Professional use of Process Mining for analyzing Business Processes

Josef K.J. Martens

Capgemini, Cluster TRAILS, Reykjavikplein 1,
3543 KA Utrecht, The Netherlands
jef.martens@capgemini.com
j.k.j.martens@alumnus.utwente.nl

Abstract: A professional application of Process Mining has been established in the context of a methodology as defined by a consultancy firm. The results of the research show where in the context of consultancy Process Mining is used and how clients can benefit from expertise and standardized work.

Keywords: Process Mining, Consultancy, Business Analysis, SEMBA, Business Process Management.

1 Introduction

1.1 Reason for research

Luftman et al. analyze IT management issues and multiple topics have been identified and known to shift in importance over time [1][2][3][4][5]. Results by Luftman and Ben-Zvi [1] shows that the topic of *Business productivity and cost reduction* is the most important issue for C-level management in 2010. From the top 5 topics from 2010, three issues (*Business productivity and cost reduction, IT and Business alignment, IT reliability and efficiency)* are presented in some form in the BPI challenge 2013 [6]. Process Mining allows for analysis of raw data sets to discover process flows and analyze the important elements related to such flows [7].

1.1.1 Business productivity and cost reduction

Business productivity is measured by Key Performance Indicators (KPI) which are common in mature businesses, aligned to strategic and tactical goals and drive the decision making processes [8]. Common KPIs are constructed by evaluating data against a benchmark value about input, output and throughput of business processes and their related waste and outage.

Having an excellent productivity performance with maximized effectiveness of expenses on the operation of assets and employees allow for margin translates to a maximization of profit in the case of for profit organizations. To achieve this

optimum state, it is essential to know what the operational performance is of the complete- and sub-system of the organizational processes.

1.1.2 IT and Business Alignment

Bridging the gap between Business and IT is one of the most challenging activities for IT and Business professionals, it has been "…a top concern of IT managers for almost 30 years" [1]. Blum et al. [9] researched the position of the information manager, a role which is concerned with many of the issues as described by Luftman et al. and concluded that the organizational maturity defines the position and importance acceptance of higher management to solve such issues.

The Business Analyst (BA) is the role that inhibits a set of competences that allow BA professionals to close the gap [10]. Therefore, Business Analysts of Capgemini present you these research results for the BPI Challenge 2013.

1.1.3 IT reliability and efficiency

The case for 2013 in the BPI Challenge is based on information of an IT system which is part of the IT department of Volvo [11] responsible for problem and incident management in combination with a call center. Call centers are the de facto standard for efficiency studies [12] and their performance is highly reliant on supporting IT systems for providing applicable knowledge. Because the Volvo call center as subject of analysis is the problem solving unit for incidents, reliability and efficiency are applicable topics of research for this analysis.

1.2 Aim

There are multiple aims for this BPIC '13 research:

1. To position Process Mining in the collection of competences of Business Analysts in relation to Business Process Management
2. To position the research method characteristics in the context of Business Analysis
3. To provide proof that Process Mining is beneficial in a methodological approach of analysis in context of Business and IT gaps and the SEMBA method
4. To provide insight in how Business Analysis is applied and what next steps are with Process Mining outcomes.

The aims as presented are positioned by answering the questions as stated by Volvo [11], where we assume that the author is the requestor.

1.3 Added value

There are two sides for the added value of this paper. For professionals, this paper shows how Process Mining and facilitating tools can be applied and how the rationale is defined when handling complex customer cases. For Science, this paper allows to

relate the insight of the requirements of consulting professionals for non-standard expertise and how Process Mining is validated as an important method for Business Analysts and their profession.

1.3.1 Professionals – positioning in practice

The consulting profession is a field of expertise that is highly reliant on academic work and evidence driven. Customers grow in their insights and requirements and demand factual decision making solutions, from one-off decisions to continuous business management tooling.

Capgemini established the Structured Expert Method for Business Analysis (SEMBA) as method for Business and IT analysis. SEMBA consists of four phases (Focus & Direction, As-Is Analysis, To-Be Design and Migration Design) and multiple streams (Business Context, Business Process, Information, Application Landscape and Requirements Engineering), depicted in figure 1. SEMBA is established with predefined deliverables, which allow for a consistent, predictable outcome of complex analysis. The method is a standard, however, the content and interpretations are customer tailored. The combination of evidence driven tools with a standardized methodology of analysis resulting in predictable delivery is developed for client satisfaction [13].

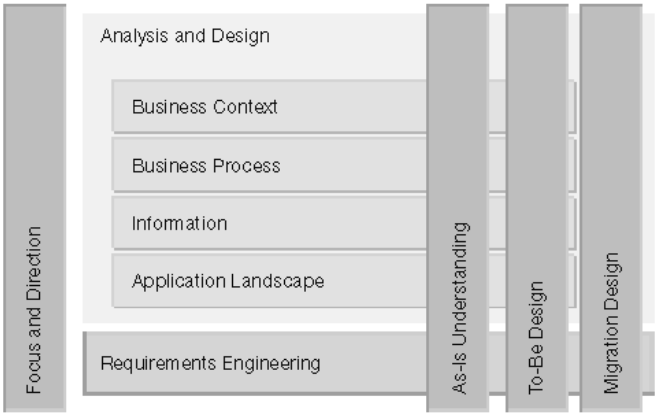

Fig. 1. The Structured Expert Method for Business Analysis (SEMBA)

1.3.2 Science – positioning in literature

The BPI Challenge is a great way to present and combine science with application in practical settings. Where the BPI Challenge is a challenge for academics and professionals, this paper is a presentation of analysis and professional positioning for anyone interested in the field of process mining and business process management. As Business Analysts bridge the gap between Business and IT, this paper bridges the gap between science and business by applying research findings.

2 Research Design

2.1 The case positioned in SEMBA

For this analysis and the deliverable (this paper) the limitations are based on time restrictions and client contact. The time restriction is 18 hours and there cannot be client interaction because of the design of the BPI challenge.

2.1.1 Focus and Direction

In the focus and direction phase, there are seven steps followed. As SEMBA is a Capgemini proprietary approach to analysis, not all details are presented in this case. The common result is that the problem is defined as the combination of questions as stated in the Volvo case description in the context of the Volvo IT department related to incident and problem management with the use of a call center in multiple

countries. The client expects answers to the stated questions with the use of the provided input: datasets and descriptions of the dataset and the system where the dataset is obtained from.

The approach is described from paragraph 2.2, normally this phase describes the approach and scope of the activities. The outcome of this phase is a formal and exhaustive overview of what is to be done, who does it, how activities are done, when activities take place, where and why. The scope for this research has no objective to capture requirements, therefore the stream Requirements Engineering will be left out of this paper.

2.1.2 As-Is Analysis

For the BPI Challenge 2013, the As-Is situation is established for some of the streams.

Business Context

The business context is related to the IT department of Volvo. The unit of analysis is the functioning of the VINST system, its users and the registrations in the system across a limited timeframe. The VINST system is used globally by multiple support organizations.

Due to limitations in report size this item is kept condensed and refer to the VINST context description [11] and VINST user guide [XX] for more detail.

Business Processes

The business processes for analysis are related purely to the registration of activities within the VINST system. There are no satellite systems or procedures in scope. The higher hierarchy process could be captured under “Incident and Problem Management”. The classes of activities can be defined as Incident solving and Problem solving activities. Activities can be handled by first, second and third line support employees. Support employees have each a specific area of expertise related to technology.

Information

Information is stored and transformed in the VINST System. Information is related to the employees working for Volvo on a global scale, their position in the organizational hierarchy, their expertise, products and geographical position. Furthermore, information is assumed to be present in the VINST system which enables knowledge transfer for storing, retrieving and adding solutions to problems related to products and services.

Application Landscape

There is no formal description available other than the VINST system. No peripheral system, interface or other element is mentioned in the context other than an e-mail facility.

2.1.3 To-Be Design

The phase To-Be design is not applicable for this research. For applicable cases, the outcomes and decisions of the As-Is phase are used to create at least one To-Be design. The design elements can be prescreened to design only one most feasible solution, or multiple scenarios are considered. In case of multiple scenario's, the individual scenarios are scored using multiple optional activities such as Multi-Criteria Decision Analysis (MCDA) techniques covering related aspects or Simulation for business performance, for example. The outcome is a so called Gap Analysis which covers the difference between the As-Is situation and the To-Be design(s). The Gap Analysis covers each of the aforementioned streams: *Business context, Business Processes, Information* and *Application Landscape*.

Creation of To-Be designs can be accelerated by usage of reference models such as the Supply Chain Operations Reference (SCOR®) [17], Process Classification Framework by APQC [18], Frameworx [19] and the Banking Industry Architecture Network (BIAN) [20].

2.1.4 Migration Design

The phase "Migration Design" is not applicable for this research. For applicable cases, the outcome of the To-Be design phase is used to review the methods of how the As-Is situation can be migrated to the To-Be design. Common scenarios are i.e. Big-Bang, Pilot location, Linear migration and Exponential migration amongst others.

2.2 Research method

As discussed in chapter 2, the basic steps for Process Mining are followed as described by van der Aalst et al., to cover the exploratory element of this research. Then professional insights on what to analyze or ask the problem owner in a next activity to proceed towards a To-Be phase or suggestions for improvement.

2.2.1 Research design

Because of the characteristics of Process Mining mainly consisting of exploratory research, the limited interaction for research by the researchers, the data type being Quantitative and a setting which resembles a Laboratory, the research method is determined as Non-reactive research, as presented in table 1.

Method	Setting	Data Type	Researcher Role
Action Research	Field	Qualitative	Active
Case Study	Field	Qualitative	Passive
Experiment	Laboratory	Quantitative	Active
Non-reactive	Laboratory	Quantitative	Passive
Survey	Field	Quantitative	Passive

Table 1. Non-reactive research selected as research method based on multiple criteria

2.2.2 Exploratory research

Process mining for processes is mainly exploratory research [7]. First, the researcher needs to get a feeling for what the data represents. Second, assumptions and statements about the dataset need to be stated to test which part of the data is relevant for the desired answer. The two aforementioned elements are attained iteratively by doing small experiments and testing.

For the research on the stated (sub)question, three basic topics will be stated: Research scope, Filters used and the results of the research with optional elaboration for each of the topics.

Research scope

The research scope limits the unit of analysis to least possible number of attributes to consider with the relevant subset of the data. The research scope is limited through some elements: the dataset, the assumption, the method and a threshold. The dataset element shows which dataset is used. The assumption is the description of which assumption(s) would lead to the right subset of the data. The method element describes how the assumption is translated into the subset. The threshold is set for limiting the results as presented for this research.

Filters used

The filters used give a description on how the tool was set and which settings were set to obtain the subset results.

Results

The results show in either figure or table form the results using the aforementioned limitations, settings and scope.

2.2.3 Explanatory research

The explanatory element in the research is highly limited, due to no client interaction, no strategic and tactical information about the company and no baseline information about performance or access to operational teams and systems.

Possible explanations will be provided as suggested research topics based on previous commercial engagements of the researcher. These explanatory contents are

not tested to be applicable in the Volvo situation and should be tested with a proper hypothesis which is refined by client interaction.

2.3 Data usage

Provided data sets Incidents [14], Open Problems [15] and Closed Problems [16]. Sets are provided in the XES data format, however for this research the prepared combined dataset by Fluxicon is used for the Disco tool [22]. Because a non-primary source is used for the data, a comparison has been run between exports of the Fluxicon dataset and the provided XES dataset. There have not been found any inconsistencies.

2.3.1 Assumptions

On the topic of data, there are many possible issues resulting in an incomplete or sometimes unusable dataset. Because of the nature of the BPI Challenge and the available prepared data, the assumption is that the dataset is fit for research purposes.

2.4 Tooling

In this paragraph multiple toolsets for analysis of the process mining category are discussed. Four software candidates are discussed about features and applicability for use in this research case.

2.4.1 ProM 5

ProM [23] is the acronym which stands for Process Mining. The tool is open source and mainly aimed at researchers and scientific application. It is a collection of custom written plugins for various insights that can be obtained from datasets. Version 5 is the last version that has a certain interface which is more complex but powerful for the experienced user.

2.4.2 ProM 6

ProM 6 [23] is a continuation of the ProM application which has been overhauled on the UI and activity design so analysis is more straightforward and entry-user friendly. The package is a platform which can be upgraded with multiple plug-ins for several types of analysis depending on the requirements of the user.

2.4.3 Fluxicon Disco

Fluxicon is the company which creates the commercial tool Disco for process mining analysis of datasets [22]. Disco is capable of delivering quick analysis results on desktop computers and is optimized for the areas of process discovery and a set of

statistical overviews. It has multiple options to filter data into subsets and quickly trim sets for specific analysis.

2.4.4 Perceptive Process Mining

Perceptive is the company which creates the product Perceptive Process Mining (PPM) for analysis of datasets [24]. PPM is capable to analyze datasets in both social network and process flow methods, using cloud technology to provide performance beyond desktop computers. The tool is powerful and feature rich but requires more experienced researchers to use the tool to its maximum effectiveness.

2.4.5 Tool selection

The tool(s) will be selected based on the availability of the tool, the user friendliness and timely analysis results whilst working with the tool.

Based on the access to the tool, Perceptive Process Mining is not used, it would require a license or accredited access to the tool for analysis. Due to time restrictions the researcher did not contact Perceptive to consider this opportunity.

Based on previous experiences with ProM 5 and 6, the applications are not used for this research.

The research tool for this paper is Disco by Fluxicon, the demo product with the prepared dataset made available.

2.5 Process discovery methodology

2.5.1 Social Network Analysis

Social network analysis is a representation of the dataset which uses the people or departments as the unit of analysis instead of the events. This allows for another dimension of outlier and deviant activity analysis.

2.5.2 Process Network Analysis after Process Discovery

Process Network analysis is the analysis of sequential events that form some sort of network based on the number of similar cases and flows of events. The flow of events is constructed using Process Discovery, in this research based the fuzzy mining technique. Some tools allow for automated generated models to be derived from datasets for further use. There is a limitation on the discovered processes in such forms, as events are the result of a process, not the process itself.

2.5.3 Methodology selection

Due to the restricted timeframe as discussed in paragraph 2.1 and the dismissal of ProM, Social Network analysis will not be applied for this research. Process Network

analysis will be applied with the notion that the discovered processes might not be the processes but the sequences of end-states per process step.

3 Results

3.1 Questions

In this paragraph, the questions are answered in the described methodology from chapter 2.

3.1.1 Q1.1 Push to front: For what Products is the push to front mechanism most used and where not?

Research scope: For what Products Push to Front is used

Element	Description
Dataset	Incidents
Assumption	Events have a specific sequence and the scope is limited to these events.
Method	Analyze the distribution of products
Threshold	All results with a relative percentage of <1% of the cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Attribute	Org:group	NOT ({A..Z}{1..99} 2nd OR {A..Z}{1..99} 3rd)
Endpoints	Activity – Mode discard cases	Start event values: All End event values: Completed / In Call

The resulting set has N=1854 cases

Results

Product	Relative Frequency	Absolute Frequency*	Cumulative Percentage**
PROD424	14,57%	270	14,57%
PROD660	11,00%	204	25,57%
PROD566	5,61%	104	31,18%
PROD494	5,16%	96	36,34%
PROD13	3,84%	71	40,18%

PROD453	3,16%	59	43,34%
PROD321	2,76%	51	46,10%
PROD544	2,06%	38	48,16%
PROD832	1,96%	36	50,12%
PROD253	1,80%	33	51,92%
PROD369	1,46%	27	53,38%
PROD104	1,39%	26	54,77%
PROD434	1,37%	25	56,14%
PROD363	1,27%	24	57,41%
PROD328	1,24%	23	58,65%
PROD423	1,20%	22	59,85%
PROD815	1,17%	22	61,02%
PROD698	1,00%	19	62,02%

* Absolute frequency is calculated using the N=X number and multiplied by the relative frequency. The resulting set (with disregard of the set threshold) can add up to another number than N=X.

** Cumulative percentage is obtained by adding up the various (rounded) results from the relative frequency. The resulting set (with disregard of the set threshold) can add up to another number than 100%.

Research scope: For what Products Push to Front is NOT used

The standard assumption would be that you would select the product list from the previous sub-question and would subtract it from the total list of products. However, when doing that, you will never discover whether there are duality issues with the same product(s). Therefore, the following research approach is followed:

Element	**Description**
Dataset	Incidents
Assumption	Events have a specific sequence and the scope is limited to these events.
Method	Analyze the distribution of products
Threshold	All results with a relative percentage of <1% of N cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	**Filter by:**	**Event values:**
Attribute	Org:group	{A..Z}{1..99} 2nd AND {A..Z}{1..99} 3rd
Endpoints	Activity – Mode discard cases	Start event values: All End event values: Closed / {Cancelled, Closed, Resolved}

The resulting set has N=2198 cases

Results

Product	Relative Frequency	Absolute Frequency*	Cumulative Percentage**
PROD424	5,56%	122	5,56%
PROD542	4,86%	107	10,42%
PROD698	3,24%	71	13,66%
PROD607	3,00%	66	16,66%
PROD802	2,73%	60	19,39%
PROD805	2,14%	47	21,53%
PROD660	1,95%	43	23,48%
PROD604	1,84%	40	25,32%
PROD617	1,62%	36	26,94%
PROD243	1,37%	30	28,31%
PROD253	1,33%	29	29,64%
PROD544	1,23%	27	30,87%
PROD325	1,19%	26	32,06%
PROD631	1,18%	26	33,24%
PROD267	1,08%	24	34,32%
PROD337	1,06%	23	35,38%

* Absolute frequency is calculated using the N=X number and multiplied by the relative frequency. The resulting set (with disregard of the set threshold) can add up to another number than N=X.

** Cumulative percentage is obtained by adding up the various (rounded) results from the relative frequency. The resulting set (with disregard of the set threshold) can add up to another number than 100%.

Q1.2 Where in the organization is the push to front process most implemented?

Research scope: differences between organizations A2 and C

Element	Description
Dataset	Incidents
Assumption	Events have a specific sequence and the scope is limited to these events.
Method	Compare input and output by case numbers.

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:

Attribute	Organization involved	Org line A2 [set 1] Org line C [set 2]
Endpoints	Activity – Mode discard cases	Start event values: Accepted / In Progress OR Queued / Awaiting Assignment End event values: Completed / Closed OR Completed / In Call

Results

The most important organization elements are C and A2, as stated in the VINST data set description are confirmed by the data set with a distribution by organization of 67% of the cases handled by organization C and 17% of the cases handled by organization A2. The resulting breakdown of how many cases are receiving the status "Completed / In Call" show a difference (not calculated for significance) of 35,5% cases solved by organization C and 2,4% by organization A2.

Organization	**A2 (17%) [set 1]**		**C (67%) [set 2]**	
Input	Cases (N=)	%	Cases (N=)	%
Accepted / In Progress	N=595	46,20%	N=4445	87,80%
Queued / Awaiting Assignment	N=694	63,80%	N=619	12,20%
Total	N= 1289	100,00%	N=5064	100,00%
Output	Cases (N=)	%	Cases (N=)	%
Completed / Closed	N=1258	97,60%	N=3264	64,50%
Completed / In Call	N= 31	2,40%	N= 1800	35,50%
Total	N=1289	100,00%	N=5064	100,00%

3.2 Q1.3 What functions are most in line with the push to front process?

Research scope: the functions which solve first line support calls

Element	**Description**
Dataset	Incidents
Assumptions	All support desks are investigated (not only A2 and C) Status Completed / In Call is the correct end state
Method	Research on attribute org:role
Threshold	All results with a relative percentage of <1% of the cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Attribute	Organization involved	All
Endpoints	Activity – Mode discard cases	Start event values: All End event values: Completed / In Call

The resulting set has N=1882 cases

Results

Organizational Role (Function)	Percentage	Absolute Frequency*
V3_2	91,78%	1727
A2_1	6,24%	117
(unknown)	1,67%	31
A2_4	0,10%	2
E_6	0,10%	2
A2_2	0,10%	2

* Absolute frequency is calculated using the N=X number and multiplied by the relative frequency. The resulting set (with disregard of the set threshold) can add up to another number than N=X.

3.3 Q2: Ping pong behavior

2.1: What are the …

Research scope: responsibles for ping-pong behavior in incidents?

Element	Description
Dataset	Incidents
Assumptions	Cases with >8 events are ping-pong cases (based on sampling ping pong case flows 1-627819166 / 1-621825480 / 1-650013051) Cases cannot have the end-state Completed / In-Call
Method	Research on attributes functions / organizations / support teams / products
Threshold	All results with a relative percentage of <4% of the cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Performance	Number of events	Minimum number of events = 9 Maximum number of events = 124
Endpoints	Activity – Mode discard cases	Start event values: All End event values: NOT Completed / In Call

Results: functions

The resulting set has N=2601 cases.

Organizational Role (Function)	Percentage	Absolute Frequency
V3_2	43,01%	1119
A2_1	18,35%	477
(unknown)	10,61%	276
E_10	9,47%	246
A2_2	4,67%	121

Results: organizations

Organization involved	Percentage	Absolute Frequency
Org line C	61,46%	1599
Org line A2	20,57%	535
Org line B	8,18%	213

Results: support teams

Support teams	Percentage	Absolute Frequency
G97	15,96%	415
G96	5,82%	151

Results: products

Product	Percentage	Absolute Frequency

PROD424	15,36%	400
PROD660	4,23%	110
PROD542	4,01%	104

Research: responsibles for ping-pong behavior in closed problems?

Element	Description
Dataset	Closed Problems
Assumptions	Cases with >8 events are ping-pong cases (based on sampling ping pong case flows 1-736351127 / 1-653989471 / 1-563477371) Mean time is >=23d (1st block is <23 days and averaging 7d which is non-exceptional)
Method	Research on attributes functions / organizations / support teams / products
Threshold	All results with a relative percentage of <4% of the cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Performance	Number of events	Minimum number of events = 9 Maximum number of events = 36
Performance	Case duration	Minimum duration = 23 d Maximum duration = 6 years, 87 days

The resulting dataset has N=114 cases

Results: functions

Organizational Role (Function)	Percentage	Absolute Frequency
(unknown)	20,48%	23
A2_2	14,07%	16
E_10	13,09%	15
C_6	12,10%	14
E_8	6,90%	8

Applying another filter to the dataset of N=114 cases where the Function (unknown) has been selected results in a subset of N=43 cases.

The characteristics of these cases are (with the <4% threshold still applied)

Group		**Org. Country**		**Org. Involved**		**Product**	
Value	%	Value	%	Value	%	Value	%
Org line G3	79,73	Us	77,66	G199 3rd	79,73	PROD97	37,80
Org line G4	14,78	Se	19,59	G51 2nd	5,50	PROD98	27,84
				S30 2nd	4,81	PROD96	14,09

Results: organizations

Organization involved	**Percentage**	**Absolute Frequency**
Org line C	49,19%	56
Org line A2	25,26%	29
Org line G3	16,33%	19
Org line B	5,21%	6

Results: support teams

Support team	**Percentage**	**Absolute Frequency**
G199 3rd	16,33%	19
G21 2nd	7,18%	9

Results: products

Products	**Percentage**	**Absolute Frequency**
PROD97	11,26%	13
PROD98	8,02%	9
PROD802	5,91%	7
PROD96	4,15%	5

Open problems are not reviewed due to their incomplete state. However, the dataset can be handled in the same manner as presented in the previous two exhibitions to obtain the results.

3.4 Q3 Wait User

3.4.1 Q3.1: Who is making most use of the state Wait / User?

Research: Most use of state Wait / User

Element	Description
Dataset	Incidents
Assumptions	A subsequence is present which is Accepted / In Progress followed in time by Accepted / Wait User
Method	Research on attributes impact / support teams / products
Threshold	All results with a relative percentage of <4% of the cases is not represented

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Follower	Activity Reference event must be eventually followed by	Reference event value: Accepted / In Progress Follower event values: Accepted / Wait User

The resulting dataset has N=2485 cases

Results: Who is making most use of this status?

The answer from the selection and breakdown to the resource that uses this function the most provided only because we assume that the name is related to a system name: Siebel. This is furthermore assumed as the status changes produced by this user are mainly at a specific time (01:19 – 01:22) each day, which is assumed to be an automated script.

Because of limitations of sharing personal details based on legal protection of employees and their performance assessment [21] we will not publicize the names of individual employees making use of the Accepted / Wait User status.

Results: Impact

Impact	Major	High	Medium	Low
Cases	0	83	1216	1186

3.4.2 Q3.2: What is the behavior per

A: Support team
B: Function
C: Organization

No results have been produced because this part is best researched using Social Network analysis. As described in paragraph 2.5.3, Social Network analysis is not included in this research.

3.4.3 Q3.3: Is there overuse of the Wait / User state by location?

Research: overuse of the Wait / User state

Element	Description
Dataset	Incidents
Assumptions	A subsequence is present which is Accepted / In Progress followed in time by Accepted / Wait User
Method	Compare the wait / user resulting country breakdown to the general country breakdown of cases

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Follower	Activity Reference event must be eventually followed by	Reference event value: Accepted / In Progress Follower event values: Accepted / Wait User

The resulting dataset for Wait / User has N=2485 cases, the total dataset has N=7554 cases

The results for the breakdown of the Wait / User state by location is presented in table x.

Country	Wait / User	Total	Distance	If > 1,2647 = Yes
se	0,2852	0,3214	0,8874	No
pl	0,2649	0,2341	1,1316	No
in	0,1484	0,1047	1,4174	Yes
be	0,0966	0,0907	1,0650	No
us	0,0648	0,0774	0,8372	No
fr	0,0406	0,0482	0,8423	No

br	0,0316	0,0406	0,7783	No
nl	0,0221	0,0198	1,1162	No
cn	0,0178	0,0181	0,9834	No
kr	0,0098	0,0088	1,1136	No
gb	0,0029	0,0041	0,7073	No
ca	0,0029	0,0061	0,4754	No
SE	0,0028	0,0083	0,3373	No
ru	0,0024	0,0053	0,4528	No
jp	0,0023	0,0023	1,0000	No
de	0,0016	0,0008	2,0000	Yes
au	0,0014	0,0029	0,4828	No
my	0,0009	0,0019	0,4737	No
0	0,0009	0,0037	0,2432	No
th	0,0003	0,0004	0,7500	No

There is a rudimentary evaluation applied with respect of the 'distance' between the percentage of the handled incidents by country versus the usage of the Wait / User state. Distance is calculated by dividing the Total incident percentage by the Use Wait/User percentage. A threshold of 1,26 (126%) is used to distinguish if further analysis into the Wait / User state use for that particular country would be useful. The threshold is based on the average of the Distance (0,8548) added with one Standard Deviation (0,4099).

The resulting countries for further analysis are "in" and "de".

3.5 Q4 Process Conformity per Organization

Research: Do organization A2 and C work in the same way?

Element	**Description**
Datasets	Incidents & Closed Problems
Assumptions	Test this for the top product with issues, as this will be the largest impact if optimized. Only successfully closed cases are compared. This allows for standard flow to be expected.
Method	Compare the process flows of A2 and C. Research on attributes organization AND product AND endpoint
Threshold	No threshold is applicable in process flow analysis

Filters used on attributes of the dataset (in sequence)

Filter name	Filter by:	Event values:
Attribute	Organization	A2 | C
Attribute	Product	PROD424
Endpoint	Filter by activity, discard cases	Start event values: All End event values: Completed / Closed

The resulting dataset has N=7 cases for A2 and N=448 for C

Results:

These resulting sets did not give enough information to be conclusive if both organizations have similar operations, because the spread in the case variants was huge, as well as a too small subset for Organization A2. The attribute Product was removed from the filterset, resulting in other datasets: N=1258 for Organization A and N=3274 for Organization C.

Results:

The short answer to the question is no; the organizations do not work in the same way. When comparing the process flows derived from the resulting datasets, settings for Activities at 100% and Paths for 0%, a direct difference is visible:

Organization A2 (figure 2) starts recording cases using the status Queued / Awaiting assignment, whereas Organization C (figure 3) starts recording cases with the status Accepted / In Progress. Furthermore, the event Wait / User is used more frequently in Organization C3. This is a different way of working.

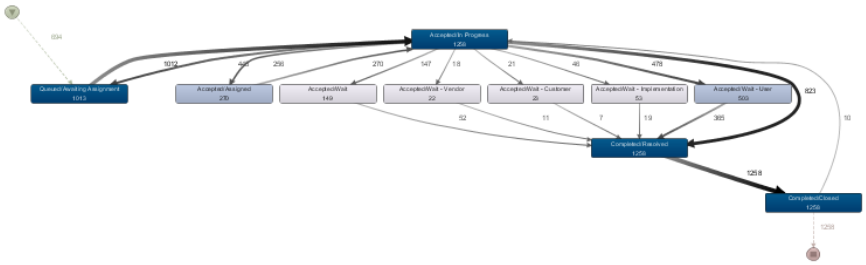

Fig. 2. Process flow for organization A2

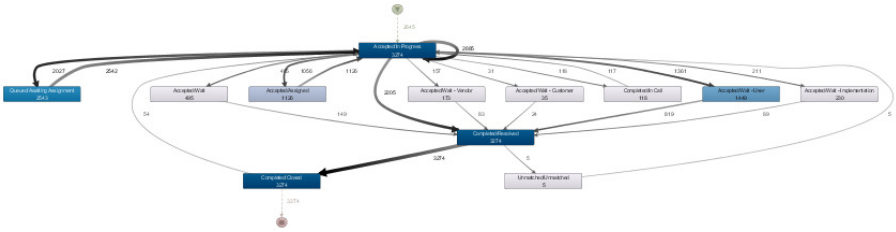

Fig. 3. Process flow for organization C

4 Conclusions & Recommendations

This chapter discusses the conclusions and recommendations, about the research, the aims for this research paper and the limitations about the results.

4.1 Conclusions

The data provided in this challenge was prepared well, but not yet cleaned, as some result sets show empty fields. The conclusion is that in such cases a choice should be made such as rework should be done to augment the dataset to become most effective, or strip the inconclusive cases for example. However, the researcher will best use the most pure dataset and remark deviations found.

The questions as asked by the problem owner are answered, except for the social network analysis question, behavior in various ways. From the results of the various questions, there can be concluded that there is not ‘one’ way of working around the globe whilst using the VINST system. Furthermore, within a product line there are differences as concluded from the most and least use of the push to front mechanism.

There are manu loops found in the dataset, of unknown cause. Analysis on the timespans shows that there are short periods of time between looping steps in most cases.

The SEMBA approach can benefit from Process Mining in creating an unbiased insight in how processes are used within a company. The Process Mining technique is therefore a valuable competence for Business Analysts, to complement basic modeling based on anecdotal or more formal process registrations.

Process Mining can be used to do exploratory and in lesser fashion explanatory research. In the situation that there is no interaction, it is not possible to do explanatory research. Process Mining is highly related to practical use, based on the source of the data. However, the interpretation of the data owner and data manipulating operators is required to make founded statements and conclusive hypothesis testing feasible.

4.2 Recommendations

These recommendations stem from the insights and musings coming from the research. They show some insight in where consultants would go next and where further research is recommended.

The first recommendation is to get access to the problem owner and system operators in order to provide meaningful answers. The feeling that arose during this research is that there might be questions behind these questions of higher importance. For example, issues with outsourcing could be driving the questions, or Service Level Agreement performance.

The second recommendation is to inquire about the procedural approach between the different organizations. The loops as presented in the data are a typical symptom of a missing event or misuse by employees of a status. The other approach would be to strip the dataset of the first iteration of “Accepted / In Progress” to make better sense of the volumes of cases and events.

The third recommendation is about the data itself. In some cases fields did not contain content or double entries or otherwise inconsistent content. This could be user errors, however a choice should be made upon including or excluding such cases.

The fourth recommendation is about the system automated actions. During the research multiple mutations were found which were executed by the user Siebel which is assumed to be the system. There are many events where it is unclear if the system ever should be able to set a case to such a state, i.e. “Waiting / User”.

4.2.1 Practical Implications and Limitations

The question behind the question.

Improving outsourcing? Contract issues? Performance improvement? As mentioned at the first recommendation, the question behind the question is something more valueable to be answered. Using the Process Mining technique researchers and consultants get the opportunity to spot ‘the elephant in the room’ and work towards a situation which has high client benefit.

What research can do.
Research in the field of Process Mining is still in high flux. The role of such a developing field is that it gives guidelines on how to proceed and discover what works and what not. The author would like to thank and encourage anyone pushing forward to set the boundaries, however (in)feasible the results.

What research cannot do.
Provide specific answers, sometimes. This is exactly why Business Analysis is up and coming as a profession. Business Analysis allows for business improvement and setting up the requirements for what the desired state would be. But moving towards a To-Be design is always limited by time, money or quality, so there will never be a best fitting solution. Specific and a high fit between Business demand and IT delivery are only possible when the requirements are identified, assessed and put into context. Then the result will be realistic on the elements of time, budget and required quality. This realism allows for high customer satisfaction with limited resources.

About the author:

Jef Martens is a Business Analyst for Capgemini and Business Process Management consultant. He is the curator of reference material for Capgemini on an international scale.

References

1. Luftman, J., Ben-Zvi, T.: Key Issues for IT executives 2010: Judicious IT Investments Continue Post-Recession. MIS Quarterly Executive, Vol. 9, No.4, pp. 263-273 (2010)
2. Luftman, J., Ben-Zvi, T.: Key Issues for IT executives 2009, MIS Quarterly Executive, Vol. 9, No. 1, pp. 49-59. (2010)
3. Luftman, J., Kempaiah, R., Rigoni, E.H.: Key Issues for IT Executives 2008, MIS Quarterly Executive, Vol. 8, No. 3. (2009)
4. Luftman, J, Kempaiah, R.: Key issues for IT executives 2007. MIS Quarterly Executive Vol. 7, No. 2 pp. 151-159 (2008)
5. Luftman, J.: Key Issues for IT Executives 2005, MIS Quarterly Executive Vol. 4 No.2, pp. 269-286, (2006)
6. Van Dongen, B. et al.: Business Processing Intelligence Challenge (BPIC) – Third International Business Process Intelligence Challenge http://www.win.tue.nl/bpi2013/doku.php?id=challenge (2013)
7. van der Aalst, W. M. P.: *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin (2011)
8. Bonakdar, Amir, et al.: Transformative Influence of Business Processes on the Business Model: Classifying the State of the Practice in the Software Industry, System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE (2013)
9. Blum, K., Landkroon, D., Tjon Tjauw Liem, J.: Informatiemanagersurvey 2010-2011 – Are you in the driver's seat of your information vehicle? Whitepaper by Capgemini Nederland B.V. (2012) http://www.nl.capgemini.com/sites/default/files/resource/pdf/Informatie_Managers_Survey_0.pdf
10. Schreiner, K.: The Bridge and Beyond: Business Analysis Extends Its Role and Reach. IT Professional vol.9,iss.6, pp. 50-54. IEEE Computer Society (2007) http://dx.doi.org/10.1109/MITP.2007.122
11. Steeman, W., Volvo IT: VINST Data Set - VINST information needed to understand the dataset, pp. 1-12 (2012)
12. Batt, R., Moynihan, L.: The viability of alternative call centre production models. Human Resource Management Journal, Volume 12, Issue 4, pages 14–34 (2002) DOI: 10.1111/j.1748-8583.2002.tb00075.x
13. Capgemini Nederland B.V.: SEMBA Structured Expert Method for Business Analysis, Whitepaper by Capgemini Nederland B.V. pp. 1-4, (2011) http://www.capgemini.com/sites/default/files/resource/pdf/SEMBA_Structured_Expert_Method_for_Business_Analysis.pdf
14. Steeman, W.: BPI Challenge 2013, incidents. Ghent University. Dataset. http://dx.doi.org/10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee (2013)
15. Steeman, W.: BPI Challenge 2013, open problems. Ghent University. Dataset. http://dx.doi.org/10.4121/uuid:3537c19d-6c64-4b1d-815d-915ab0e479da (2013)
16. Steeman, W.: BPI Challenge 2013, closed problems. Ghent University. Dataset. http://dx.doi.org/10.4121/uuid:c2c3b154-ab26-4b31-a0e8-8f2350ddac11 (2013)
17. Supply Chain Council, What is SCOR? Retrieved 10-07-2013 https://supply-chain.org/scor
18. APQC, Process Classification Framework, Accessed 10-07-2013, http://www.apqc.org/process-classification-framework
19. TM Forum, Frameworx, eTOM and other Frameworks, Accesed 10-07-2013, http://www.tmforum.org/TMForumFrameworx/1911/home.html
20. Banking Industry Architecture Network, About BIAN, Accessed 10-07-2013 http://bian.org/about-bian/

21. van den Brand, P.: Training Perceptive Process Mining, inhouse for Capgemini and Capgemini Consulting, Training date 26-04-2012, Utrecht
22. Fluxicon, Fluxicon Blog Entry, BPI Challenge 2013, Accessed 02-06-2013 https://fluxicon.com/blog/2013/06/bpi-challenge-2013/
23. Process Mining group, Accessed 10-06-2013 http://processmining.org
24. Perceptive Software: Perceptive Process Mining, Accessed 10-06-2013 http://www.perceptivesoftware.com/products/perceptive-process/process-mining