Process Mining and the ProM Framework: An Exploratory Survey - Extended report

Jan Claes and Geert Poels

Department of Management Information Science and Operations Management,
Faculty of Economics and Business Administration
Ghent University, Belgium
{jan.claes, geert.poels}@ugent.be

1   Introduction

We conducted a survey to get insights in the process mining community's opinion about the process mining field and tools. The survey was held by sending out a questionnaire with 5 questions about process mining, 5 questions about the ProM framework and 5 questions about the respondent[1]. The questionnaire was put online at 2012-3-18 and was closed at 2012-5-1. In total 336 different browser sessions were registered of which 90 respondents completed all questions.

At 2012-4-7 we added three additional questions for which we counted 48 respondents. These extra questions asked about the opinion on the most used ProM plug-ins (at 2012-4-5 we calculated next formula for every plug-in and selected only the plug-ins with a value equal to or larger than 10 as 'most used ProM plug-in').

$$5 \times FU + 2 \times OU + 1 \times TO$$

(FU being the number of respondents that indicated 'frequent use' of the plug-in, OU stands for 'occasional use' and TO stands for 'tried it once').

Because all questions were optional, we had a different number of respondents for each question. The number of respondents for each question can be derived from Fig. 1. The minimum number of respondents was 28 (question 10), the maximum was 119 (question 2) and the average was 97 respondents. We approached possible respondents by mail and by social media (LinkedIn and Twitter). Fig. 2 shows the number of respondents per day.

---

[1] see http://processmining.ugent.be/survey.php

The questions were not derived from theory; but were selected based on their perceived relevance. In this report we discuss answers and hope to add value by not only presenting the results, but also adding our comments to the observed characteristics.
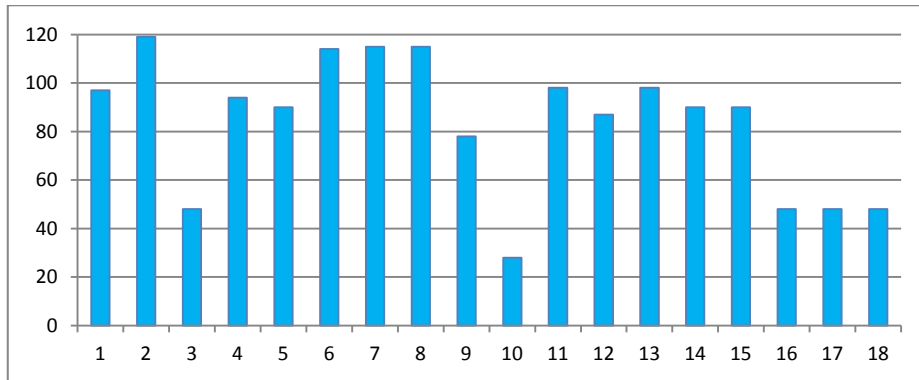

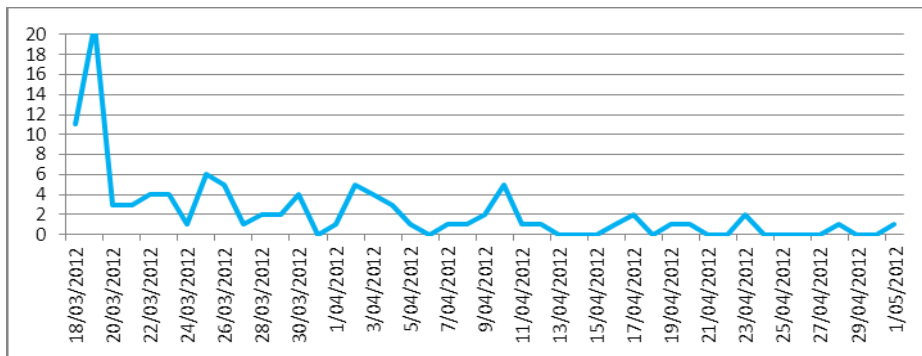**Fig. 1.** Respondents per question

Fig. 2. Respondents per day

2   Results

This section lists the results of the survey. For your convenience we tried to present the data in an easy form. The raw data set can be downloaded from http://www.janclaes.info/papers/PMSurvey.

2.1   Used source to introduce process mining

Question 1.   If you need to introduce process mining to other people: what article, blog, website, ... would you refer to? (97 respondents)

Many researchers, but also many consultants, do not find it easy to convince people of the possible benefits of process mining (see Section 2.5: "hard to sell"). Therefore we were

interested in which sources are used to introduce the field to other people. Next to the official ProM website, Process Mining Manifesto and Process Mining Book, the blog from Fluxicon seems to be a popular source.

Table 1. Sources to introduce process mining

http://www.processmining.org	42
"Process Mining Manifesto"	24
Process Mining Book	17
http://www.fluxicon.com/blog	15
ProM website	8
Academic articles	4
Own documentation	4
Process mining: a research agenda	3
"Business process mining: An industrial application"	2
http://www.promtools.org	2
http://www.wikipedia.org	2
"Workflow Mining - Discovering process models from event logs"	2
"Workflow mining: A survey of issues and approaches"	2
A Process Case study ("Process Mining - Ana Aeroportos de Portugal")[2]	1
Blogpost: "A Process Mining Project"[3]	1
Blog: Gartner, Keith Swenson, Sandy Kemsley	1
"Discovering Concurrency; Learning (Business) Process Models from Examples"	1
http://www.google.com	1
http://www.logistiek.nl[4]	1
http://www.managementsite.nl[5]	1
http://www.process-intelligence.com[6]	1
http://wwwis.win.tue.nl/~wvdaalst	1
IEEE Task Force on PM	1
http://www.linkedin.com	1
http://perceptivesotware.com	1
"Process Diagnostics: a Method Based on Process Mining"	1
"Process Mining – bring speed to your improvement efforts"[7]	1
Process Mining in one minute - promo film produced by Pallas Athena	1
Process mining tv	1
http://www.processmining.com	1
http://processmining.ugent.be	1
ProM forum	1
ProM Tool	1
ProM tutorial	1
QPR product papers	1
Van der Aalst	1
http://www.processgold.de	1
http://www.qpr.com	1
XESame master thesis	1

[2] http://www.bptrends.com/publicationfiles/BPTrends%20Article%20Process%20Mining%20-%20Ana%20Aeroportos%20de%20Portugal-Alberto%20Manuel%20%20V01-Final.pdf

[3] http://ultrabpm.wordpress.com/2011/12/19/a-process-mining-project

[4] http://www.logistiek.nl/experts/id12244-Process_Mining_sneller_en_beter_inzicht_in_logistieke_proces

[5] http://www.managementsite.nl/4528/ict-internet/process-mining-snel-inzicht-bedrijfsprocessen.html

[6] http://www.process-intelligence.com/en/Process-Analysis/176644.html

[7] http://ultrabpm.wordpress.com/2011/09/2

2.2   Process mining tools

Question 2.   Which process mining tools do you use/know? (119 respondents)

The most popular process mining tool is definitely ProM. Also the tools to prepare files for ProM are very popular (ProM import, Nitro and XESame). The at the moment of the survey unreleased tool Disco completes the top 5.

Table 2. Process mining tools

	Frequent use	Occasional use	Tried it once	Didn't use but heard about it	Never heard about it
ARIS Process Performance Manager (Software AG)	1	3	12	51	64
BPMOne (Pallas Athena)	3	8	12	51	57
Disco (Fluxicon)	10	6	9	29	77
Futura Reflect (Futura Technology)	4	5	11	35	76
Interstage Automated Process Discovery (Fujitsu)	0	0	3	37	91
Nitro (Fluxicon)	11	21	23	24	52
ProM (Academic)	55	33	9	6	28
ProM Import (Academic)	18	34	23	17	39
QPR ProcessAnalyzer/Analysis (QPR Software)	2	1	3	27	98
XESame (Academic)	6	12	13	29	71
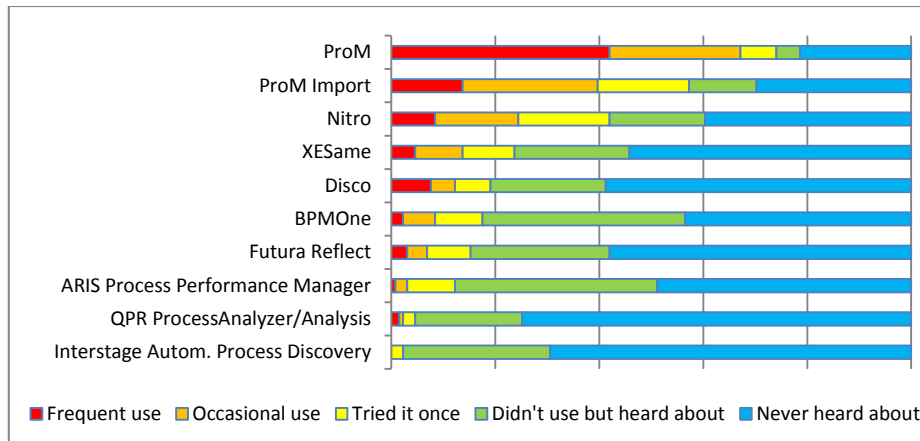
Fig. 3. Process mining tools

2.3   Used metaphor to describe (part of) process mining

Question 3. If you would use a metaphor to describe (a certain part of) process mining, what metaphor would you use? For example a black box on a plane is sometimes used as a metaphor for an event log. (48 respondents)

Many people use metaphors to describe (a certain part of) process mining. We were curious which metaphors are popular. Taking an x-ray of the organization, visualize a process using techniques of geographical maps, sifting sand (data) to find gold (valuable information on processes) form the top 3 answers to this question.

Table 3. Used metaphor to describe (part of) process mining

X-ray	10
Map	7
Sift sand for gold	6
A peephole in the black box	3
GPS	3
Desire lines	2
Camera/picture	1
Caved paths	1
Conformance checking: playing *railroad* and see whether your train jumps the rails	1
Diagrams beat Tables	1
Find one string in spaghetti	1
Find forest between trees	1
How heavy is the sack	1
Judge (compliance)	1
Lights on versus flash	1
Medical exams	1
Mining of diamonds	1
Painting room (lots of preparation)	1
Protein synthesis	1
Reflection	1
Walking on a factory floor	1

2.4   Benefits of process mining

Question 4. What are, according to you, the main benefits of process mining or reasons to use process mining techniques? (94 respondents)

The main perceived benefit of process mining is its objectivity (the use of real process data assures a certain degree of objectivity of analyses). Another benefit is speed. Once you have the right data in place, most process mining techniques manage to get fast results. A popular application that is named as a benefit of process mining, is the possibility to find exceptions and check conformance. Not only process mining is praised for finding errors, but it also helps in identifying the causes for certain deviations. The possibility to take different views on the same process or data is also much appreciated.
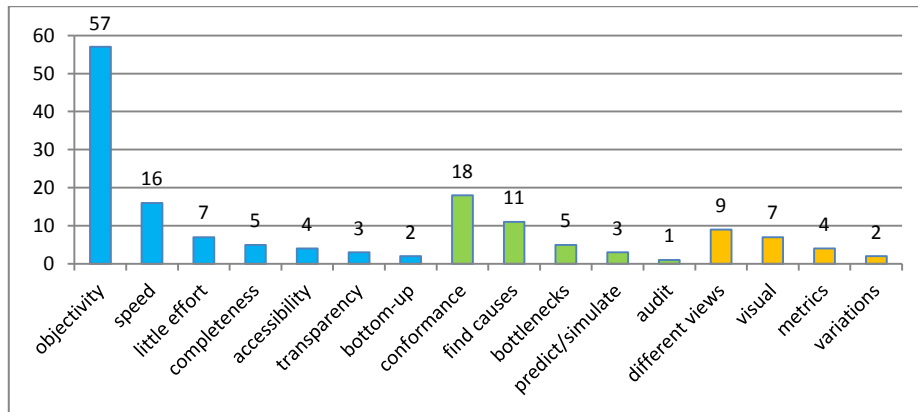
Fig. 4. Benefits of process mining
(blue: characteristic, green: application, orange: representation)

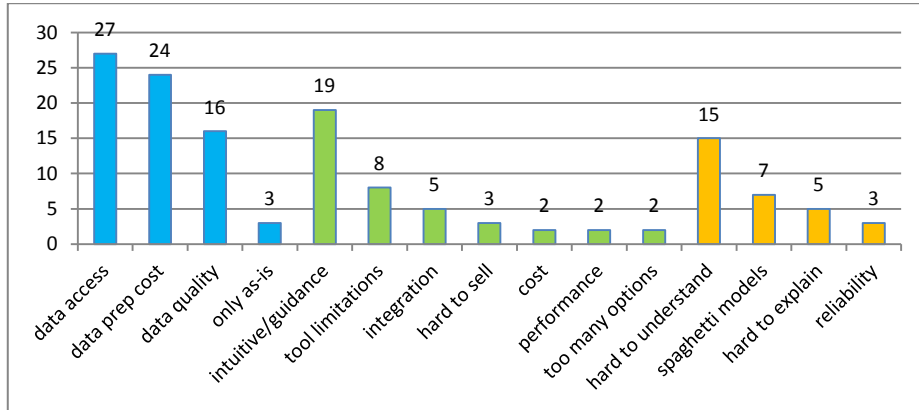## 2.5   Disadvantages of process mining

Question 5.   What are, according to you, the main disadvantages of process mining or reasons to not use process mining techniques? (90 respondents)

The big problem with process mining seems to be the search for data. It is hard to find the right data of the right quality and to fit it in the right structure. Efforts have to be made by both research and practice to improve the access to proper data. Recently the IEEE Task Force on Process Mining[8] developed the event log open standard xes[9]. Now it is important that the community tries to persuade software developers to use this format for logging. That would help solving the problem of the data issues in process mining projects.

Another drawback of current techniques is a lack of documentation and intuitiveness. Most process mining techniques are very complex, but to be really useful this complexity should be hidden behind simple and beautiful user interfaces with guidance for the inexperienced users. Should it be difficult to hide certain parameters or to assist a user in its choice, at least some clear documentation should be available. Whether this is the task of researchers or (commercial) developers is unclear. In our opinion, it is at least an opportunity for commercial organizations to make profit by making the techniques and tools available and useful for a larger audience.

---

[8] See http://www.win.tue.nl/ieeetfpm
[9] See www.xes-standard.org/

Fig. 5. Disadvantages of process mining
(blue: input, green: techniques, orange: output)

2.6 Used versions of ProM

Question 6. Which versions of ProM did you ever use? (114 respondents)

It seems logical that older versions of ProM are less used. Still we observe that the previous version ProM 5 is more used than the newer ProM 6. We think the reason is that ProM 6 is a totally redeveloped tool and because many existing and trusted techniques that have proven their value are still not converted to the new ProM 6, many users choose for stability and stick to ProM 5. The message is clear: If ProM 6 wants to be seen as the successor of ProM 5, developers should convert the popular plug-ins in a stable manner. But this seems easier than it is: because many developers are researchers and their research might have shifted focus, it is possible they cannot be convinced to give priority to this conversion operation (if they are still interested in (this type of) research at all).

Table 4. Used versions of ProM

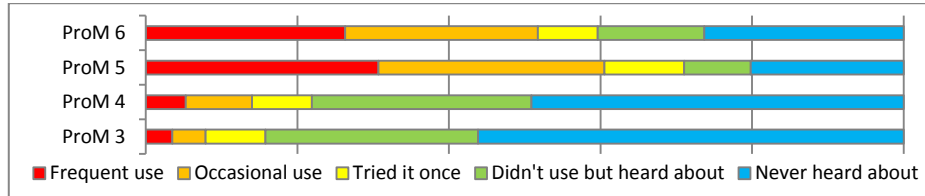|	Frequent use	Occasional use	Tried it once	Didn't use but heard about it	Never heard about it
ProM 3	4	5	9	31	64
ProM 4	6	10	9	32	56
ProM 5	35	34	11	10	23
ProM 6	30	29	8	16	30

Fig. 6. Used versions of ProM

2.7   ProM 5 plug-ins

Question 7. Which ProM 5.2 plugins do you use (select up to 10 plugins)? (115 respondents)

Question 16. Please indicate for each mentioned ProM 5.2 miner plugin if

Question 17. Please indicate for each mentioned ProM 5.2 analysis plugin if
  you ever used this plugin: used
  For the plugins you ever used, then you can also indicate if
   you think the use is intuitive (enough): intuitive
   you understand the results: understand
   you trust the results (enough): trust
   you find it is fast (enough): fast
   you always use the default settings: default
  (48 respondents)

The big question for many ProM users is which plug-in to use. In a first attempt to shed some light on this problem, we did not try to theoretically derive which plug-ins give the most correct results, but we just asked the community which plug-ins they used and how these plug-ins are rated. We are aware that the answer to which plug-ins are most popular might not point to the proper plug-ins to use, but it might provide some direction (if a plug-in is used by lots of people, it might be a good first choice if one has no other clue).

For the most popular plug-ins we also asked respondents to indicate if they find them intuitive to use, if they understand the results, if they trust the results, if they like the speed and if they mainly use default settings or not. As a main conclusion we wonder why the indicated most used plug-ins are not the ones that score best on our improvised quality metrics. We marked every cell in Table 6 were less than 75% of the respondents agreed to place them into the indicated category.

Table 5. ProM 5.2 plug-ins: use

	Frequent use	Occasional use	Tried it once
Fuzzy Miner	26	16	5
Heuristics miner	22	14	7

Social network miner	14	9	2
Dotted Chart analysis	11	8	6
Alpha algorithm plugin	10	12	6
LTL Checker	10	9	2
Log Summary	8	2	
Alpha++ algorithm plugin	7	7	2
Geneteic algorithm plugin	7	9	3
Basic Performance Analysis	6	9	6
Organizational Miner	6	2	2
Performance Sequence Diagram Analysis	5	5	1
Transition System Generator	3	1	
SCIFF Checker Plugin	3		2
Originator by Task Matrix	3	2	1
Advanced Dotted Chart Analysis	3	2	1
Basic Log Statistics	2	7	1
Sequence Clustering	2	1	
Semantic LTL Checker	2		
Region miner	1	1	
Tsinghua-alpha algorithm plugin	1		
Performance Metrics in Ontologies	1		1
Parikh Language-based Region miner	1	1	
Pattern analyzer	1	1	2
Control Flow Benchmark	1	1	2
Trace Clustering	1	4	2
Prediction Miner	1	1	
Workflow patterns miner	1	1	1
Semantic Organizational Miner	1		
Multi-phase Macro Plugin	1		
Explicit Model Miner	1		
Case data visualization	1		1
Log Clustering		1	
DecMinerPlugin		3	
Role Hierarchy Miner		4	
FSM Miner		2	2
Log Splitting		1	
Association Rule Miner		1	3
Enhance Log With History			1
Frequency abstraction miner			2
Process Instance Inspector		2	1
Activity Clustering Miner		1	2

Table 6. ProM 5.2 plug-ins: evaluation

|	default settings	fast	trust	understand	intuitive
Genetic algorithm plugin	63%	83%	79%	17%	58%
Trace Clustering	50%	60%	70%	90%	40%
Sequence Clustering	60%	60%	60%	80%	60%
SCIFF Checker Plugin	50%	83%	83%	83%	50%
Transition System Generator	50%	83%	100%	83%	50%
Semantic LTL Checker	40%	100%	100%	100%	40%
Fuzzy Miner	64%	95%	79%	90%	56%
Perf. Sequence Diagram Analysis	62%	86%	86%	86%	67%
Dotted Chart analysis	72%	88%	80%	88%	60%
LTL Checker	59%	86%	95%	91%	59%
Advanced Dotted Chart Analysis	83%	92%	92%	75%	50%
Alpha++ algorithm plugin	56%	94%	89%	83%	78%
Alpha algorithm plugin	65%	90%	74%	94%	81%
Heuristics miner	67%	89%	89%	97%	69%
Organizational Miner	73%	82%	82%	95%	82%
Social network miner	72%	96%	88%	88%	76%
Originator by Task Matrix	78%	94%	94%	94%	78%
Basic Performance Analysis	90%	100%	95%	90%	76%
Log Summary	86%	100%	100%	100%	95%
Basic Log Statistics	87%	100%	100%	100%	96%
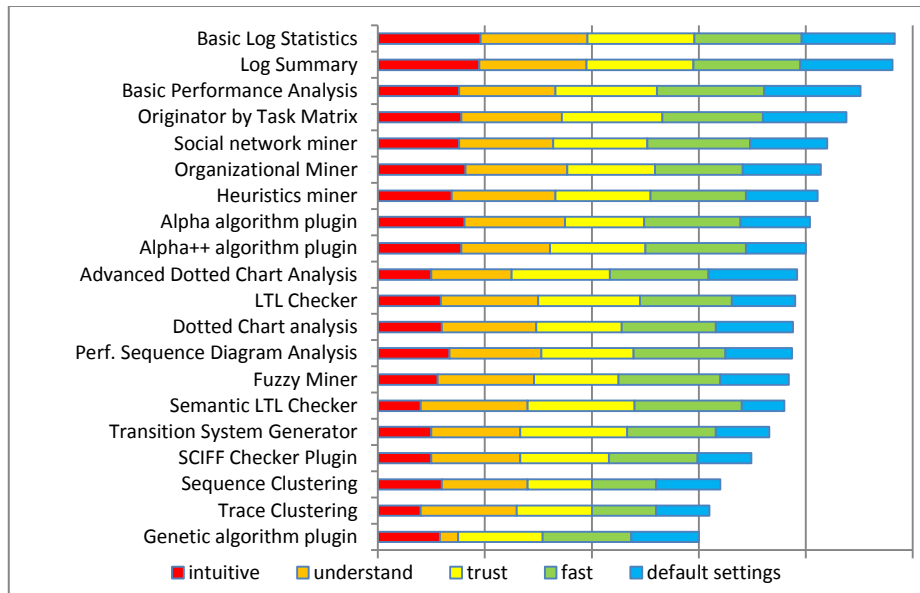
Fig. 7. ProM 5.2 plug-ins: evaluation

2.8   ProM 6 plug-ins

Question 8. Which ProM 6.1 plugins do you use (select up to 10 plugins) (115 respondents)

Question 18. Please indicate for each mentioned ProM 6.1 plugin if
> you ever used this plugin: used
>
> For the plugins you ever used, then you can also indicate if
>> you think the use is intuitive (enough): intuitive
>> you understand the results: understand
>> you trust the results (enough): trust
>> you find it is fast (enough): fast
>> you always use the default settings: default
>
> (48 respondents)

Similar to the ProM 5.2 plug-ins, we asked for the ProM 6.1 plug-ins which ones respondents use and how they are rated. We observe equivalent plug-ins are popular, although some of the most used plug-ins of ProM 5.2 are not listed as popular ProM 6.1 plug-in: e.g., Basic Performance Analysis (simply because it does not exist in ProM 6.1).

Table 7. ProM 6.1 plug-ins: use

	Frequent use	*Occasional use*	*Tried it once*
Heuristics Miner	14	7	5
Mine for a Fuzzy Model	11	5	5
Analyze using Dotted Chart	6	4	2
Replay a Log on Petri Net for Conformance Analysis	5	1	
Animate Event Log in Fuzzy instance	4	2	
Add Artificial Events	4	1	1
Mine for a Handover-of-Work Social Network	3	4	
LTL Checker Default	3	1	
Trace Alignment (with Guide Tree)	3	2	2
Filter Log Using Simple Heuristics	2	2	
LTL Checker	2	7	
Mine for a Petri Net using Alpha-algorithm	2	8	1
Mine Transition System	2	1	
Flexible Heuristics Miner	2	5	3
Analyze Transition System	2	1	
Mine for a Working-Together Social Network	2	2	
Mine for a Petri Net using ILP	1	2	1
Check Conformance using ETConformance	1	2	
Analyze Petri Net using LoLA	1		1
Convert Petri Net to Causal Net	1		

Convert LTL To Automaton	1		1
Check CTL* Model on Petri Net	1	1	
Replay a Completed Case on Petri Net for Conformance Analysis	1		
Mine Heuristic Net using Genetic Miner	1	1	1
Convert Log to Key/Value Set	1		
Mine for a Similar-Task Social Network	1	2	
Animate Transition System	1		1
Declare Miner	1	2	1
MoBuCon LTL	1		
Declare Miner Default	1	1	
Operational Support Service	1		
Rename/Merge Events	1	1	
Calculate Log Meta Data	1		
Mine for a Reassignment Social Network	1		
Mine for a Subcontracting Social Network	1		1
Convert Causal Net to Petri Net		1	2
Genetic Miner - from initial population		6	2
Analyze with Woflan		1	
Show PomPom View		1	1
Construct Coverability Graph of a Petri Net		1	
Construct Reachability Graph of a Petri Net		1	
Analyze Structural Property of Petri net		1	1
Analyze Behavioral Property of Petri net		1	
Filter on Timeframe		1	
Convert Declare to LTL		1	1
Add Missing Events		1	1
Convert Petri Net using Regions		1	
Convert EPC to Petri net		1	2
Replay a Log on Causal Net for Conformance Analysis		1	
Convert Heuristics net into Flexible model			1
Analyze Model Using Uma		1	1
Simplify Mined Model Using Uma		1	
Reduce Silent Tranistions		1	
Convert Heuristics net into Petri net		2	2
Compute Fitness			2
Guide Tree Miner		3	
Merge two Event Logs using AIS algorithm		1	
Unfold Petri Net with Data to Petri Net			1
Concept Drift			1
Operational Support Annotation Provider			1

Table 8. ProM 6.1 plug-ins: evaluation

|	*default settings*	*fast*	*trust*	*understand*	*intuitive*
Trace Alignment (with Guide Tree)	40%	60%	60%	60%	40%
Genetic Miner - from initial population	67%	67%	67%	67%	67%
LTL Checker	27%	82%	100%	100%	45%
Mine for a PN using ILP	71%	86%	71%	71%	71%
Mine Transition System	20%	80%	100%	100%	80%
Flexible Heuristics Miner	64%	82%	73%	100%	73%
Replay Log on PN for Conf. Analysis	67%	92%	100%	75%	67%
LTL Checker Default	63%	88%	100%	100%	50%
Heuristics Miner	75%	80%	80%	100%	75%
Mine for a Fuzzy Model	79%	89%	95%	89%	63%
Animate Event Log in Fuzzy instance	83%	83%	100%	75%	92%
Analyze using Dotted Chart	100%	100%	92%	77%	77%
Mine for a PN using Alpha-algorithm	100%	100%	82%	91%	82%
Add Artificial Events	100%	100%	100%	100%	88%
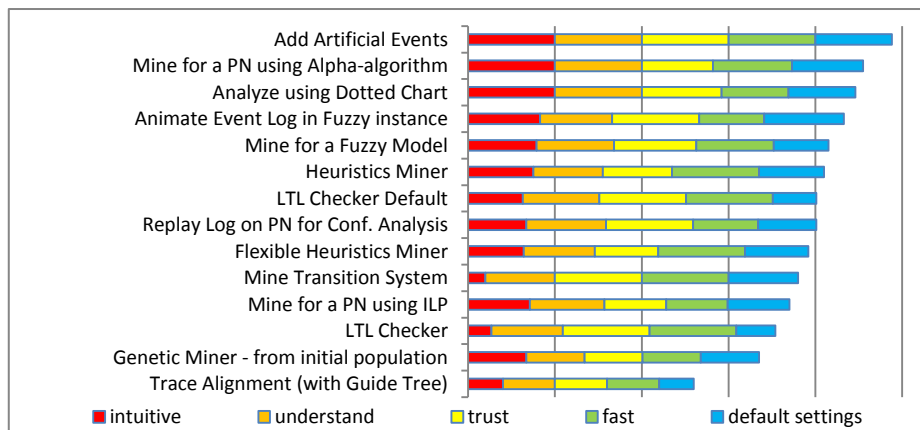
Fig. 8. ProM 6.1 plug-ins: evaluation

2.9   Benefits and drawbacks of ProM

Question 9. What is the main positive and the main negative thing in ProM? (78 respondents)

The principal benefit of the ProM framework is the many plug-ins. A whole set of different techniques are bundled in one process mining framework (many of them once existed in a separate tool). But the large number of tools might also have led to the main

drawback of the ProM tool: its limited ease of use (and lack of intuitiveness and guidance)? It is not only the number of plug-ins, but also the implementation of some specific plug-ins that limits the ease of use (see raw data set for individual answers in category 'ease of use').

Discussion. Many process mining techniques are complex by nature and maybe developers didn't succeed to hide this complexity for end users. Notice that ProM is mainly an academic tool and for researchers it might not be a goal to make user friendly plug-ins as they might be only interested in testing the theoretical correctness of their technique. The question can be raised if these plug-ins should be contained by an official release that is also meant for practical use (maybe releasing two versions of Prom might bring a solution: a limited one with only user friendly plug-ins and an extended one for research and experts only?)
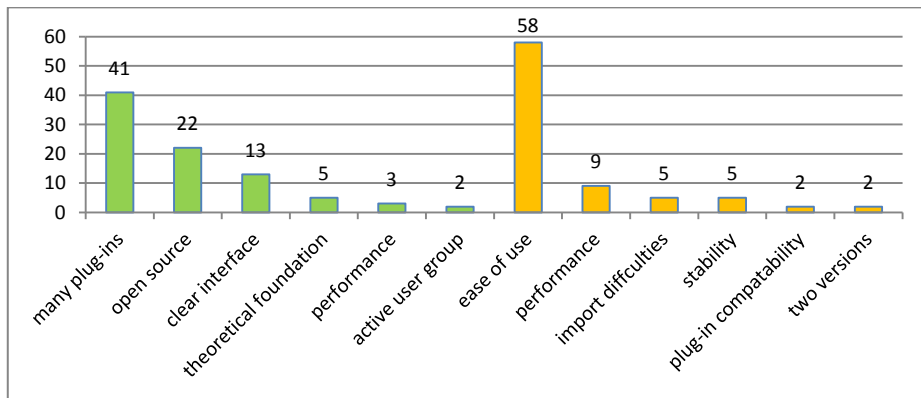
Fig. 9. Benefits and drawbacks of ProM
(green: benefits, orange: drawbacks)

2.10   Plug-ins of ProM 5.2 that are missed in ProM 6.1

Question 10. What plugins do you miss in ProM 6.1 These could be plugins from ProM 5.2 or new ones (please describe) (28 respondents)

We already noticed that the older ProM 5.2 version is still heavily used. One of the main reasons for this is probably because some very popular plug-ins are missing in the newest version ProM 6.1. Therefore, we asked the respondents which functionality they deemed missing in the newest version. We observe mainly a performance analysis plug-in and a model and/or log editor are requested.

Table 9. Plug-ins that are missed in ProM 6.1

Existing plug-ins from ProM 5.2	*New or enhanced plug-ins*
Advanced filters (5x)	Robust performance analysis (2x)
Conformance Checker (3x)	Log/Model editor (2x)
Basic Performance Analysis (3x)	Security analysis (2x)
Performance Sequence Diagram Analysis (2x)	*Better* process discovery techniques

|	
Alpha Algorithm(s)	*Better* performance analysis plugin
Trace Clustering	Medical analysis plug-in
Region Miner	Self organising maps
Pattern Sequence Analyser	Export to image option in *all* plugins

2.11 Demographical information

Question 11. In which role do you/would you use process mining? (98 respondents)
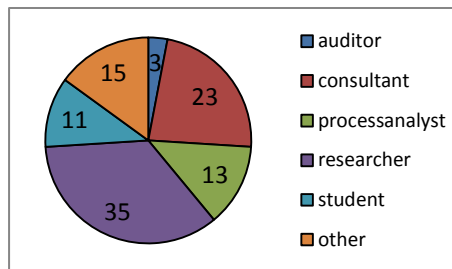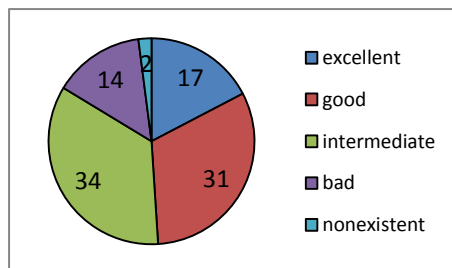

Fig. 10. Profession of respondents

Question 12. In which context do you/would you use process mining? If you are a researcher, think about the context/application of your research. (87 respondents)
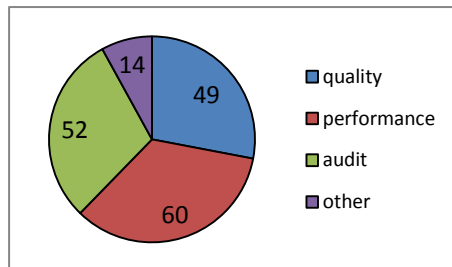
Fig. 11. Application domain of respondents

**Question 13.** How do you rate your own expertise in process mining? (98 respondents)

Fig. 12. Experience of respondents

Question 14. In which country do you work? (90 respondents)
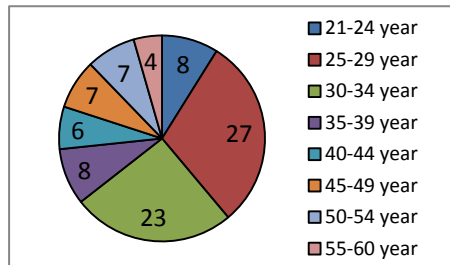
Table 10. Country

Netherlands	27
Belgium	16
Germany	7
Israel	5
United States	5
Austria	3
China	2
Australia	2
United Kingdom	2
Chile	2
Portugal	2
Switzerland	2
Spain	1
India	1
Poland	1
France	1
Peru	1
Japan	1
Ireland	1
Finland	1
Slovakia	1
Turkey	1
Norway	1
Sweden	1
Italy	1
Greece	1
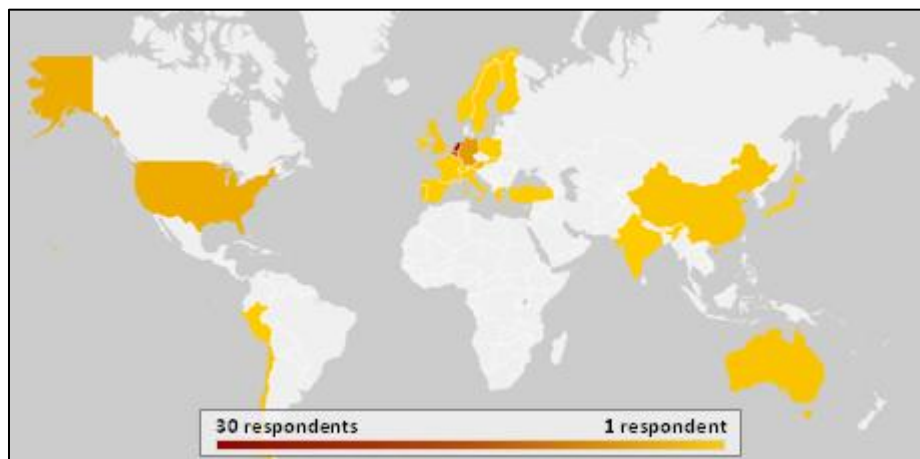
Fig. 13. Country where respondents work

Question 15. How old are you? (90 respondents)

Fig. 14. Age of respondents

3 Conclusion

As a short conclusion, we think that many known 'facts' are confirmed by this survey. We summarize the points we find most interesting.

- There is an active community working on process mining projects consisting of developers, researchers and process analysts (see Fig. 1, Fig. 2, Section 2.11, and Section 2.4)
- It is difficult to introduce and sell process mining techniques. Many people use a metaphor (see Section 2.3) or refer to different sources (see Section 2.1).
- ProM is still the most used process mining tool, although the survey indicates that Disco might also get very popular (see Section 2.2).
- The main big problem with process mining is that the preparation phase (finding, structuring, improving quality of data) is undersupported (see Section 2.5). It is not clear if this is a problem that can be resolved by research, but there is in our opinion at least room for improvement.
- ProM 5.2 is still used a lot (see Section 2.6). Especially practitioners seem to like this version for its stability and completeness. Maybe this is an opportunity for researchers or master students to convert the plug-ins to ProM 6.1 and to make this version and its plug-ins more stable (see also Section 2.10).
- Section 2.7 and 2.8 give an overview of which plug-ins are mostly used and what the respondents think of these plug-ins.

Although we think, due to the many respondents, this survey gives a good indication to the true answers to the questions, there is still need for more research to prove the answers we derived from the collected data. In any case the survey revealed or confirmed a lot of points in the field that can be improved. We think, it gives substantial input for researchers to derive hypotheses to corroborate, for developers to pick priorities and for users to see the current process mining domain in the right perspective.