longer possible after 12/05/2015 as seen in the last part of the control flow; hence there is a change in the process.

Table 11: Overview of the control flow of 'before' and 'after' 12/05/2014 when the assumed location combination started in municipality 5.

Before 12/05/2014	*After 12/05/2014*
	

4.5 Answer to Question 4

Reviewing the structure of the dataset, there is one variable named "caseProcedure" which has specific naming conventions. We assume that this is the attribute to review for the assessment for outsourcing. An overview of the figures related to each municipality is provided in Table 12 and was obtained by using the Pivot Table function in *Excel* using the complete dataset (not filtered) at first. Then, the columns are defined by the resource IDs, the rows are defined by "caseProcedure" and the Activity count is provided in the sigma Values. Our assumption is that the variety "Regulier" which translates to "Common" procedures will be outsourced.

Unfortunately, the number of Common cases is quite low, whereas there is limited classification for the bulk of cases, as marked "Blank". Without further interaction with the customer, we cannot conclude the full impact of the outsourcing activities. From practical experience however we can provide insight in the fact that outsourcing work means that in most cases the work performed by employees will decrease and sometimes that roles are removed from the organization as known in the As-Is situation. For the municipalities this means in M3 that resource 3069866 will not have

work in future for Common cases. This resource only performs 22 actions out of a total of 59.681 in case of this municipality and for the five municipalities total of 262.628 actions shared across the total amount of 72 resources the social impact is low. As outsourcing will remove all common cases, this means that there is more focus possible on procedures in the "Uitgebreid" category. This might have an impact, ideally improving one of the three common variables Time, Quality or Money. One important remark is that with the unclear cases marked "Blank" there is a potentially hidden set of cases that might require adjustment of the statements above.

Table 12: Overview of the split in case procedure types as registered per municipality, with counts based on the Activity

caseProcedure	*M1*	*M2*	*M3*	*M4*	*M5*
Regulier	818	2.792	1.309	230	902
Uitgebreid	4.861	10.834	6.805	6.629	5.390
Blank	46.538	30.728	51.567	40.434	52.791
Total	52.217	44.354	59.681	47.293	59.083

4.6 Answer to Question 5

As stated in paragraph 3.2.5, the throughput times are retrieved and compared between municipalities. The dataset applicable is the subset of "Kap" permits, created with a filter in *Disco* in the filter settings on an attribute, for the "(case) parts" attribute where the value is "Kap". The cases are then in the Statistics, "Overview view exported as a .CSV file and loaded in *Excel* to be converted into number of hours for the total throughput time. Standard descriptive formulas are used on the data of all municipalities, stating the number of cases, the minimum throughput time in hours, the maximum throughput time, the mean and the median values. The resulting overview is provided in Table 13. Based on the differences of the extreme values, the mean and median values between the municipalities, a quick overview whether there is a normal distribution found in the cases was performed. The resulting histograms are found in Figure 2. Concluding from the Mean values between the municipalities, M1, M4 and M5 show similarities. Municipality M3 seems to be very efficient in the throughput time judging from the Mean, Median and Maximum values. Judging from the high value for the Minimum value and the Histogram for Gaussian distribution, there seems to be a normal distribution applicable for M3.

Table 13: Overview of descriptive statistics for all municipalities on throughput time.

Dataset (h)	*M1*	*M2*	*M3*	*M4*	*M5*
Cases (N=)	90	95	262	154	67
MIN	15,44	17,51	58,15	12,01	57,36
MAX	6.719,00	6.682,45	4.357,46	6.609,12	32.255,00
Mean	1.941,62	1.428,62	554,75	1.806,87	2.252,87
Median	1.980,00	1.282,87	456,00	1.728,50	1.574,41

The process flows are now reviewed in *Disco* and analyzed on the Performance view, with Activities and Path sliders set to 0%. The performance is reviewed on the "Mean duration", to identify differences in the process. The results are presented in Table 14.

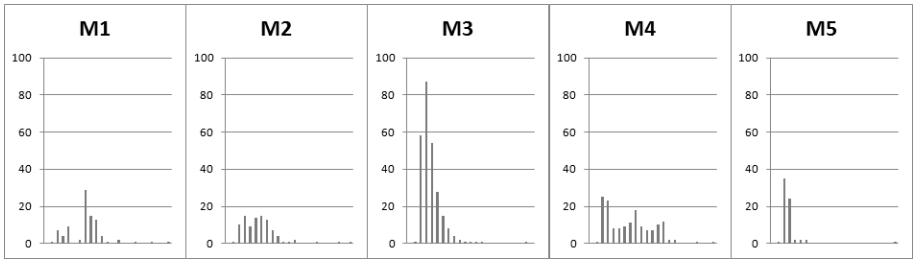


Figure 2: Histogram per municipality to determine whether a Gaussian (Normal) distribution is applicable for the throughput time of cases.

Table 14: Comparison of process flows and delays in the Mean duration per municipality

	M1	*M2*	*M3*	*M4*	*M5*
Location 1	End	Start	Start	End	Middle
From	Start decision phase decision permitting sent	Enter send date acknowledgement	Enter send date acknowledgement	Phase decision sent	Create publication document
To	Register Deadline	Create sub cases completeness	start WABO procedure	Read publication date field	Phase archived case
Location 2		End			
From		Enter senddate decision environmental permit			
To		Registration date publication			

From the results in Table 14 it becomes clear that each municipality has a different point in the flow where there are delays. There is one commonality in the activity for M2 and M3 showing the "Enter send date acknowledgement" activity for starting point of the delay. There seems to be a difference in the start, the middle or the end of the set of activities. Explanations in the start are possibly related to incomplete cases and required further input from the applicant, but in other cases it is possible that assigning the correct person or determining the correct types of sub applications is causing delay. The delay found in the middle for M5 might actually not be a real delay; in The Netherlands there is an obligation to make applications for permits publicly known, where there is a period of time allowed to state objections from the public. This is typically something to discuss with the client. Delays in the end might be related to the decision that is sent and the need to record a final statement or archiving step is less pressing. In most cases as researched for questions 1-4, it is found that the final archiving step is only applied in a few cases, where one might assume that all cases are archived. High numbers of missing trace elements might indicate that the software registering these steps is wrongly configured.

4.7 Answer to Question 6

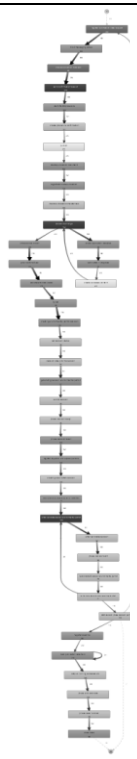
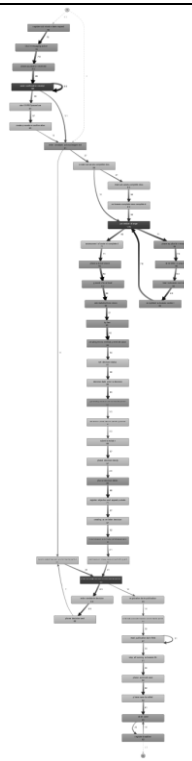
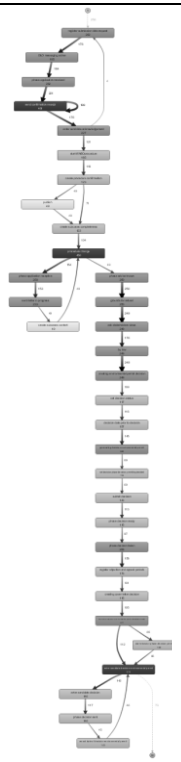
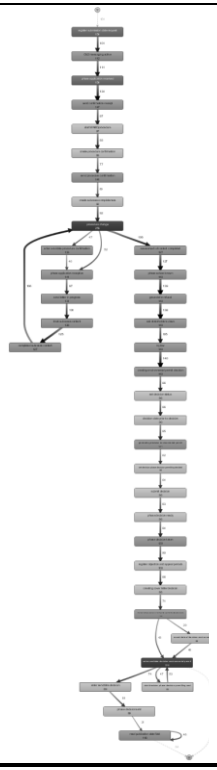
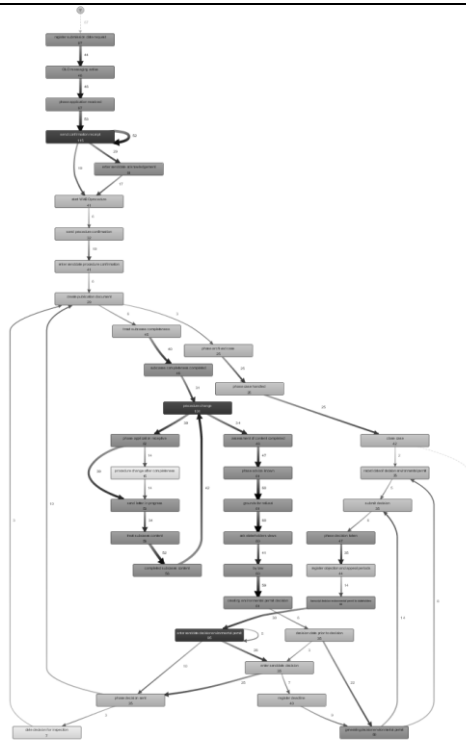
Following the steps identified in paragraph 3.2.6, the control flow per municipality is analyzed by obtaining the control flow from *Disco*. The dataset preparation is equal to the previous answers, focusing on "Kap" permits by applying a filter on attribute for the "(case) parts" attribute where the value is "Kap". Additionally a filter on attribute is applied filtering for the process starting with "01_HOOFD_*" process steps on attribute "concept:name". The control flows are found in **Error! Reference source not found.**. The observed differences between the five municipalities are listed in Table 15.

Table 15: Observed defining attributes in control flows

Observations	*M1*	*M2*	*M3*	*M4*	*M5*
Sub Route	Y	Y	Y	Y	
Self-Loops		Y	Y		Y
Long sequence	Y	Y	Y	Y	
Shortcut in top path		Y	Y	Y	
Number of activities in Sub Route	3	4	3	5	-

Comparing the flows of the five municipalities in Table 16, we see a lot of comparable elements in the main flows for M1 through M4, such as the existence of a sub-route, looping back to the main flow and a long sequential flow of activities. The fifth municipality has a different approach for the control flow. This might be related to the lower number of cases as can be seen in Table 13, creating more influence for a few variants. However, in case of uniform distribution of cases, this should not be the case as the ratio of cases would be equal. Furthermore, there are two distinctive characteristics for municipalities being self-loops and Shortcuts in the top path. Self loops are found in M2, M3 and M5. A shortcut in the starting sequence is found in M2 through M4 found respectively.

Table 16: Process flow per Municipality for “Kap” and filtered on 01_HOOFD_*

	M1	*M2*	*M3*	*M4*	*M5*
Process flow					

4.8 Conclusions for the results

The limitations as mentioned in paragraphs 3.1.2 through 3.1.4 allow only for quite straightforward research and answer generation. The datasets for the municipalities are fit for quantitative research, but less fit for applying comparable visual process flow analysis due to a large number of activities recorded and obfuscation of actual processes due to the high number of variants.

In this part of the paper, the answers as found have been provided, and in multiple cases the absence of a possibility to interact with the client has lead to results that are of less impact than anticipated beforehand.

Providing the answers in a structured manner in combination with the approach described in chapter 3 per question allowed for starting points for highly reusable approaches of providing insights related to similar situations and/or data.

5 Conclusions

5.1 Answers to the questions

The six questions from the municipalities are answered in multiple aspects. The roles as identified in answers to question 1 show that there might be a need to review the segregation of duties for municipalities 1, 2 and 5, as is also concluded as an improvement point in answers to question 2. The number of employees as described in answers to question 1 is different for each municipality and in municipalities 1, 2 and 4 there is one person handling most of the activities in relation to “Kap” permits.

Work is shared more in municipalities 3 and 5, where in 3 the process is assumed to be quite efficient judging from the answer to question 5. Regarding points for improvement, there might be a benefit to elaborate on what happens in municipality 3 to be shared in the location where employees for municipality 2 and 5 are located. As we have seen in answers for question 3 there is an impact in the process possible after working together in one location.

Improvements might be unknown to be proven by the outsourcing decision for 2018 as requested as an answer for question 4. The dataset shows that there is more required to assess “Blank” procedure cases whether they are of one type, or comprise of unclassified cases, obfuscating the actual number of “Regulier” versus “ Uitgebreid” cases. Furthermore, there are indications of combined cases which can be seen in the naming convention applied in the variable “(case) parts”.

Further analysis is beneficial for answering question 5 in detail, as there is unclear whether the delays found are customer induced, part of standard procedure or lack of execution of archiving activities, or possibly a configuration error of the recording software. In the subset of “Kap” permits, there is a clear difference to be found between the first four municipalities and the fifth municipality which is the answer to question number 6, differences are furthermore found in shortcuts and self-loops.

5.2 Business value

With this report, the business use of addressing customer questions has been provided. It is a showcase of how business analysis is applied in the context of a SEMBA approach for the current As-Is situation and the future To-Be situation. Furthermore, it shows that Process Mining has a good fit in the set of the tools and techniques applicable for Business Analysis.

The "Spaghetti processes" identified in the datasets in combination with the unclear definitions of process phases, incomplete case procedure registrations and combinations of naming conventions show clear evidence for the need of CMMN as a methodology to address non-Straight Through Processes (STP). Case Management is found mostly in public scenarios, where each case can be different from what has been processed before. An indicator for Case Management is suggested to be the ratio between variants and cases being close to 1.

BPM systems are undervalued in The Netherlands (position 10 vs. average 4 in EU) as indicated by research by Luftman, whereas there is high value if there is true unification of processes possible by using clear process definition. The amalgamation of municipalities has a high correlation to business unit process harmonization across countries for the for-profit industries. Having a process that is predictable and controllable has impact on the service level, business process maturity and henceforth the experience by the municipality inhabitant for applications in common but life changing building permits. Therefore, using the best practices which might be found in municipality 3 to be shared with all other municipalities might provide such benefits.

5.3 Scientific Value

In chapter 2 there are multiple overviews provided to support in the development of the research for Process Mining for professionals and researchers. The stepwise approach and best practices found by assessing previous BPI Challenge entries provide clear guidance and overviews for methodologies, tools, filters and functions applied.

In this report a clear connection between research and business practice has been established bringing both worlds closer together to show the need to log data, to provide valuable insights and showing the requirement of sound definitions so all participants can understand what is happening in complex situations.

6 Recommendations

The recommendations are related to the professional field of Process Mining as a practice, not the dataset for the municipalities. As lessons learned from this BPI Challenge 2015, we would like to put the attention to a formal template to perform research per domain of questions. The SEMBA industrialized method performs well in a business environment because of repeatability and stability of the outcomes when

applied. Judging from the reviewed submissions for the past BPI Challenges, we see a wide spectrum of answers and methods which, depending on the viewpoint is great (in case of exploration) or impractical (in case of uniform results). The current success in business adoption of Process Mining is in our vision found in a structural reporting approach such as applied in the Coursera course. This is a great start but lacks a formal research template for preparing data, classification of research steps for certain types of analysis and in some cases, the cross over points on how to use which kind of toolset to assess results.

In order to support the development of Process Mining problem solving, Taxonomy of research approaches is a possibility to consider, synchronized with the refined process mining framework.

As a final recommendation, we would like to challenge the setup of the BPI Challenge, to try and come with a situation that will drive for more comparable results between participants' submissions, for example by introducing one question which must be answered by using one tool and one or two filters, which is then discussed as a final topic in the closing words of the proceedings by the members of the jury.

Disclaimer

References

01. ISO (International Organization for Standardization): ISO/IEC 19510:2013, Object Management Group Business Process Model and Notation (2013) Retrieved from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=62652
02. OMG (Object Management Group): Case Management Model and Notation (2014) Retrieved from: http://www.omg.org/spec/CMMN/Current
03. Capgemini Academy: Curriculum voor Business Analysis – Professional (2014) Retrieved from: http://academy.capgemini.com/binaries/content/assets/capgemini-academy-nl/curricula/curricula-2014/curriculum_map_ba.pdf
04. van der Aalst, W.M.P.: Process Mining: Data science in Action. (2014) Participated through Coursera: https://www.coursera.org/course/procmin
05. van der Aalst, W.M.P.: Process mining: discovery, conformance and enhancement of business processes, Springer, Heidelberg ISBN: 978-3-642-19345-3 (2011)
06. Luftman, J.: Influential IT Management Trends: An International Study. (2014) Retrieved from: http://www.globaliim.com/app/download/961894350/2014+Global+IT+Trends+Paper.pdf
07. Luftman, J., Derksen, B.: European key IT and management issues & trends for 2014. CIONET Europe and Business & IT Trend Institute (2014). Retrieved from: http://blog.cionet.com/wp-content/uploads/2014/02/ITTrends_2014print.pdf
08. van Dongen, B.F.: BPI Challenge 2015 dataset. Accessed on 18-05-2015. Retrieved from: http://dx.doi.org/10.4121/uuid:31a308ef-c844-48da-948c-305d167a0ec1

09. Merriam-Webster online dictionary: Productivity| Definition of Productivity by Merriam-Webster (2015) Retrieved from: http://word.com/dictionary/productivity
10. Wynia, S.: Er wart een spook door onze gemeenten. Elsevier, 20-11-2014 (2014), Retrieved from: https://blendle.com/i/elsevier/er-waart-een-spook-door-onze-gemeenten/bnl-elsevier-20141120-18639_er_waart_een_spook_door_onze_gemeenten
11. Allers, M.A., Geertsema, B.: The effects of local government amalgation on public spending and service levels. Evidence from 15 years of municipal boundary reform. Groningen : University of Groningen, SOM research school, 2014. SOM Research Reports; Vol. 14019-EEF (2014), Retrieved from: http://irs.ub.rug.nl/ppn/382645367
12. Ghoshal, S.: Bad management theories are destroying good management practices. Academy of Management learning & education 4.1 pp.75-91. (2005)
13. Roest, N., Theuns, M., Pijnenburg, H.: Informatiemanagersurvey 2012-2013, Capgemini (2014), Retrieved from: https://www.nl.capgemini.com/resource-file-access/resource/pdf/informatie_management_survey.pdf
14. Bose, J.C., van der Aalst, W.M.P.: Analsyis of Patient Treatment Procedures. BPI Challenge 2011 (2011) Retrieved from: http://www.win.tue.nl/bpi/_media/2011/bose.pdf
15. Caron, F, Vanthienen, J., De Weerdt, J., Baesens, B.: Beyond X-Raying a Case-Flow: Adopting Different Focuses on Care-Flow Mining. BPI Challenge 2011 (2011) Retrieved from: http://www.win.tue.nl/bpi/_media/2011/caron.pdf
16. Varvaressos, G.: Semantic Process Mining. BPI Challenge 2011 (2011). Retrieved from: http://www.win.tue.nl/bpi/_media/2011/varvaressos.zip
17. Adriansyah, A., Buijs, J.C.A.M.: Mining Process Performance from Event Logs – The BPI Challenge 2012 Case Study (2012) Retrieved from: http://www.win.tue.nl/bpi/_media/2012/adriansyah.pdf
18. Bautista , A.D., Wangikar L., Kumail Akbar, S.M.: Process Mining Driven Optimization of a Consumer Loan Approvals Process. BPI Challenge 2012 (2012) Retrieved from: http://www.win.tue.nl/bpi/_media/2012/bautista.pdf
19. Jagadeesh Chandra Bose, R.P., van der Aalst, W.M.P.: Process Mining Applied to the BPI Challenge 2012: Divide and Conquer While Discerning Resources. BPI Challenge 2012 (2012), Retrieved from: http://www.win.tue.nl/bpi/_media/2012/bose.pdf
20. Kang, C.J., Shin, C.K., Lee, E.S., Kim, J. H., An, M.A.: Analyzing Application Process for a Personal Loan or Overdraft of Dutch Financial Institute with Process Mining Technique. BPI Challenge 2012 (2012), Retrieved from: http://www.win.tue.nl/bpi/_media/2012/kang.pdf
21. Molka, T., Gilani, W., Zeng, X.J.: Dotted Chart and Control-Flow Analysis for a Loan Application Process. BPI Challenge 2012 (2012), Retrieved from: http://www.win.tue.nl/bpi/_media/2012/molka.pdf
22. Verbeek, H.M.W.: BPI Challenge 2012: The Transition System Case. BPI Challenge 2012 (2012), Retrieved from: http://www.win.tue.nl/bpi/_media/2012/verbeek.pdf
23. Arias, M., Rojas, E.: Volvo Incident and Problem Management Behavior Analysis. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 1, Retrieved from: http://ceur-ws.org/Vol-1052/paper1.pdf
24. Bautista, A., Akbar, S., Alvarez, A., Metzger, T., Reaves, M.: Process Mining in Information Technology Incident Management: A Case Study at Volvo Belgium. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 2, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper2.pdf
25. vanden Broucke, S., Vanthienen, J., Baesens, B.: Volvo IT Belgium VINST. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 3, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper3.pdf

26. Dudok, E., van den Brand, P.: Mining an Incident Management Process. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 4, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper4.pdf
27. van Geffen, F., Niks, R.: Accelerate DMAIC using Process Mining. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 5, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper5.pdf
28. Hansen, J.: Analyzing Volvo IT Belgium's Incident and Problem Management Data Using Automated Business Process Discovery. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 6, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper6.pdf
29. Hevia, J., Saint-Pierre, C.: Analyzing Volvo Information with Process Mining. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 7, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper7.pdf
30. Kang, C.J., Kang, Y. S., Lee, Y. S., Noh, S., Kim, H. C., Lim, W. C., Kim, J., Hong, R.: Process Mining-based Understanding and Analysis of Volvo IT's Incident and Problem Management Processes. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 8, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper8.pdf
31. Martens, J.K.J.: Professional Use of Process Mining for Analyzing Business Processes. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 9, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper9.pdf
32. Paszkiewicz, Z., Picard, W.: Analysis of the Volvo IT Incident and Problem Handling Processes using Process Mining and Social Network Analysis. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 10, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper10.pdf
33. Radhakrishnan, S., Anantha, G.: Process Improvement Focused Analysis of VINST IT Support Logs. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 11, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper11.pdf
34. Van den Spiegel, P., Dieltjens, L., Blevi, L.: Applied Process Mining Techniques for Incident and Problem Management. Proceedings of the 3rd Business Process Intelligence Challenge, CEUR-Workshop Proceedings, Volume 1052, Paper 12, ISSN: 1613-0073 (2013) Retrieved from: http://ceur-ws.org/Vol-1052/paper12.pdf
35. Buhler, P., O'Callaghan, R., Aubry, S., Dejoy, D., Kuo, E., Shoup, N., Khosla, I., Ginsburg, M., Hartman, N., Mcbride, N.: Service Desk and Incident Impact Patterns Following ITIL Change Implementation. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_6.pdf
36. Suchy, J., Suchy, M.: Predictive Model for supporting ITIL Business Change Management Processes. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_8.pdf
37. Dees, M., Van den End, F.: A Predictive Model for the Impact of Changes on the Workload of Rabobank Group ICT's Service Desk and IT Operations. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_7.pdf
38. Buffett, S., Emond, B., Goutte, C.: Using Sequence Classification to Label Behavior from Sequential Event Logs (Most recent version). BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_14.pdf

39. Van den Spiegel, P., Dieltjens, L., Blevi, L., Verdickt, J., Albertini, P., Provinciael, T.: Applied data mining and process mining techniques for analyzing the impact of a change on the workload of the Service Desk and IT Operations. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_9.pdf
40. Hansen, J.:BPIC Challenge 2014 - John Hansen. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_2.pdf
41. Thaler, T., Knoch, S., Krivograd, N., Fettke, P., Loos, P.: ITIL Process and Impact Analysis at Rabobank ICT. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_4.pdf
42. Arias, M., Arriagada, M., Rojas, E., Saint-Pierre, C., Sepúlveda, M.: Rabobank: Incident and change process analysis. BPI Challenge 2014 (2014) Retrieved from: http://www.win.tue.nl/bpi/_media/2014/bpic2014_submission_15.pdf
43. Process Mining Group, Math&CS department, Eindhoven University of Technology: ProM. Retrieved from: http://www.processmining.org/prom/start
44. Db-engines.com: DB-Engines Ranking – popularity ranking of database management systems. Accessed 02-07-2015. Retrieved from: http://db-engines.com/en/ranking
45. Rozinat, A.: BPI Challenge 2015. Retrieved from: http://fluxicon.com/blog/2015/05/bpi-challenge-2015/
46. van Dongen, B.F.: BPI Challenge 2015 Municipality 1. Eindhoven University of Technology. Dataset. (2015) http://dx.doi.org/10.4121/uuid:a0addfda-2044-4541-a450-fdcc9fe16d17
47. van Dongen, B.F.: BPI Challenge 2015 Municipality 2. Eindhoven University of Technology. Dataset. (2015) http://dx.doi.org/10.4121/uuid:63a8435a-077d-4ece-97cd-2c76d394d99c
48. van Dongen, B.F.: BPI Challenge 2015 Municipality 3. Eindhoven University of Technology. Dataset. (2015) http://dx.doi.org/10.4121/uuid:ed445cdd-27d5-4d77-a1f7-59fe7360cfbe
49. van Dongen, B.F.: BPI Challenge 2015 Municipality 4. Eindhoven University of Technology. Dataset. (2015) http://dx.doi.org/10.4121/uuid:679b11cf-47cd-459e-a6de-9ca614e25985
50. van Dongen, B.F.: BPI Challenge 2015 Municipality 5. Eindhoven University of Technology. Dataset. (2015) http://dx.doi.org/10.4121/uuid:b32c6fe5-f212-4286-9774-58dd53511cf8

Appendix A: About the Authors

Jef Martens is a Senior Business Analyst for Capgemini and Business Process Management consultant. As part of the Capgemini Expert Program you can easily reach out to him here: Expert Connect

.

Paul Verheul is a Business Consultant for Capgemini specialized in process optimization.
He has a special interest in Master Data Management. LinkedIn

Appendix B: Overview of Filters and Functions used of ProM{5,6}

Author \ Filter or Function	*Fuzzy Man Miner*	*Enhanced Fuzzy Miner*	*Heuristic Miner*	*Originator by task*	*Social Network Miner (Handover of Work)*	*LTL checker*	*Log Summary*	*Pattern Abstractions*	*Performance Matrix*	*Simple Log Filter*	*Transition System Miner*	*Transition System Analyser*	*Log Filter*	*Dotted Chart*	*Alpha Algorithm*	*ILP Miner*	*Conformance Checker*	*Petri-Net Complexity Analysis*	*Resource Work Analsysis*	*Guide Tree Miner*	*Trace Alignment with Guide tree*	*Sequence Clustering*	*Trace Alignment*	*Control Flow*	*Statistical analysis*	*Footprint Matrix*	*Counting Flows*	*Markov Chain*
Bose & van der Aalst		Y																					Y					
Caron et al.			Y	Y	Y	Y																						
Varvaressos			Y																									
Adriansyah & Buijs			Y				Y		Y				Y	Y	Y	Y												
Bautista, Wangikar & Kumail Akbar																								Y				
Bose & van der Aalst	Y		Y			Y		Y									Y	Y	Y	Y	Y							
Kang et al.					Y																			Y	Y			
Molka, Gilani & Zeng	Y		Y											Y														

Verbeek				Y	Y	Y					
Arias & Rojas		Y	Y					Y	Y		
Bautista et al.										Y	
vanden Broecke et al.		Y									
Dudok & van den Brand		A	A								A
van Geffen & Niks											
Hansen										Y	Y
Hevia & Saint-Pierre	Y	Y	Y								
Kang et al											
Martens											
Paszkiewicz & Picard							Y				
Radhakrishnan & Anantha											
Van den Spiegel, Dieltjens & Blevi			Y								
Hansen											
Thaler et al.	Y	Y									
Buhler et al.											
Dees & van den End											
Suchy & Suchy											
Buffett, Emond & Goutte											
Van den Spiegel et al.											

Y = Yes, filter or function applied

A= Alternative for the same filter or function is applied

Appendix C: Overview of resources working in each municipality

Table 17: Overview of resources working in a municipality. Bordered records highlight a resource working in more than 1 municipality

Resource Number	*M1*	*M2*	*M3*	*M4*	*M5*
6	Y		Y	Y	
560427					Y
560429		Y			Y
560431				Y	
560454			Y		
560458		Y			
560462	Y				
560464	Y				
560504					Y
560519		Y			
560521		Y			
560528		Y			
560530		Y			Y
560532		Y			Y
560589	Y				
560594					Y
560596					Y
560598		Y			Y
560600					Y
560602					Y
560604					Y
560608					Y
560613					Y
560673			Y		
560696			Y		
560713			Y		
560741			Y		
560749			Y		
560752				Y	Y
560781				Y	
560796				Y	
560812				Y	
560821				Y	
560849				Y	Y
560852				Y	
560872	Y				
560881	Y				
560890	Y				
560894	Y				
560912	Y				
560922			Y		
560925	Y				

Resource Number	*M1*	*M2*	*M3*	*M4*	*M5*
560950	Y				
560999	Y				
1254625					Y
1550894				Y	
1898401	Y				
2013365			Y		
2670601	Y				
3069866			Y		
3122446			Y		
3148844			Y		
3175153	Y				
3273854	Y				
3442724			Y		
4634935		Y			
4936828	Y				
5025869			Y		
5726485	Y				
6925826					Y
6993893					Y
8492512					Y
9106499					Y
9264148	Y				
10716070	Y				
11345232	Y				
11744364	Y				
12941730	Y				
13412010					Y
13412649					Y
20987361		Y			
22445896		Y			

Appendix D: Phases and frequencies of activities in phase

Table 18: Phases and frequencies of activities per phase per municipality. Colors range from the lowest number highlighted in red, through green for the most frequent occurring activity counts. No counts are highlighted grey.

Phase \ Municipality	*M1*	*M2*	*M3*	*M4*	*M5*
fase aangepast plan gevraagd	2	2	63	2	14
fase aangepast plan ontvangen	2	1	51	1	13
fase aanvraag ontvangen	1199	829	1409	1052	1155
fase aanvraag ontvankelijk	940	707	1180	885	985
fase aanvullende gegevens gevraagd	181	102	210	109	212
fase aanvullende gegevens ontvangen	149	75	167	83	167
fase activiteit meldingplichtig	1	2	3	5	2
fase activiteit vergunningvrij	1	1	2	4	13
fase advies bekend	933	709	1238	877	987
fase beroep aangetekend	3	6		2	
fase beschikking gereed	357	301	535	425	416
fase beschikking verzonden	494	386	606	461	503
fase beslissing aangehouden	23	10	32	17	8
fase besluit genomen	742	571	896	704	747
fase besluit onherroepelijk	105	343	17	335	313
fase bezwaar ingediend	3	16	1	3	5
fase buiten behandeling gelaten	31	17	18	15	35
fase concept beschikking gereed	54	1			
fase concept ontwerpbeschikking gereed	87	148	118	85	83
fase doorgezonden aan bevoegd gezag	5	1		1	
fase hoger beroep aangetekend		2			
fase instellen	4	26	3	4	5
fase instellen besluit vernietigd	9	71	2	60	116
fase nadere gegevens gevraagd		1			
fase nadere gegevens ontvangen		1			
fase ontwerpbeschikking gereed	29	40	46	35	31
fase ontwerpbeschikking verzonden	4	4	5	3	8
fase ontwerpbesluit genomen	122	217	157	122	121
fase procedure afgebroken	75	51	31	95	84
fase procedure hervat	18	10	30	1	4
fase procedure tussentijds beeindigd	34	5	15	40	12
fase proceduretermijn hervat	3	5	9	1	3
fase proceduretermijn opgeschort	3	6	17	3	3
fase proceduretermijn verlengd	63	40	31	30	11
fase zaak afgehandeld	107	352	21	21	359
fase zaak gearchiveerd	108	352	23	23	359
fase zienswijze ingediend	16	19	16	7	12