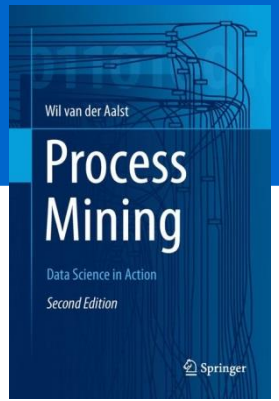
Process Mining: Data Science in Action

Process Mining Software

Available software and academic programs



prof.dr.ir. Wil van der Aalst

www.processmining.org

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

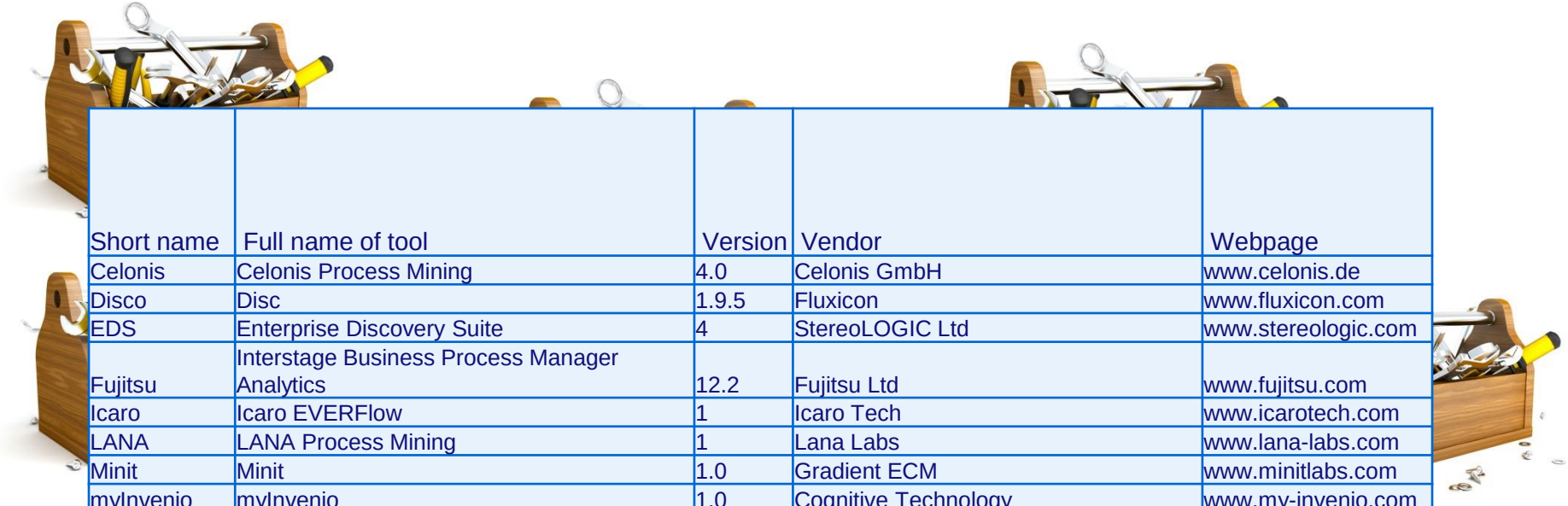
The course uses ProM and Disco

ProM
process mining workbench

Disco
by Fluxicon.

… , but there are many more process mining tools

… , but there are many more process mining tools

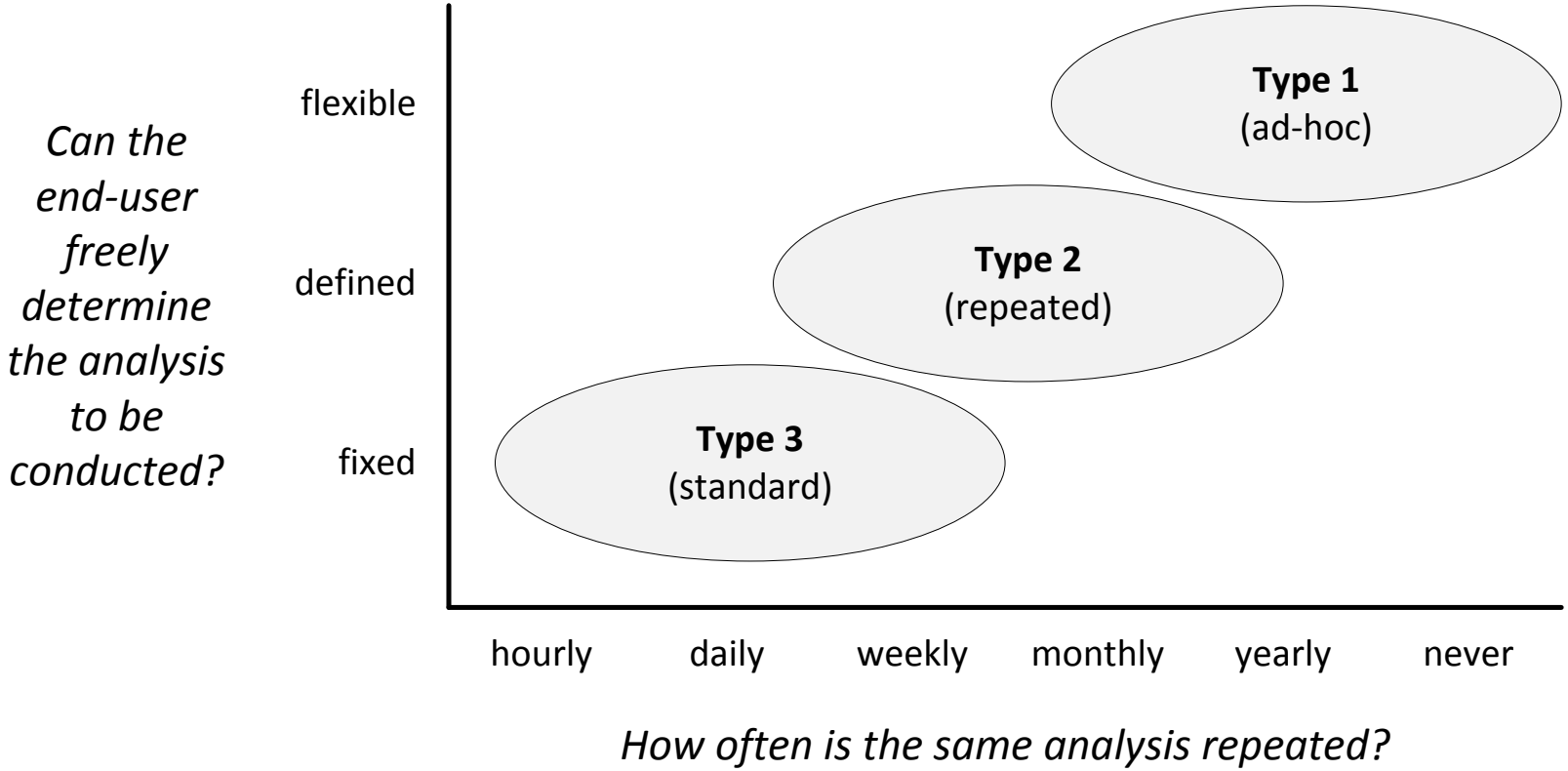


Short name	Full name of tool	Version	Vendor	Webpage
Celonis	Celonis Process Mining	4.0	Celonis GmbH	www.celonis.de
Disco	Disc	1.9.5	Fluxicon	www.fluxicon.com
EDS	Enterprise Discovery Suite	4	StereoLOGIC Ltd	www.stereologic.com
Fujitsu	Interstage Business Process Manager Analytics	12.2	Fujitsu Ltd	www.fujitsu.com
Icaro	Icaro EVERFlow	1	Icaro Tech	www.icarotech.com
LANA	LANA Process Mining	1	Lana Labs	www.lana-labs.com
Minit	Minit	1.0	Gradient ECM	www.minitlabs.com
myInvenio	myInvenio	1.0	Cognitive Technology	www.my-invenio.com
Perceptive	Perceptive Process Mining	2.7	Lemark	www.lexmark.com
ProM	ProM	6.6	Open Source hosted at TU/e	www.promtools.org
ProM Lite	ProM Lite	1.1	Open Source hosted at TU/e	www.promtools.org
QPR	QPR ProcessAnalyzer	2015.5	QPR	www.qpr.com
RapidProM	RapidProM	4.0.0	Open Source hosted at TU/e	www.rapidprom.org
Rialto	Rialto Process	1.5	Exeura	www.exeura.eu
SNP	SNP Business Process Analysis	15.27	SNP Schneider-Neureither & Partner AG	www.snp-bpa.com
PPM	webMethods Process Performance Manager	9.9	Software AG	www.softwareag.com

… , but there are many more process mining tools

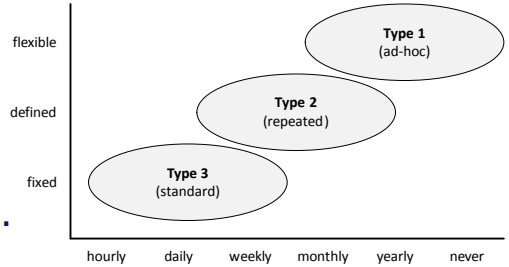
- Since 2011 many new (commercial) process mining tools emerged.
- The functionality of existing tools changed significantly (e.g., scalability, features, etc.).
- The tool chapter in the process mining book was completely rewritten to reflect this.
- These slides provide an update.

Three types of use cases



Three types of use cases

- Use cases of Type 1 (ad-hoc) require a spreadsheet-like tool: questions are ad-hoc and the user needs to have complete freedom to perform analysis. The analysis process is iterative and undefined. The results of one analysis step may lead to unanticipated additional data extractions (or transformations) to enable the next analysis step. Analysis workflows are unique and seldom repeated.
- Use cases of Type 2 (repeated) involve questions that are recurring, but at a lower frequency. Analysis workflows may be predefined but not completely fixed. Customization may be needed and the interpretation of the results requires knowledge of process mining and understanding of the data.
- Use cases of Type 3 (standard) involve routine questions that are recurring frequently. The different analysis views are fixed and no customization is possible. The user only needs to understand predefined dashboard-like views.

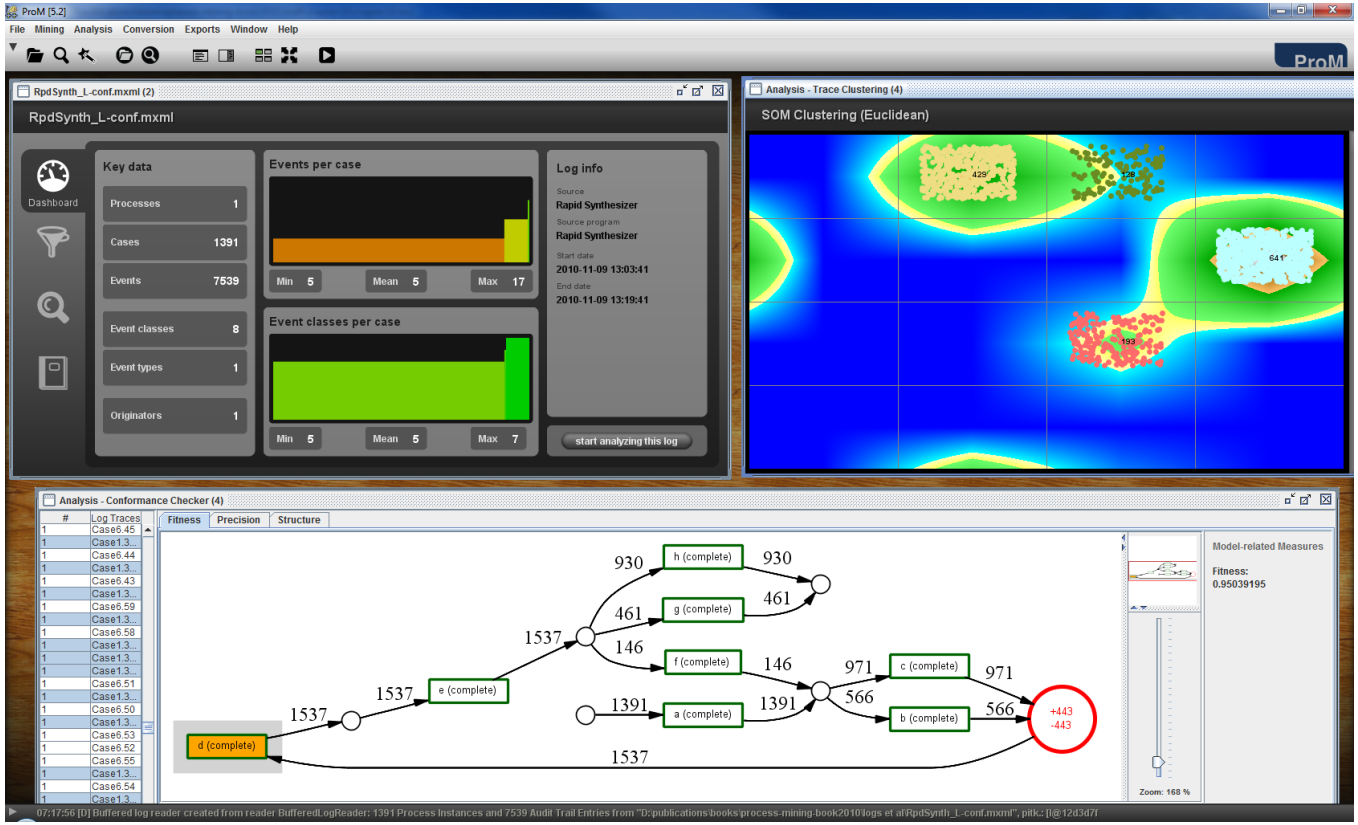


TU/e

History of ProM

- Predecessors of ProM: MiMo, EMiT, Little Thumb, …
- First version of ProM released on 2004 (only 29 plug-ins).
- Evolved into different versions until ProM 5.2 with 286 plug-ins (2009)
- ProM 6 (released in 2010) was developed from scratch.
- Current version has over 1500 plug-ins when all available packages are installed.

Screenshot of ProM 5.2

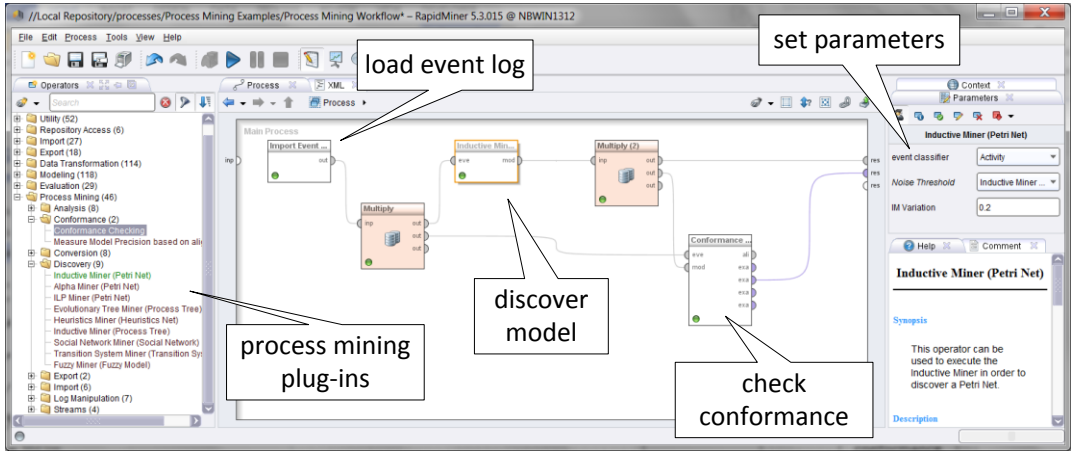


Current ProM versions

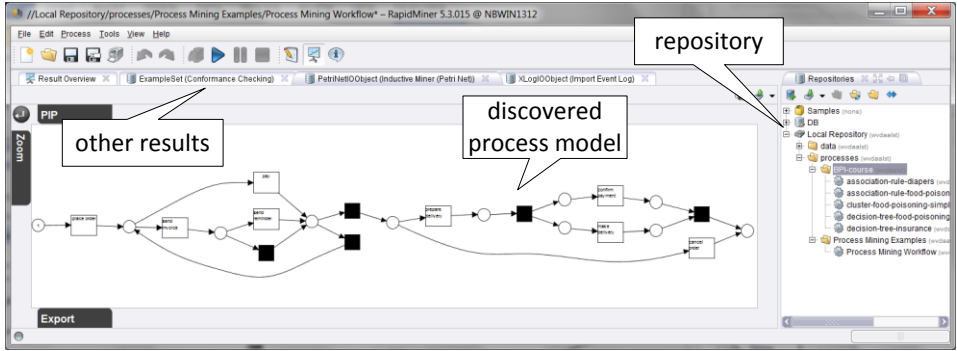
- **ProM 6.6 (complete version intended for experts)**
  http://www.promtools.org/doku.php?id=prom651
- **ProM Lite 1.1 (contains only the most used packages)**
  http://www.promtools.org/doku.php?id=promlite11
- **RapidProM (embedded in RapidMiner)**
  http://www.rapidprom.org/

Screenshot of ProM 6.6

RapidProM



- **Designed analysis workflows can be reused.**
- **Allows for the combination of data mining, text mining, etc. with process mining**



Characteristics of tools

- Dedicated process mining software versus embedded process mining software.
- Open-source process mining software versus closed-source process mining software.
- Models supported: informal process models (“boxes and arrows”), formal low-level process models (transition systems, Markov chains, episodes, sequences, etc.), and/or formal high-level process models (BPMN models, EPC models, UML activity diagrams, Petri nets, process trees, etc.).

TU/e

Commercial software

- **Examples of commercial tools include Celonis Process Mining, Disco (Fluxicon), Fujitsu Interstage Business Process Manager Analytics, Minit (Gradient), myInvenio, Perceptive Process Mining (Lexmark), QPR ProcessAnalyzer, Rialto Process, SNP Business Process Analysis, and webMethods Process Performance Manager (PPM).**
- **Disco, Fujitsu, QPR, and PPM have been around for a few years. Since 2011 many new tools/vendors emerged.**

A few examples

Disco - New project

event_log_10000_cases +17 | Map | Statistics | Cases | Academic w.m.p.v.d.aalst@tue.nl | Disco

Statistics views

- Overview — Global statistics
- Activity — Activity classes
- product — Other attribute
- prod-price — Other attribute

Overview
Global statistics

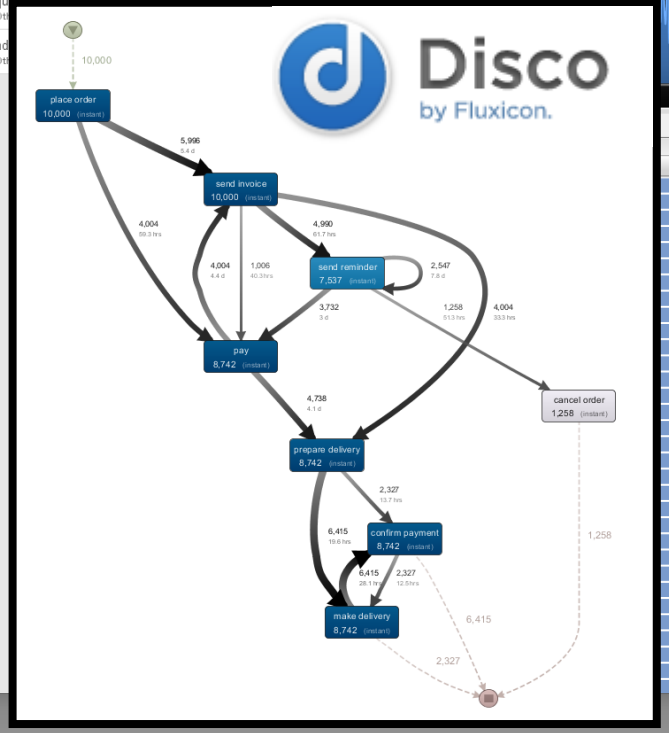
Events over time | Active cases over time | Case variants | Events per case | Case duration

Log timeline

Events	63,763
Cases	10,000
Activities	8
Median case duration	13.9 d
Mean case duration	14.9 d
Start	05.01.2015 09:00:07
End	31.12.2019 14:46:02

Cases (10000) | Variants (9)

Variant	Started	Finished	Duration
7	05.01.2015 09:00:07	26.01.2015 16:42:28	21 days, 7 hours
1	05.01.2015 10:18:21	15.01.2015 15:52:30	10 days, 5 hours
4	05.01.2015 11:54:49	09.01.2015 18:38:58	4 days, 6 hours
3	05.01.2015 14:07:45	22.01.2015 13:18:30	16 days, 23 hours
1	05.01.2015 15:33:38	12.01.2015 17:27:36	7 days, 1 hour
5	05.01.2015 17:25:23	02.02.2015 12:31:09	27 days, 19 hours
4	05.01.2015 19:08:53	15.01.2015 14:56:54	9 days, 19 hours
9	05.01.2015 21:54:00	13.01.2015 15:49:53	7 days, 17 hours
4	06.01.2015 07:25:13	15.01.2015 11:27:50	9 days, 4 hours
1	06.01.2015 10:09:51	15.01.2015 19:15:18	9 days, 9 hours
1	06.01.2015 11:37:49	14.01.2015 09:14:28	7 days, 21 hours
4	06.01.2015 13:33:45	14.01.2015 11:30:05	7 days, 21 hours
4	06.01.2015 15:25:38	13.01.2015 12:25:34	6 days, 20 hours
2	06.01.2015 17:09:23	22.01.2015 18:59:10	16 days, 1 hour
3	06.01.2015 18:36:53	22.01.2015 14:39:39	15 days, 20 hours
8	06.01.2015 21:26:54	26.01.2015 17:16:02	19 days, 19 hours
1	07.01.2015 04:42:36	16.01.2015 10:17:14	9 days, 5 hours
3	07.01.2015 10:10:58	21.01.2015 17:31:29	14 days, 7 hours
8	07.01.2015 11:40:04	28.01.2015 10:27:12	20 days, 22 hours
9	07.01.2015 13:38:15	13.01.2015 13:22:15	5 days, 23 hours
1	07.01.2015 15:34:37	19.01.2015 09:11:23	11 days, 17 hours
1	07.01.2015 17:27:21	16.01.2015 09:09:25	8 days, 15 hours
5	07.01.2015 19:12:50	03.02.2015 14:34:33	26 days, 19 hours
6	07.01.2015 22:01:54	19.01.2015 13:15:02	11 days, 15 hours
8	08.01.2015 07:12:36	28.01.2015 10:41:14	20 days, 3 hours
3	08.01.2015 09:55:59	26.01.2015 15:52:42	18 days, 5 hours
6	08.01.2015 12:10:05	15.01.2015 13:54:59	7 days, 1 hour
1	08.01.2015 13:38:17	14.01.2015 12:30:26	5 days, 22 hours
5	08.01.2015 15:34:42	02.02.2015 14:10:36	24 days, 22 hours
2	08.01.2015 17:27:31	29.01.2015 11:26:06	20 days, 17 hours
3	08.01.2015 19:13:09	26.01.2015 14:16:02	17 days, 19 hours
4	08.01.2015 22:02:32	20.01.2015 10:35:40	11 days, 12 hours



Filter | Copy | Delete | Export

Version 1.9.1

Disco - New project

event_log_10000_cases +17

Map | Statistics | Cases

Academic
w.m.p.v.d.aalst@tue.nl

Disco

Zoom: 159%

search...

Detail

Activities 100%

Paths 100%

Frequency

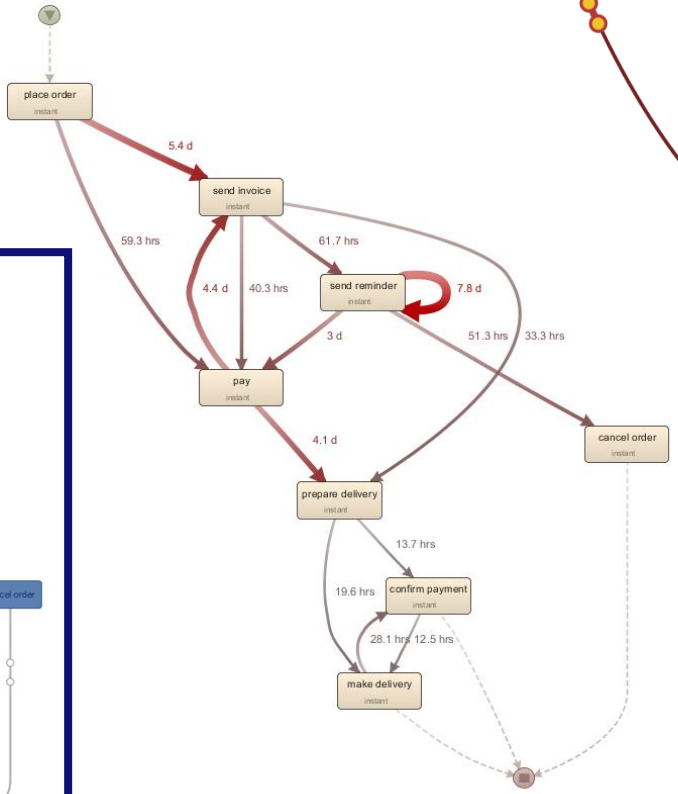
Performance

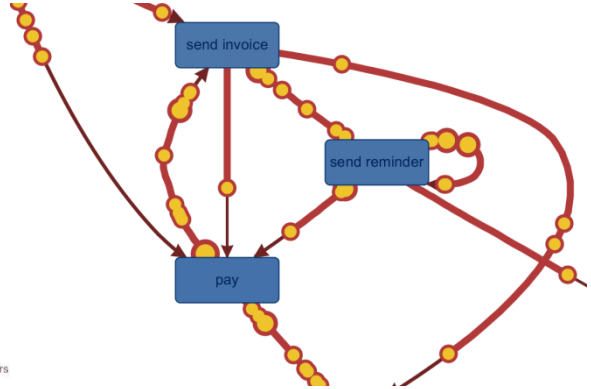
Show: Mean duration

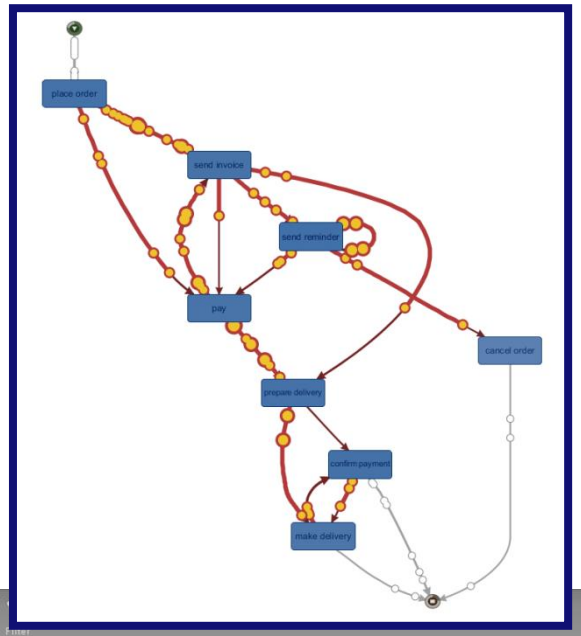
instant 6.3 d
instant 4.7 d
instant 3.1 d
instant 37.6 hrs

Add secondary metrics

Disco
by Fluxicon.







Filter

Animation

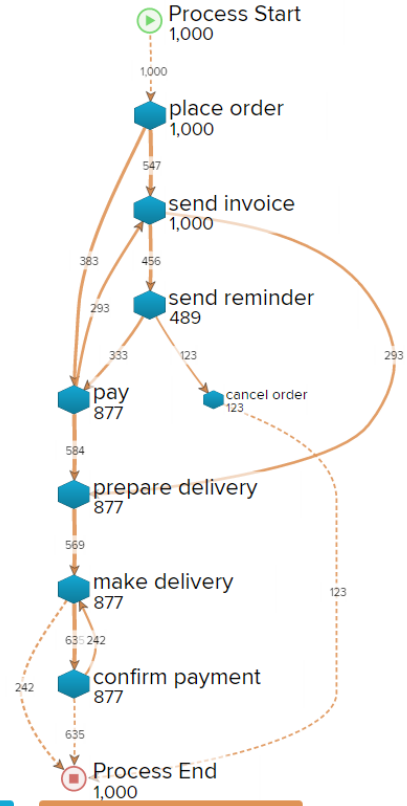
Copy | Delete | Export

Version 1.9.1

Celonis Process Mining

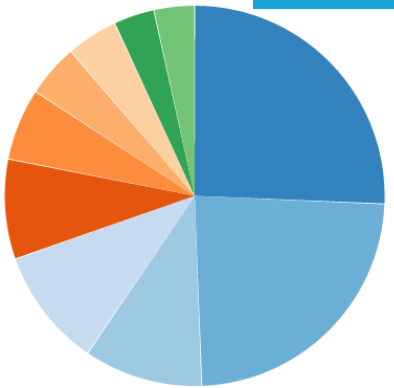
Artificial Logs Original / Spreadsheet-1000-cases
Last save on 2016-03-04 at 15:38

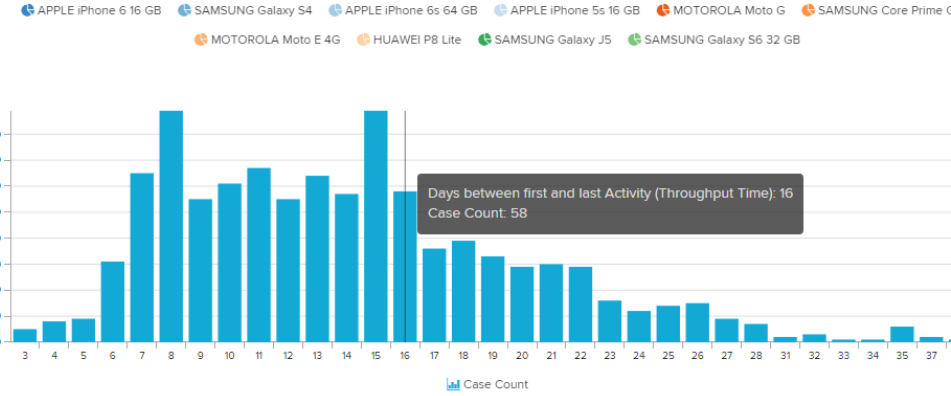
100% 1000 of 1000 cases selected



Activities 100%

Connections 87.8%



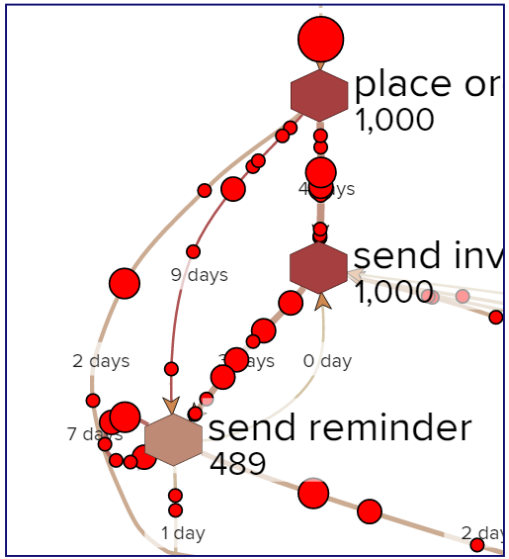


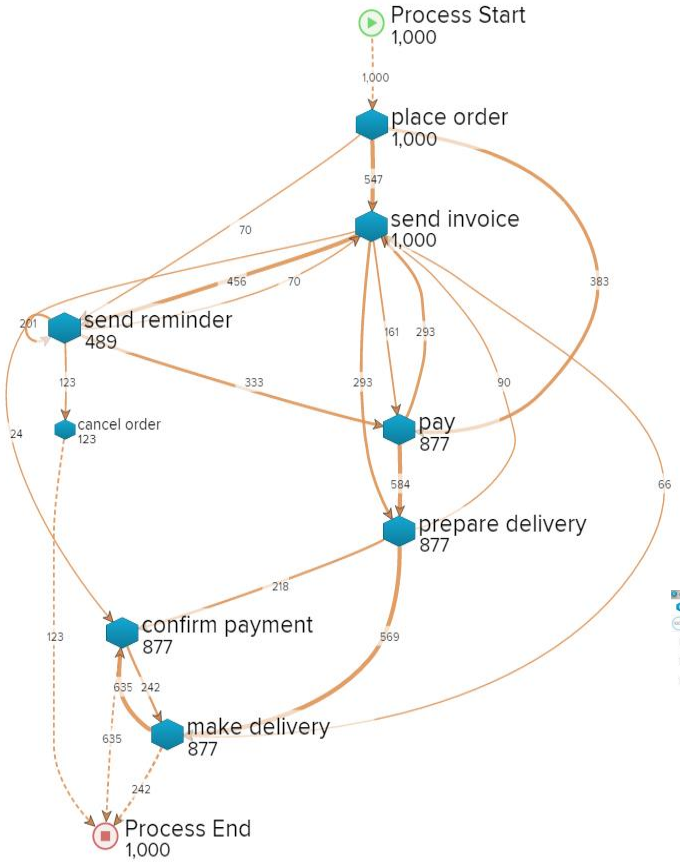
modelguidedabstraction.tex updated
"modelguidedabstraction.tex" was updated to the latest version (click to view).

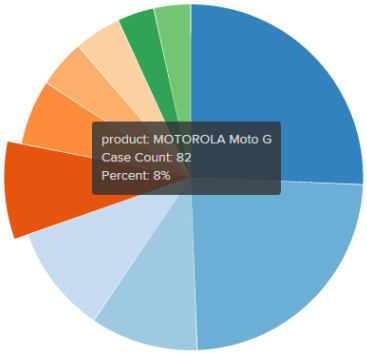
Process | Cases | Sheet 1

celonis
process mining

STORIES HELP Edit Refresh

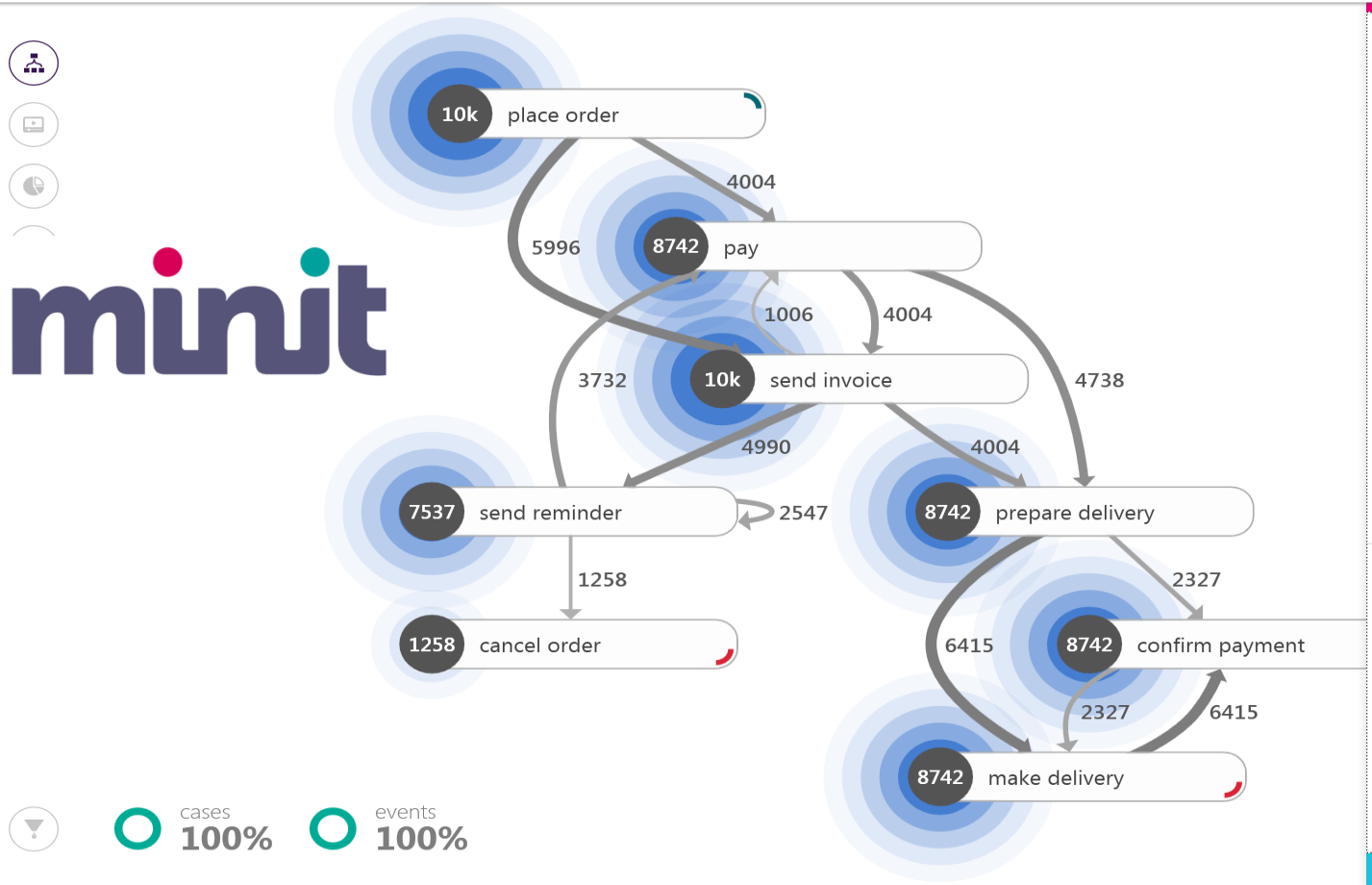






Activities 100% Connections 100%

Process Cases Sheet 1



cases 100%
events 100%

Customize

Process view | Social view

Missing attribute of type Resource.

Show terminal nodes
Snap to backbone
Left to right
Highlight predecessor/successor activities

Activities 100 %
Paths 100 %

Frequency | Performance

Event count

2000 2500 3333 5000 10k

1283 1603 2138 3207 6415

Visualize

myInvenio | reviewing

https://app1.my-invenio.com/workflow?p=reviewing#

myinvenio | Datasource | Dashboard | Model | Activity map | Social net | Filters | Export | Wil van der Aalst

reviewing

Portrait | Landscape

100% 100%

100 cases 2278 events

Resolution Tokens

Sat Dec 31, 2005 23:00:00

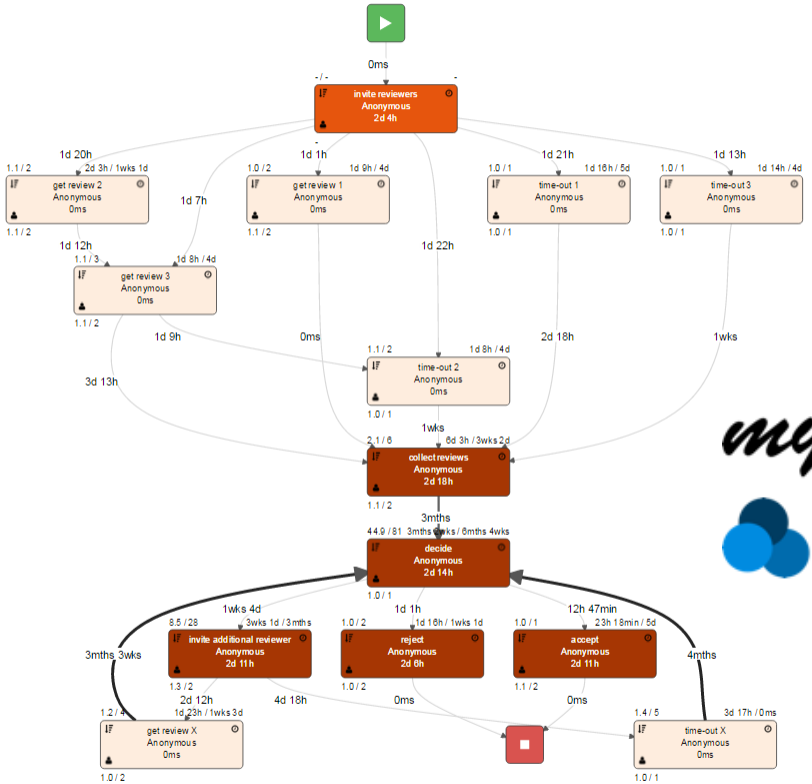
Activity

Model's details

Activities

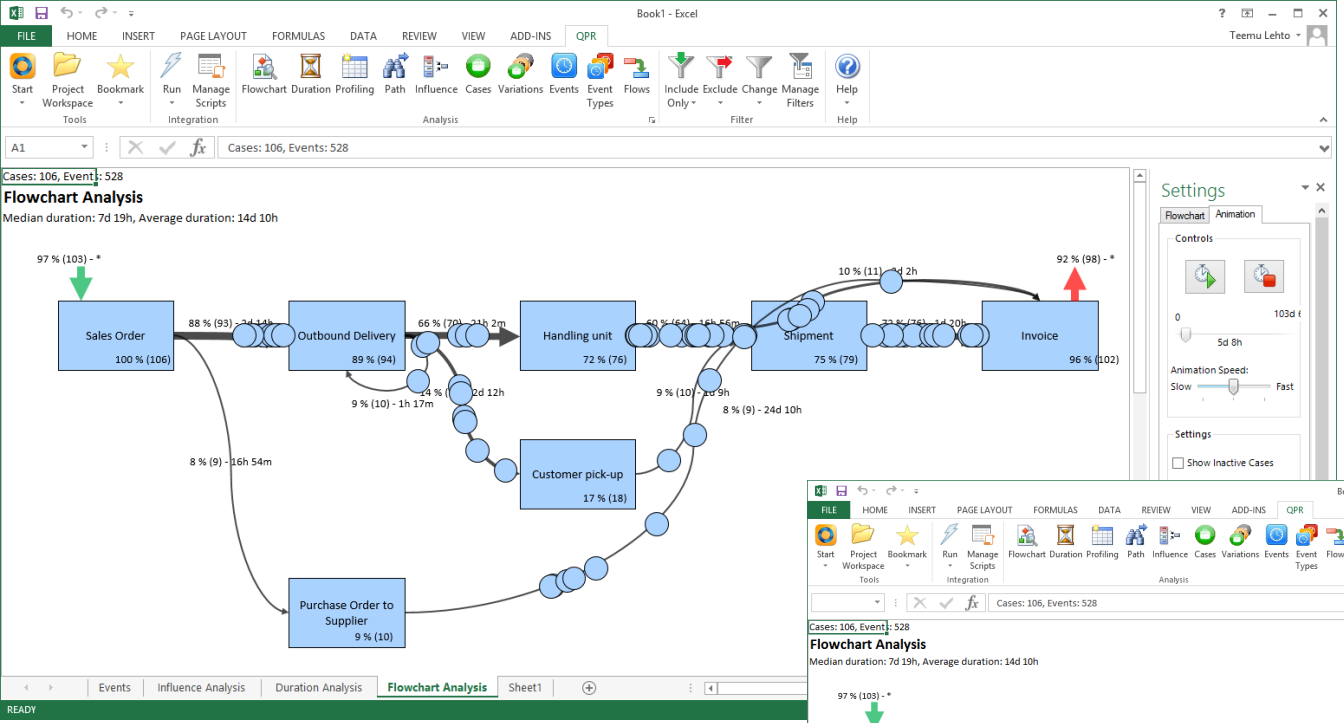
Relations

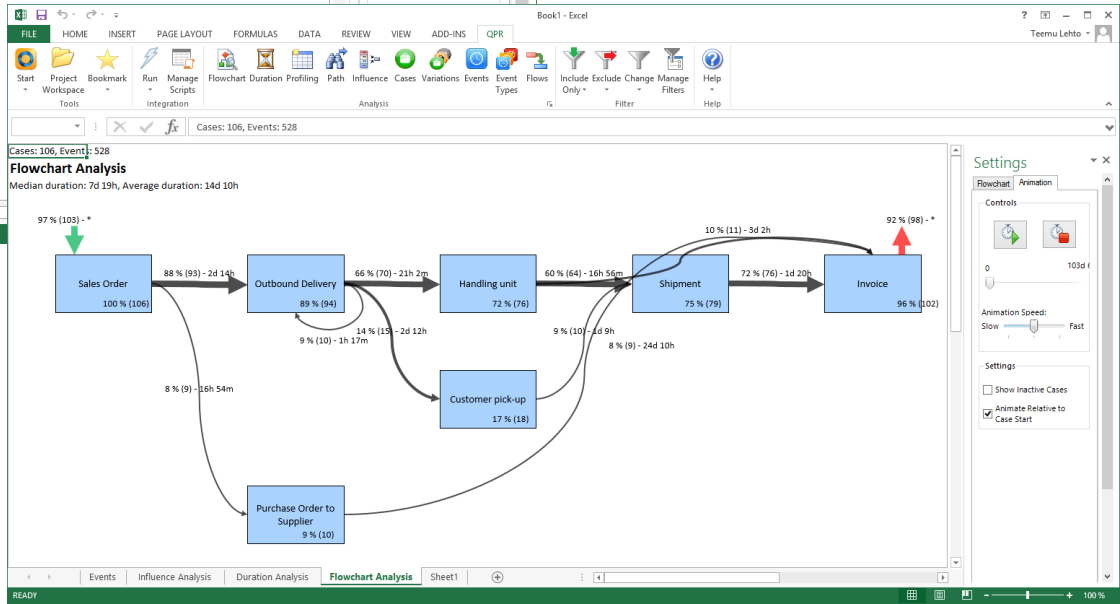
Resource focus

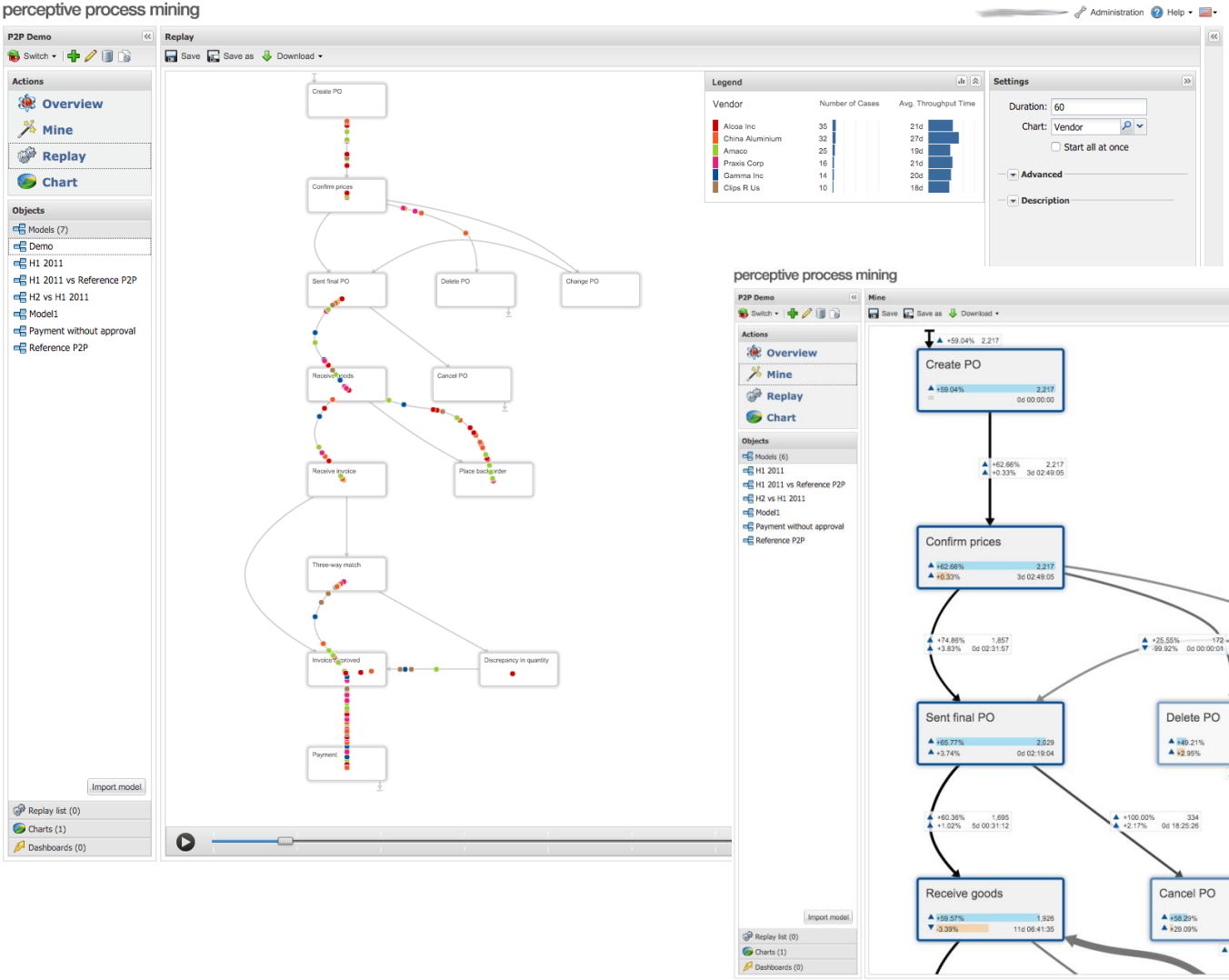


myinvenio COGNITIVE TECHNOLOGY

0ms
13h 20min
1d 2h

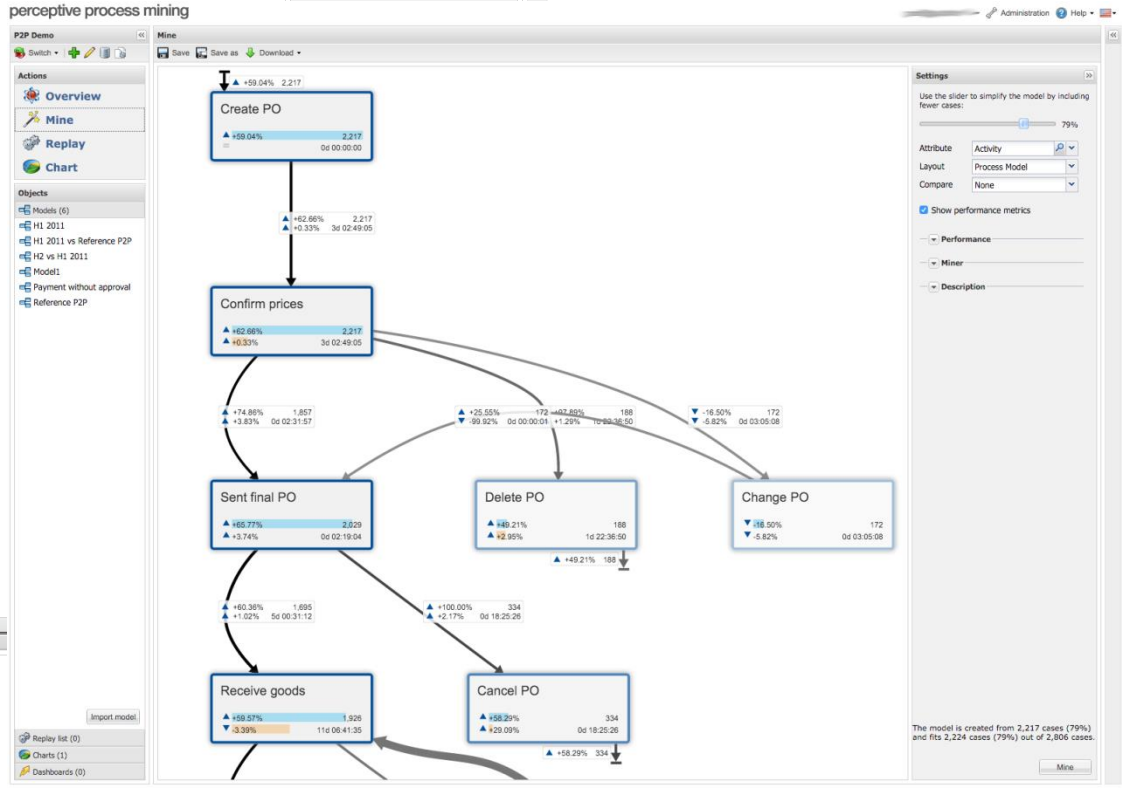

Book1 - Excel
Teemu Lehto
FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ADD-INS QPR
Start Project Workspace Bookmark Tools
Run Manage Scripts Integration
Flowchart Duration Profiling Path Influence Cases Variations Events Event Types Flows Analysis
Include Only Exclude Change Manage Filters Filter
Help Help
A1
Cases: 106, Events: 528
Cases: 106, Events: 528
Flowchart Analysis
Median duration: 7d 19h, Average duration: 14d 10h
97 % (103) - *
Sales Order
100 % (106)
88 % (93) - 2d 14h
Outbound Delivery
89 % (94)
66 % (70) - 21h 2m
Handling unit
72 % (76)
60 % (64) - 16h 56m
Shipment
75 % (79)
72 % (76) - 1d 20h
Invoice
96 % (102)
92 % (98) - *
10 % (11) - 3d 2h
9 % (10) - 1h 17m
2d 12h
9 % (10) - 1d 9h
8 % (9) - 24d 10h
8 % (9) - 16h 54m
Customer pick-up
17 % (18)
Purchase Order to Supplier
9 % (10)
Settings
Flowchart Animation
Controls
0
103d
5d 8h
Animation Speed:
Slow Fast
Settings
Show Inactive Cases
Events Influence Analysis Duration Analysis Flowchart Analysis Sheet1
READY


Book1 - Excel
Teemu Lehto
FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ADD-INS QPR
Start Project Workspace Bookmark Tools
Run Manage Scripts Integration
Flowchart Duration Profiling Path Influence Cases Variations Events Event Types Flows Analysis
Include Only Exclude Change Manage Filters Filter
Help Help
Cases: 106, Events: 528
Cases: 106, Events: 528
Flowchart Analysis
Median duration: 7d 19h, Average duration: 14d 10h
97 % (103) - *
Sales Order
100 % (106)
88 % (93) - 2d 14h
Outbound Delivery
89 % (94)
66 % (70) - 21h 2m
Handling unit
72 % (76)
60 % (64) - 16h 56m
Shipment
75 % (79)
72 % (76) - 1d 20h
Invoice
96 % (102)
92 % (98) - *
10 % (11) - 3d 2h
9 % (10) - 1h 17m
14 % (15) - 2d 12h
9 % (10) - 1d 9h
8 % (9) - 24d 10h
8 % (9) - 16h 54m
Customer pick-up
17 % (18)
Purchase Order to Supplier
9 % (10)
Settings
Flowchart Animation
Controls
0
103d
Animation Speed:
Slow Fast
Settings
Show Inactive Cases
Animate Relative to Case Start
Events Influence Analysis Duration Analysis Flowchart Analysis Sheet1
READY
100 %

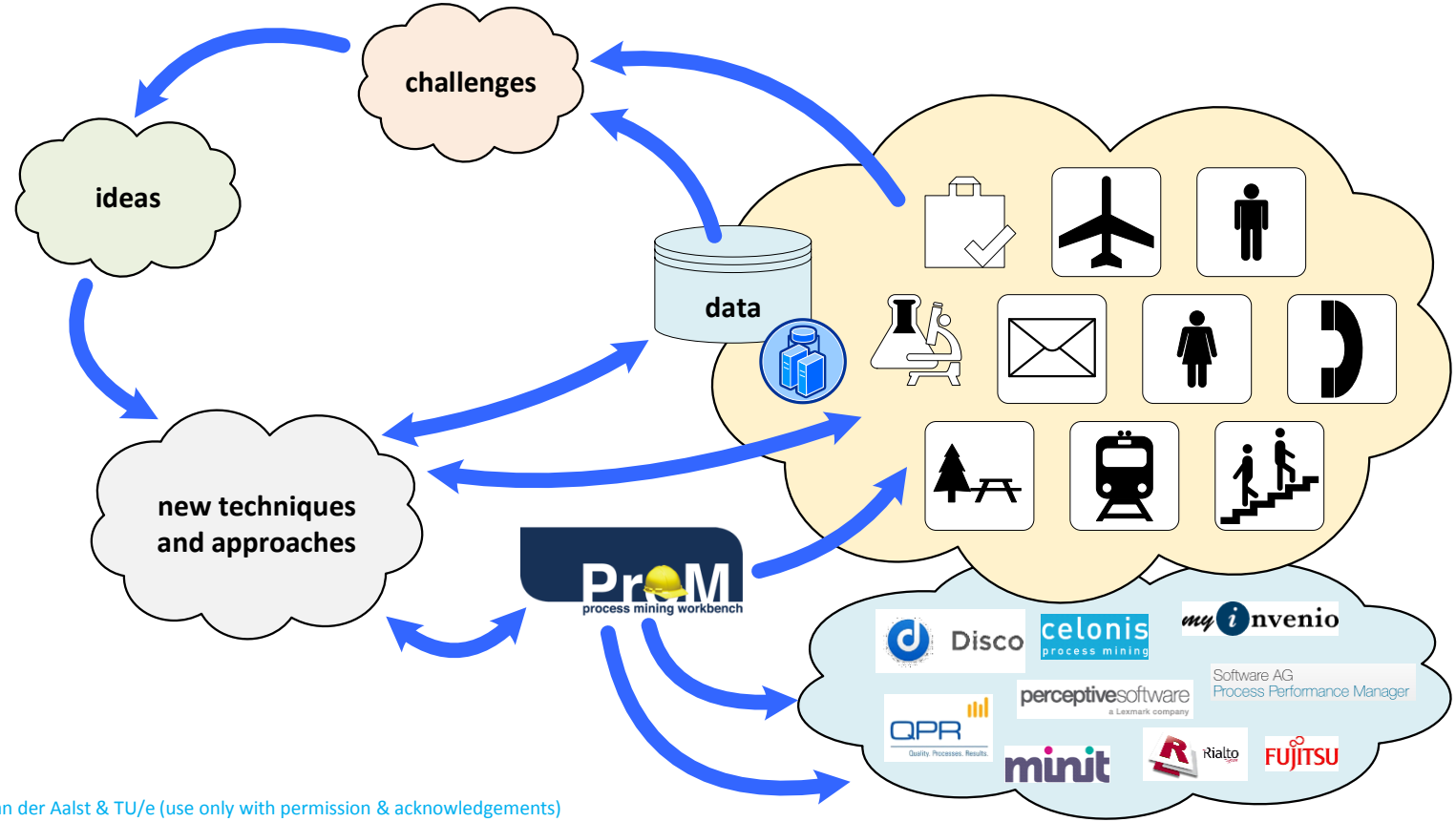


Lexmark™

Perceptive Process Mining



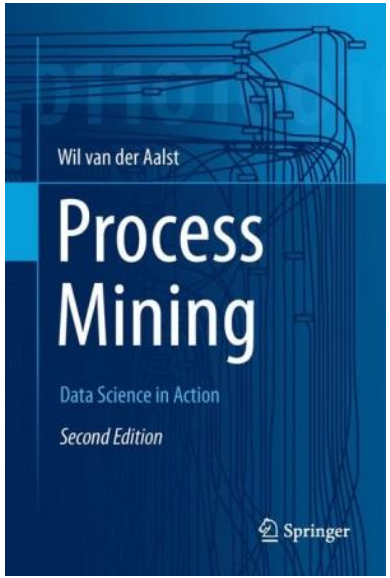
Interaction with industry



TU/e

Discussion on the strengths and weakness of commercial tools

See section 11.4.2 in book



Disclaimer: the capabilities of some tools are expanding/improving rapidly!

TU/e

Final recommendation

- Since the process mining market is developing fast, users are advised to test tools using their own event data.
- Check the quality of discovered process models using representative data.
- Carefully consider the features you would like to have related to conformance checking, performance analysis, etc.
- Compare different tools and don't be afraid to do a pilot project.
- Even when tools look similar, differences in terms of practical usability and scalability may be significant!

TU/e