



Gartner Business Process Management Summit 2009

23 – 25 February 2009 | London



Executive Keynote: Process Mining: Beyond Business Intelligence

Prof. dr. ir. Will van der Aalst, Professor of Information Systems, **Technische Universiteit Eindhoven**

This is something completely NEW, something people said wasn't possible, that the data wasn't there to allow systems that really could map out a process; they were wrong. Data is now everywhere; it is accessible, there is an abundance of data and it can provide you with insights you could never find just in interviews. The goal is to get away from workflow systems that are divorced from reality and from how people really work. Today's tools oversimplify reality when what you need is a view as close to the real world as possible.

Gartner Business Process Ma

23 – 25

This is something completely NEW, something people said wasn't possible, that the data wasn't there to allow systems that really could map out a process; they were wrong. Data is now

ent report:
ynote: **Process Mining:
ness Intelligence**

van der Aalst, Professor
of Information Systems, **Technische
Universiteit Eindhoven**

This is something completely NEW, something people said wasn't possible, that the data wasn't there to allow systems that really could map out a process; they were wrong. Data is now everywhere; it is accessible, there is an abundance of data and it can provide you with insights you could never find just in interviews. The goal is to get away from workflow systems that are divorced from reality and from how people really work. Today's tools oversimplify reality when what you need is a view as close to the real world as possible.



In 2009, I gave the keynote at the **Gartner Business Process Management Summit** in London. The title of my keynote was “**Process Mining: Beyond Business Intelligence**” and I explained what process mining was. I talked about event data, process discovery, conformance checking, evidence-based performance analysis, predictive analytics, process models as maps, process animation, and applications to Lasagna and Spaghetti processes. The slides that I presented in 2009 are attached. The audience loved it and the conference report, Gartner wrote “**Process Mining is something completely NEW, something people said wasn’t possible, that the data wasn’t there to allow systems that really could map out a process; THEY WERE WRONG**”. Yet, it took a long time to get where we are today.

In 2009, Futura Process Intelligence, founded by one of my students who worked on the ProM process mining framework, was also elected as a “Cool Vendor in Business Process Management” by Gartner. Gartner praised Futura’s work on automated business process discovery (ABPD): “Factors that differentiate Futura from many other offerings in the field of BPM include its strong focus on staying ahead of the curve by innovating and the highly intuitive way it provides insight into the historical execution of a process using a novel process animation technique”.

The look and feel of my slides reveal that this is almost 20 years ago. However, many of the ideas and algorithms were already in place and are still valid today. The internet tends to forget about these things. Details about Futura Process Intelligence and the Gartner BPM Summit 2009 are hard to find. It is a reminder that there is no such thing as a free lunch when it comes to process improvement. **The technology and data are there. However, to improve, one needs “changemakers” who are willing to address “inconvenient truths” persistently.**

Note these slides are from 2009!

Process Mining: Beyond Business Intelligence

Gartner Business Process Management Summit, February 2009, London

prof.dr.ir. Wil van der Aalst
www.processmining.org



TU / **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

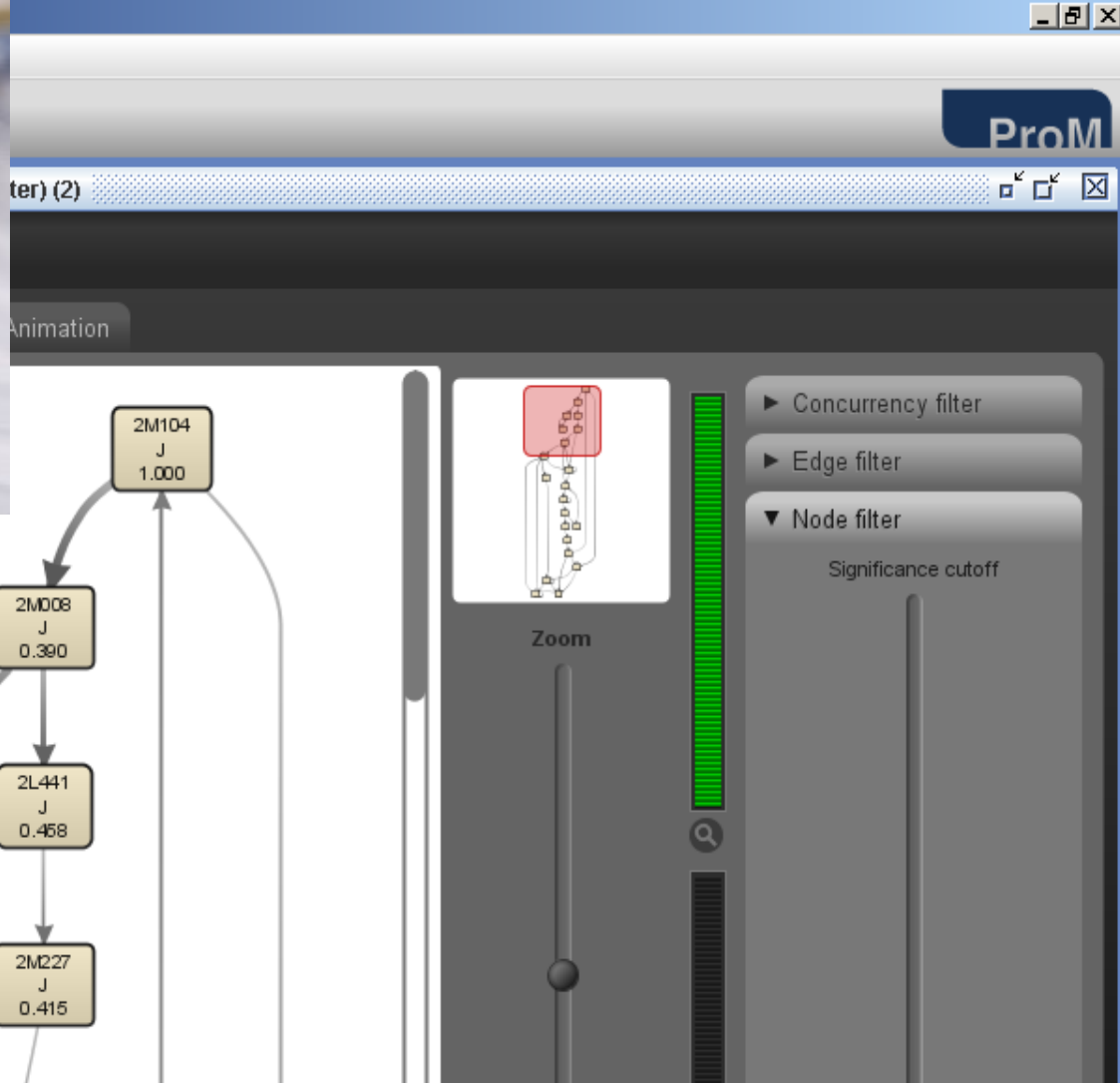


011011001111001011000100110110111011011001010010011100

Process Mining



- **Process discovery:** "What is really happening?"
- **Conformance checking:** "Do we do what was agreed upon?"
- **Performance analysis:** "Where are the bottlenecks?"
- **Process prediction:** "Will this case be late?"
- **Process improvement:** "How to redesign this process?"
- **Etc.**



- **Process discovery:** "What is the real curriculum?"
- **Conformance checking:** "Do students meet the prerequisites?"
- **Performance analysis:** "Where are the bottlenecks?"
- **Process prediction:** "Will a student complete his studies (in time)?"
- **Process improvement:** "How to redesign the curriculum?"

Outline

- **Trends in BPM**
- **Process Mining: The Basics**
 - **Input data**
 - **Discovery**
 - **Conformance**
 - **Software support**
- **Process Mining: Applications**
- **Process Mining: TomTom for Business Processes**
- **Conclusion**

Trends in BPM



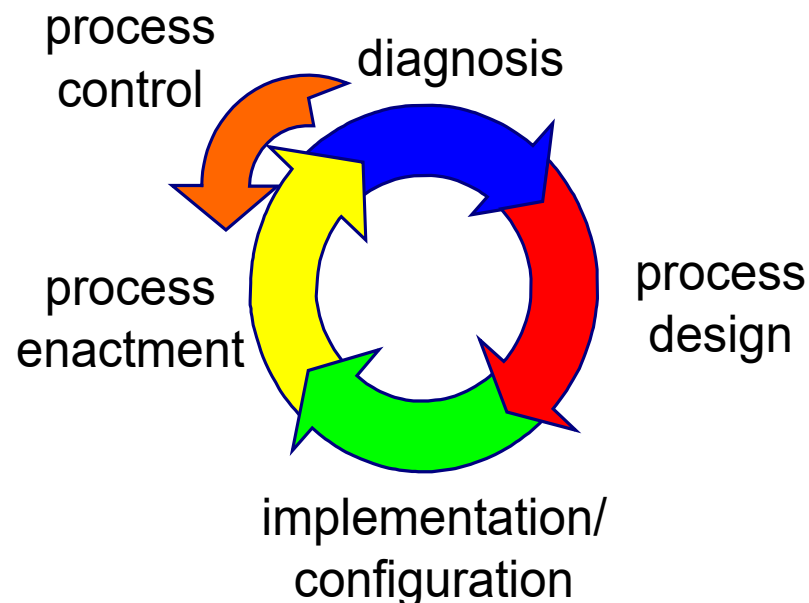
TU / **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

History

- The first workflow management systems (called "office automation systems") were implemented in seventies, cf. Petri-net-based systems such as Officetalk (Xerox Parc, Skip Ellis) and SCOOP (Wharton, Michael Zisman).
- Mid nineties: "explosion" of workflow products.
- Shift from workflow automation to business process management.





MS Workflow Foundation Global 360 BPM Suite

YAWL FileNet

InConcert

Fujitsu Interstage

Axxerion

BWise

Software AG/webMethods

XPDL IBM WebSphere

casewise

COSA

BPEL

UML ADs

Savvion BusinessManager

jBPM

BPM|one

TIBCO iProcess Suite

BPMN

EPCs

FlowConnect

SAP Workflow

Pegasystems SmartBPM Suite Ensemble

Bizagi

TeamWARE

Oracle BPEL

Ultimus BPM Suite

Promatis

BiZZdesigner

Workflow Patterns Initiative

- Initiative started in late 90-ties.
- Collections:
 - 43 control-flow patterns (process/routing)
 - 40 data patterns
 - 43 resource patterns (work distr. and organization)
 - exception, flexibility, service interaction, ... patterns
- Frequently used as a tool in selection processes.
- Influenced standards (BPMN, BPEL, etc.) and systems.
- See www.workflowpatterns.com (+/- 500 unique visitors per day)

Workflow Patterns

Problem is NOT the automation of structured processes!

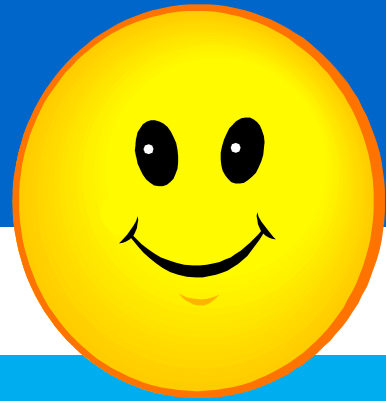
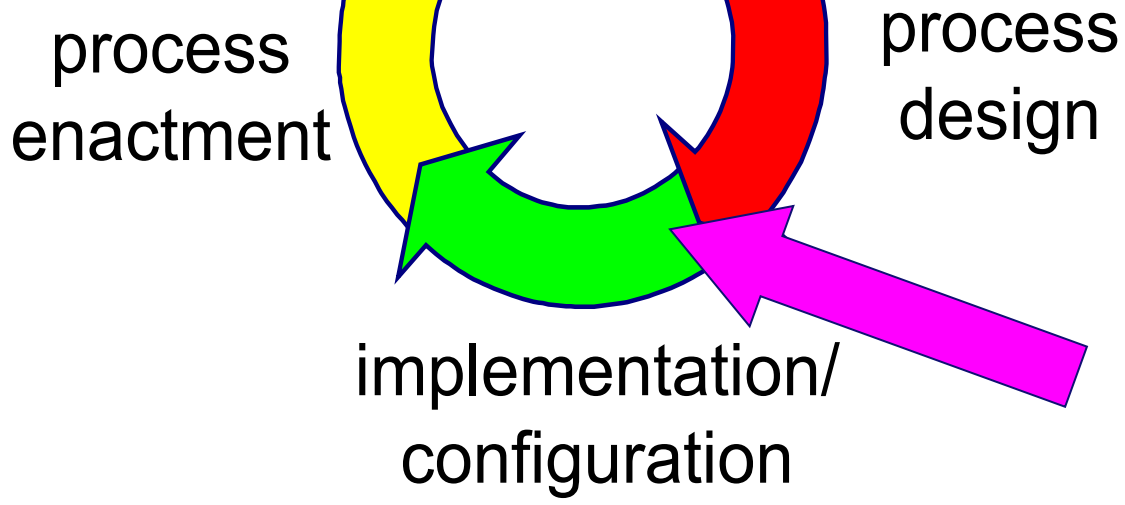
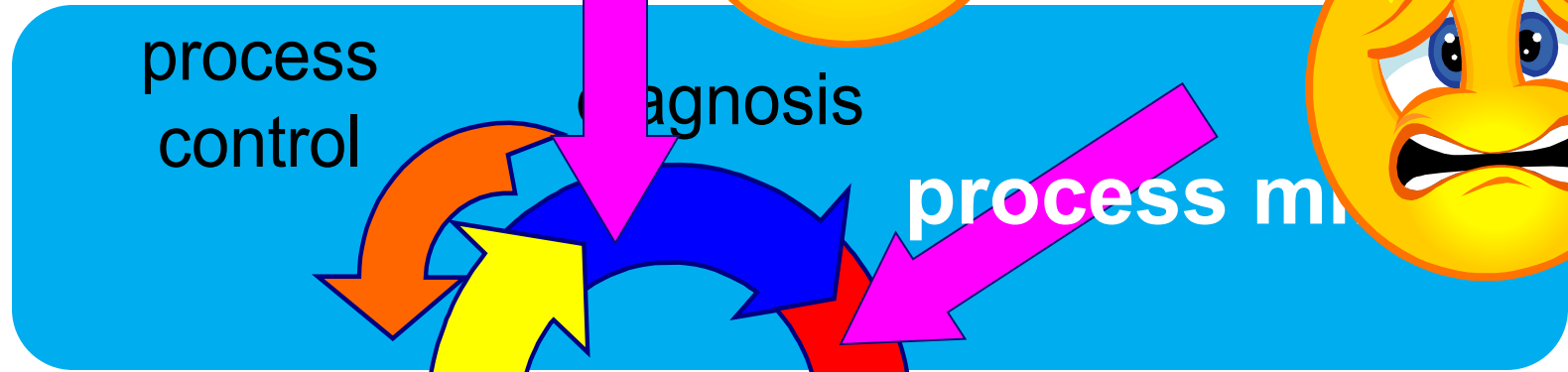
Alignment (Avoiding PowerPoint reality)

Ensuring
compliance



Supporting
flexibility

Where to start?



Process Mining: The Basics

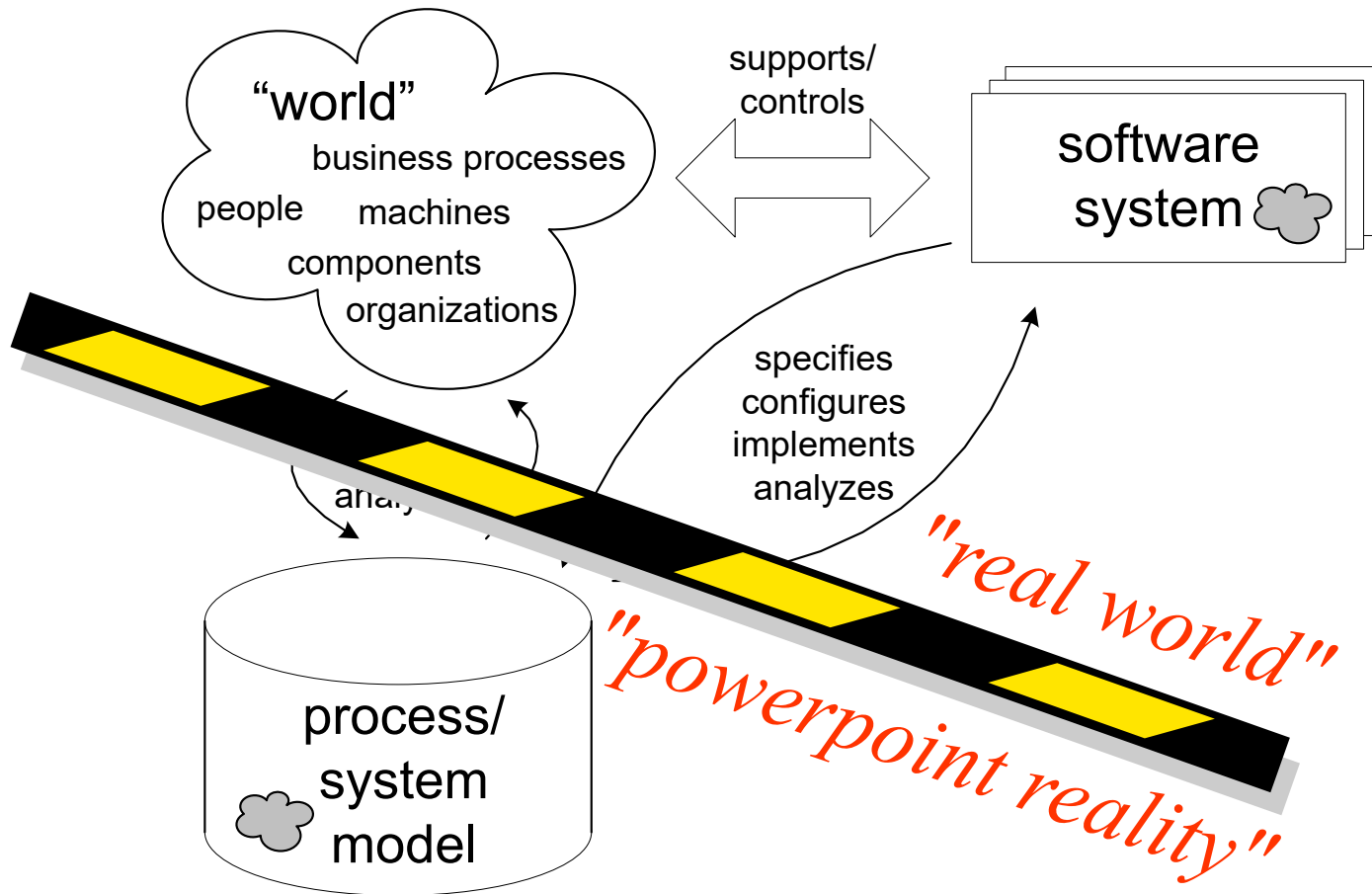


TU / **e**

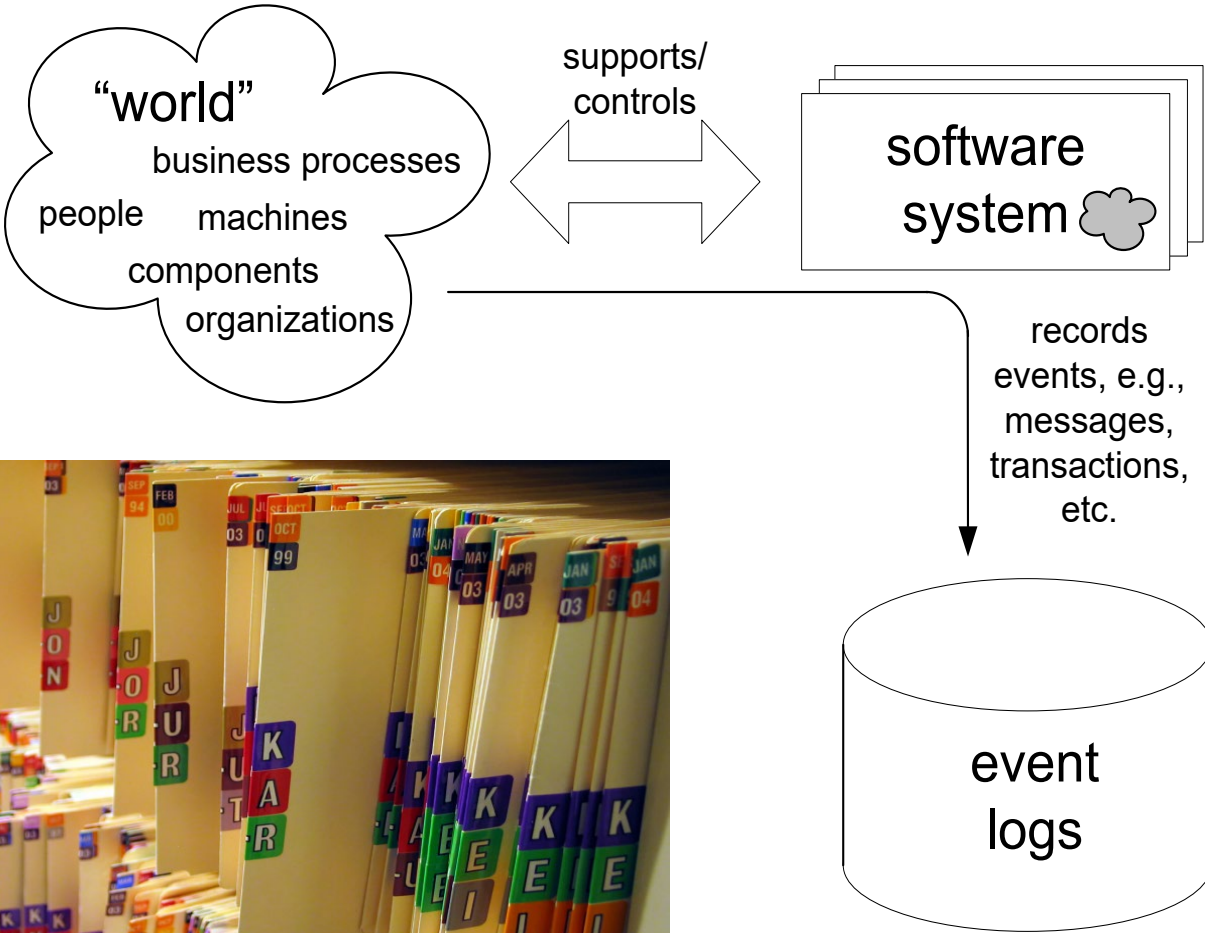
Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Role of models



Event logs are a reflection of reality



Examples:



IBM

staffware.
THE POWER OF PROCESS

FILENET
An IBM® Company

océ

Pallas Athena

PHILIPS
sense and simplicity

SAP

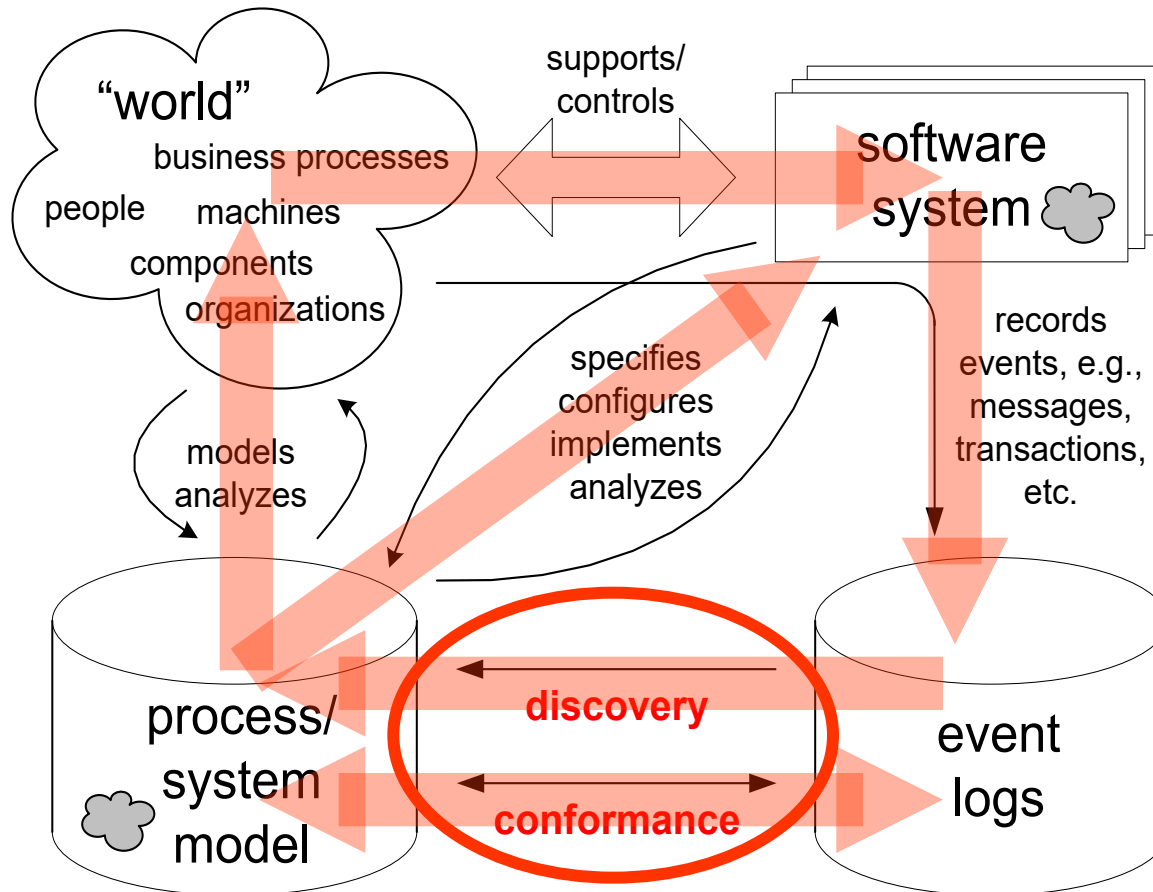
ORACLE

ASML

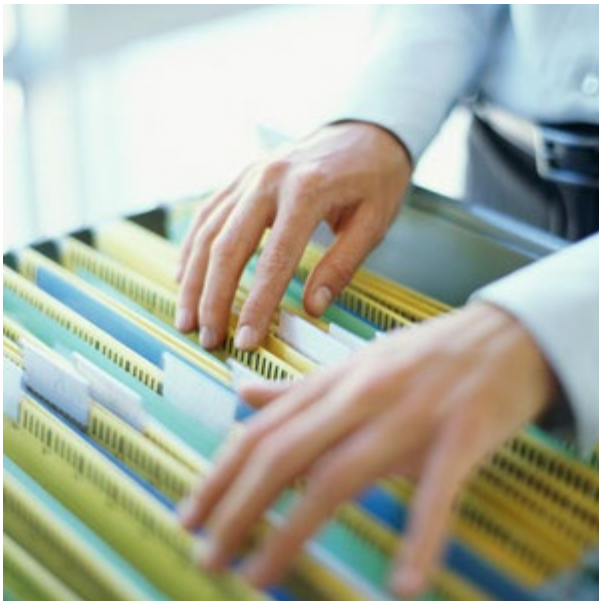
WebSphere® software



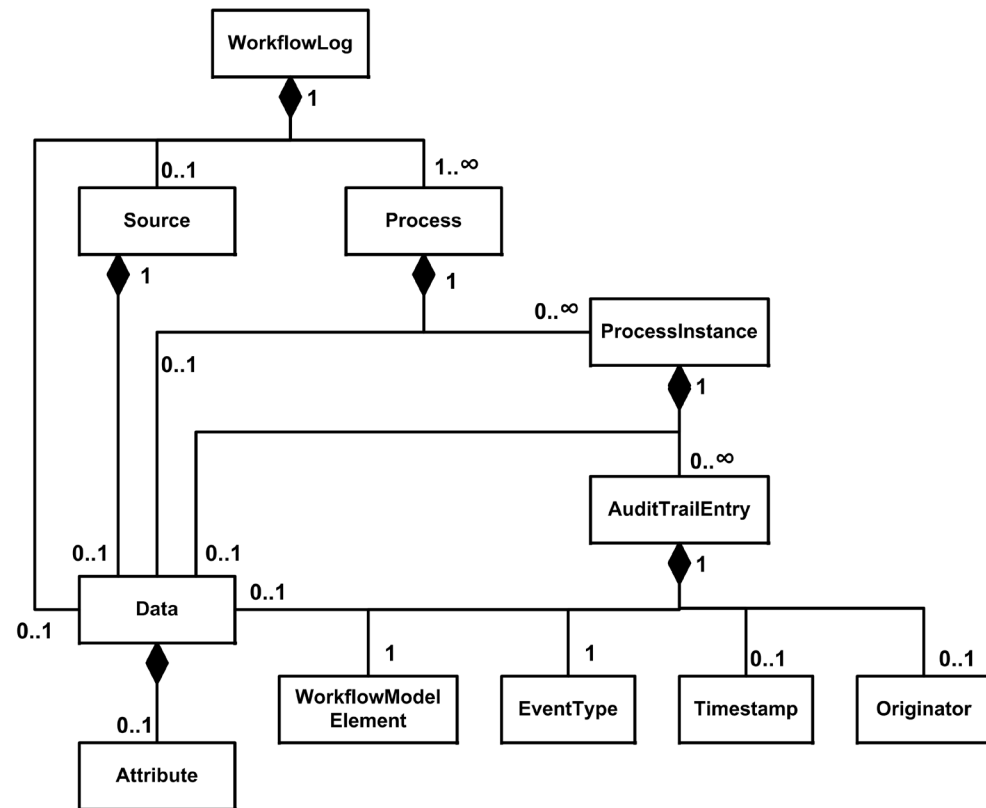
Process mining: Linking events to models



Starting point: event logs

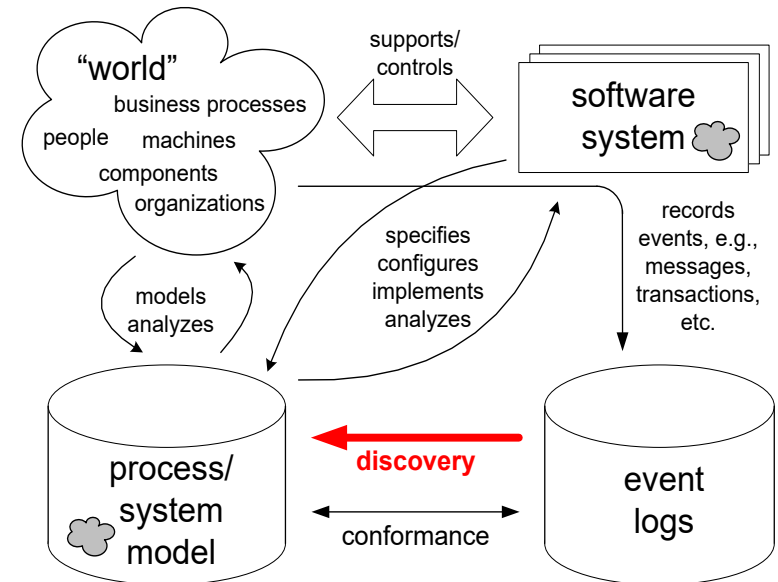


event logs, audit trails, databases, message logs, etc.



unified event log
(MXML)

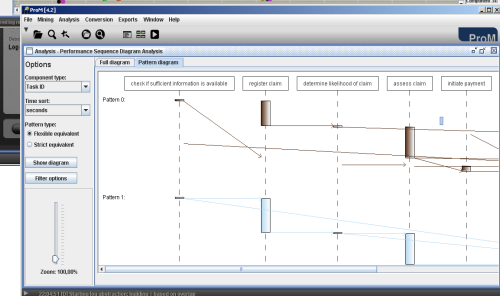
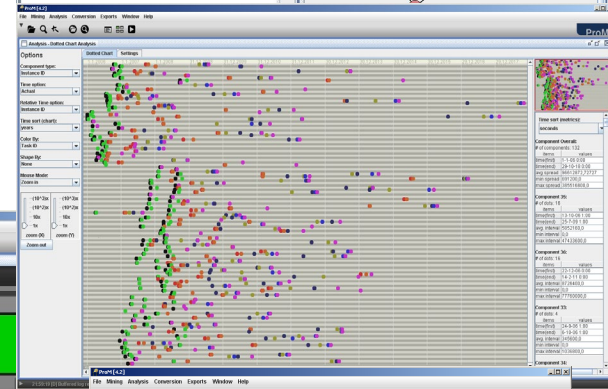
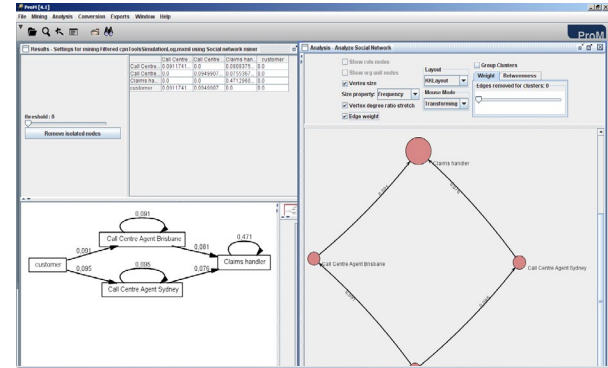
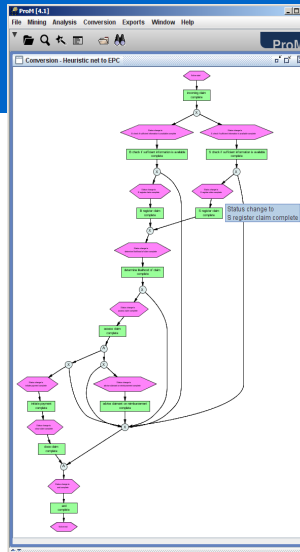
Discovery



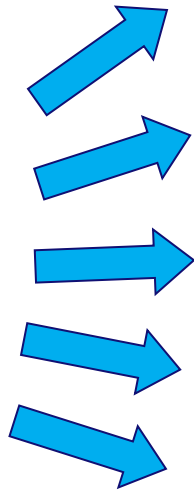
What to discover?

- process models (Petri nets, EPCs, BPMN, etc.),
- organizational models,
- social networks,
- sequence diagrams,
- business rules,
- bottlenecks,
- simulation models,
- etc.

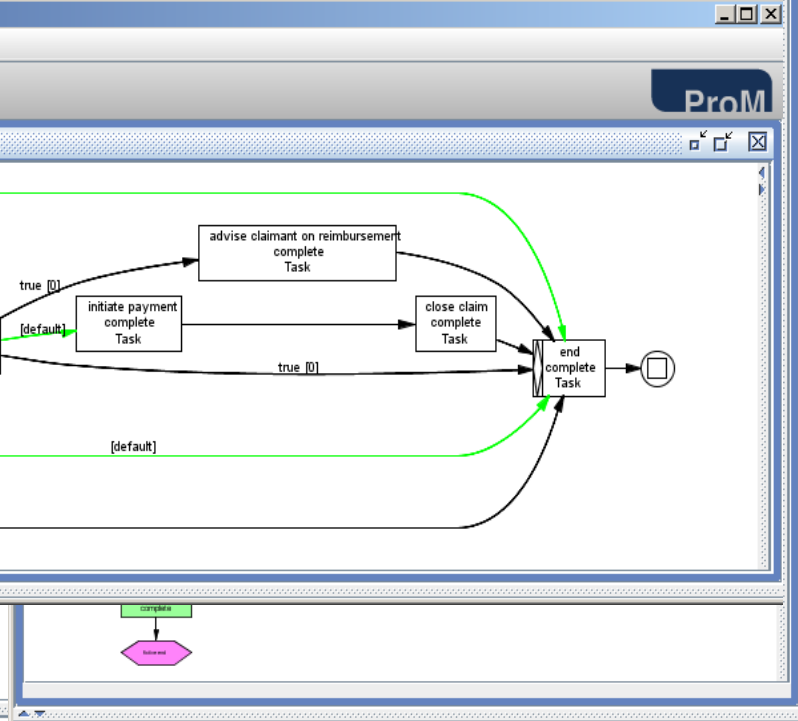
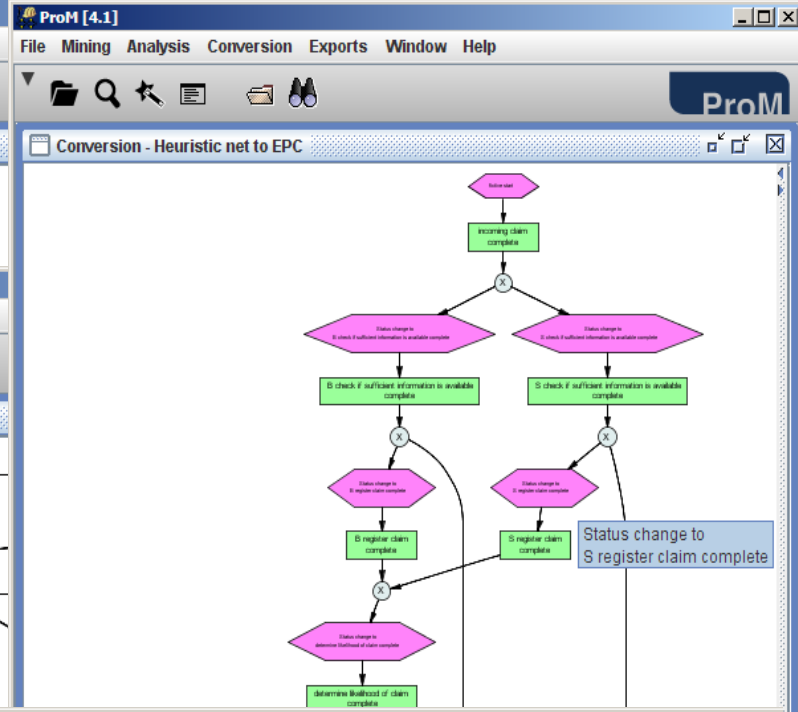
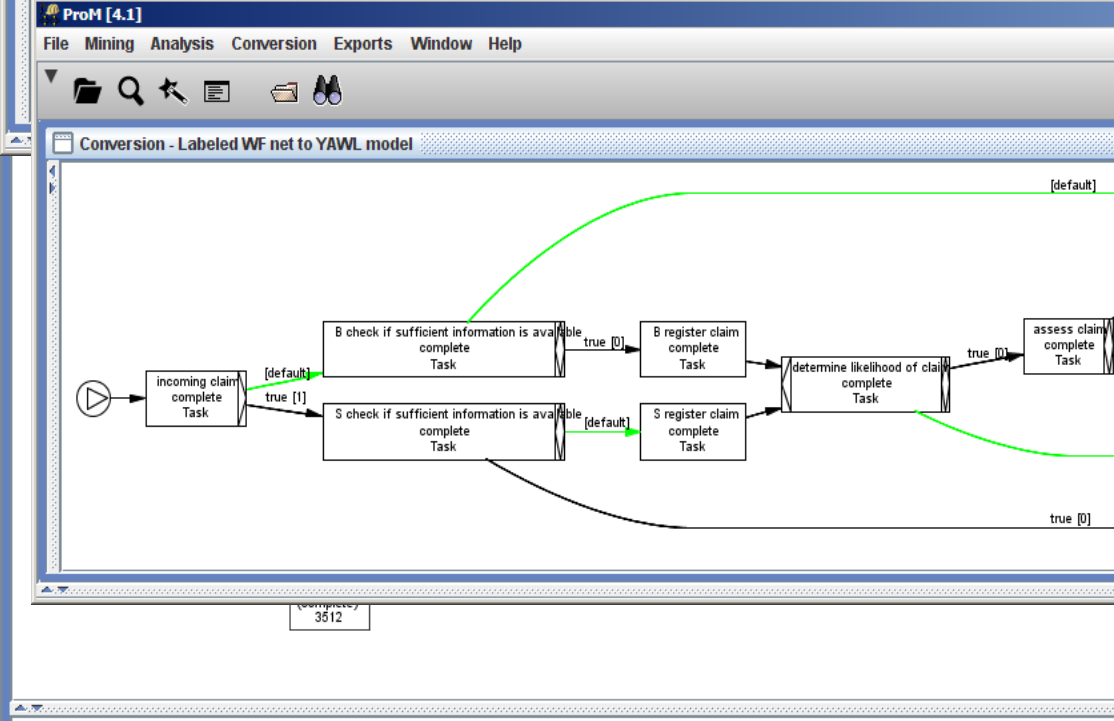
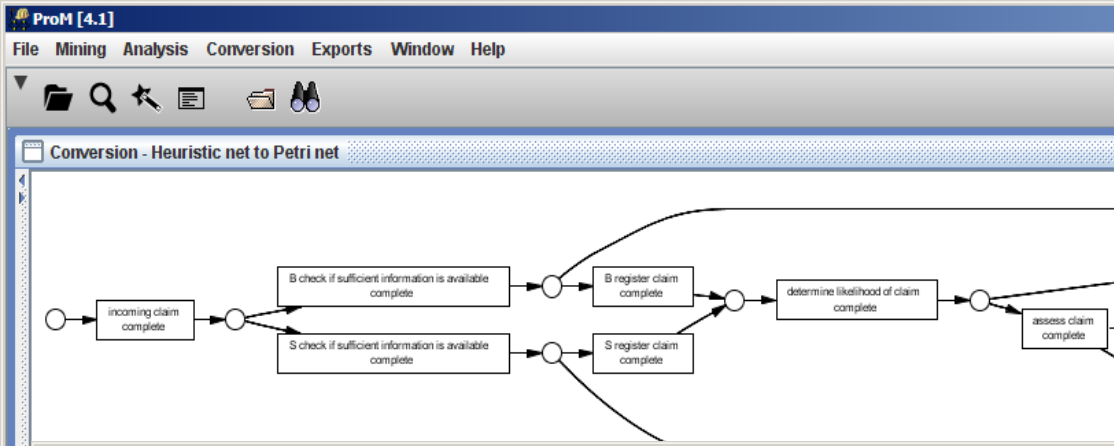
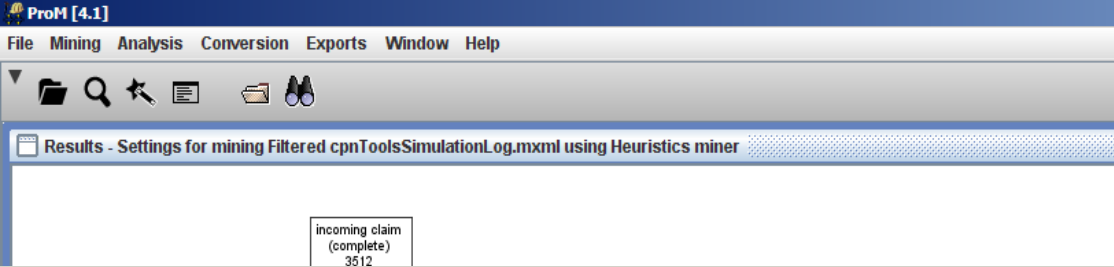
i.e., beyond "slice and dice" and showing KPIs on a dashboard ...



MXML Log
- instances: 3512
- audit trail entries: 46138



ProM supports +40 types of model discovery!



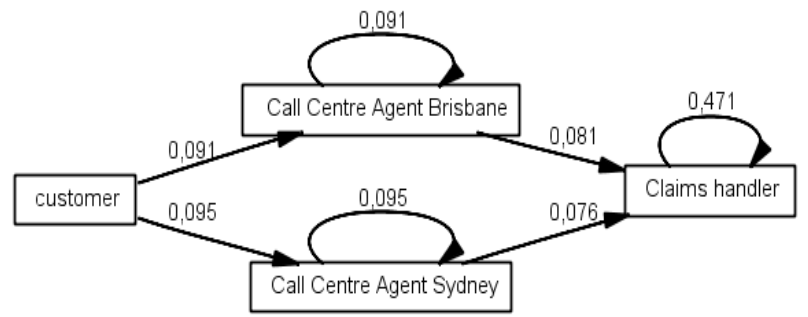


Results - Settings for mining Filtered cpnToolsSimulationLog.mxml using Social network miner

	Call Centre ...	Call Centre ...	Claims han...	customer
Call Centre...	0.0911741...	0.0	0.0808375...	0.0
Call Centre...	0.0	0.0949907...	0.0755367...	0.0
Claims ha...	0.0	0.0	0.4712960...	0.0
customer	0.0911741...	0.0949907...	0.0	0.0

threshold : 0

Remove isolated nodes

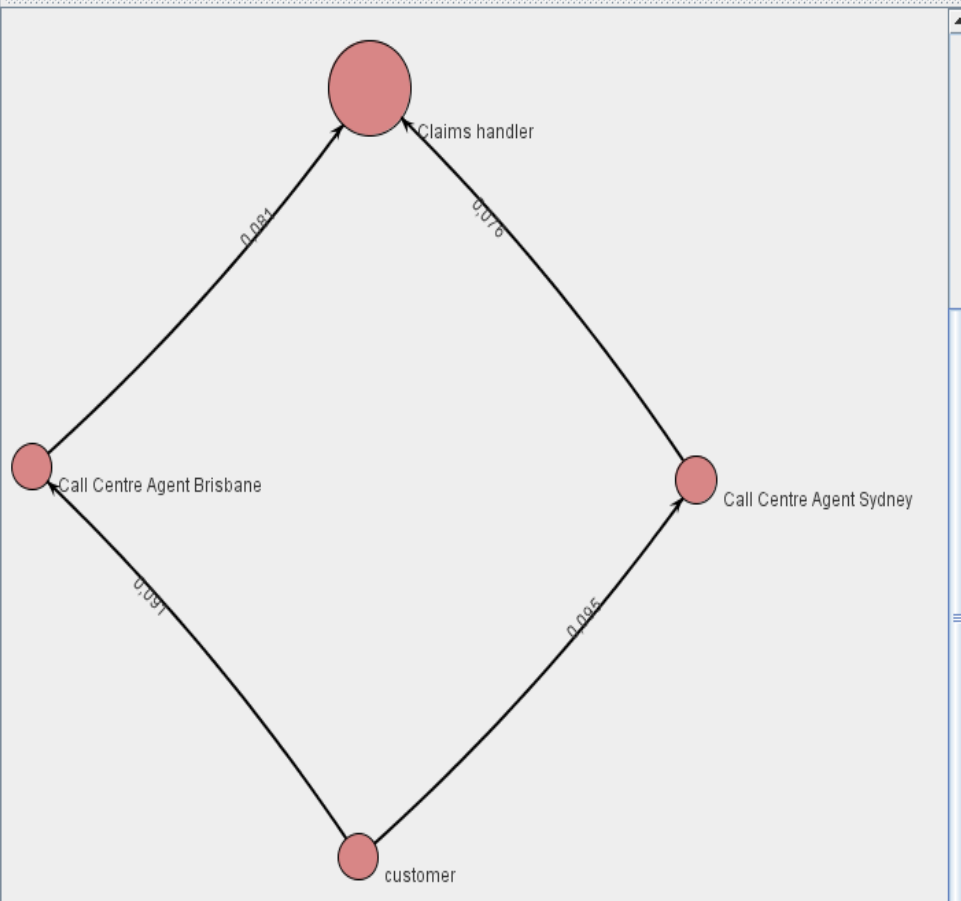


Analysis - Analyze Social Network

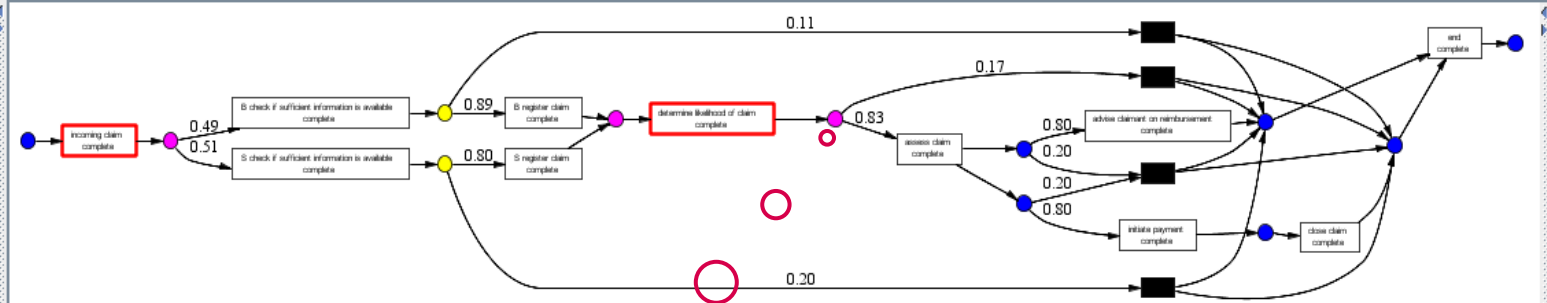
Show role nodes
 Show org unit nodes
 Vertex size
Size property: Frequency
 Vertex degree ratio stretch
 Edge weight

Layout: KKLLayout
Mouse Mode: Transforming

Group Clusters
Weight: Betweenness
Edges removed for clusters: 0



Analysis - Performance Analysis with Petri net



bottle-necks

flow time from A to B

throughput time

Process information:

Total number selected:
3512 cases
Number fitting:
3512 cases
Arrival rate:
0,12 cases per second

	Throughput time (seconds)
avg	11115,54
min	0,0
max	40704,0
stdev	8906,98
fast 25...	1379,19
slow 2...	23817,24
norma...	9632,87

Change Percentages Export Time-Metrics

Performance information of the selected transitions:

Frequency: 2950 cases

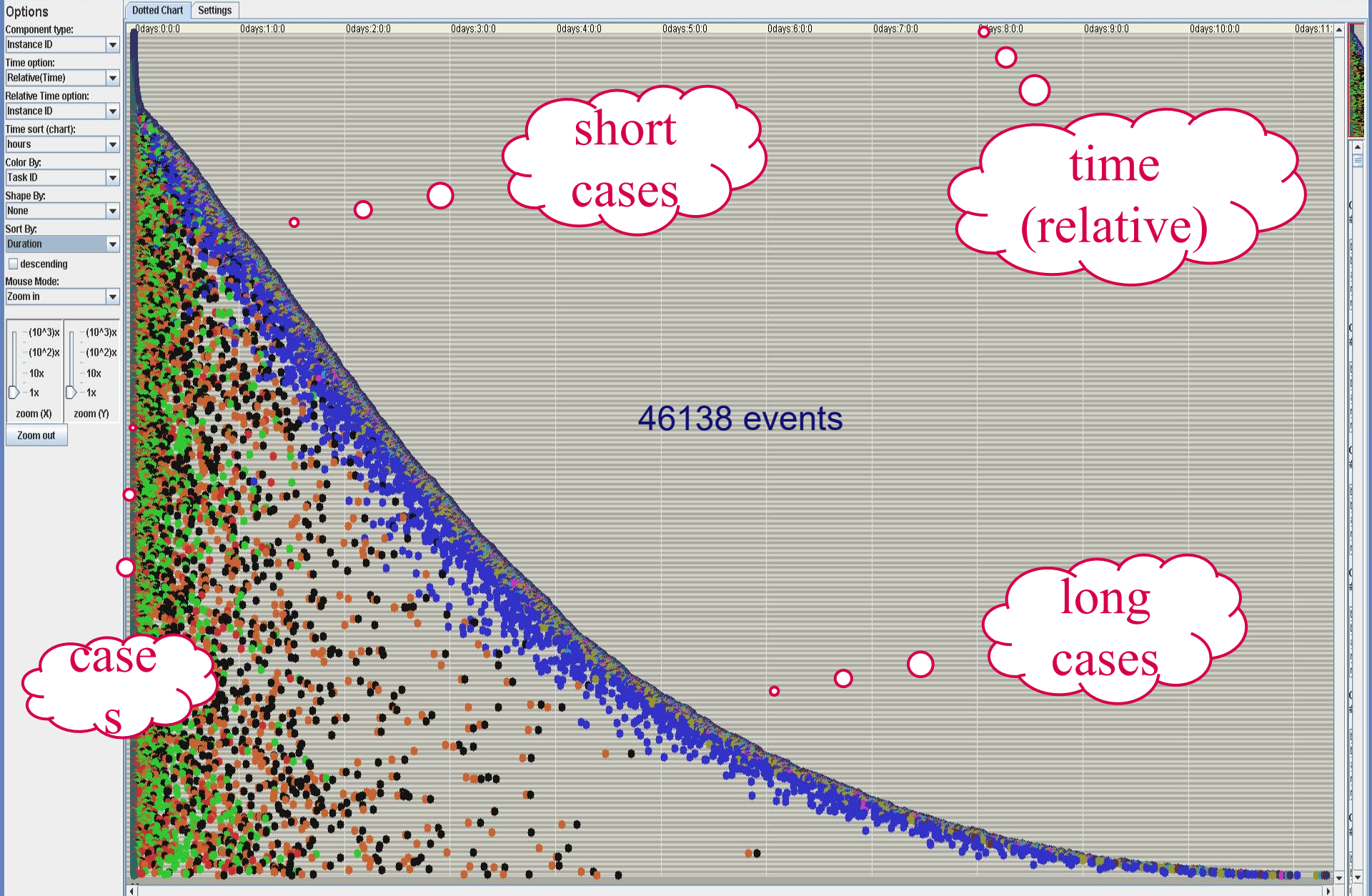
	Time in between (seconds)
avg	12248,87
min	53,0
max	39706,0
stdev	8381,14

Waiting time:

- High
- Medium
- Low

Settings

Selected:
Transition - incoming claim c...
and:
Transition - determine likeliho...



short cases

time (relative)

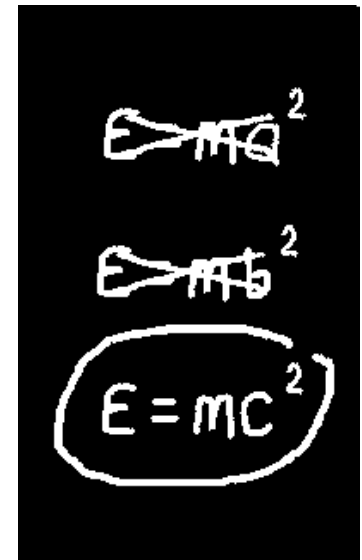
46138 events

cases

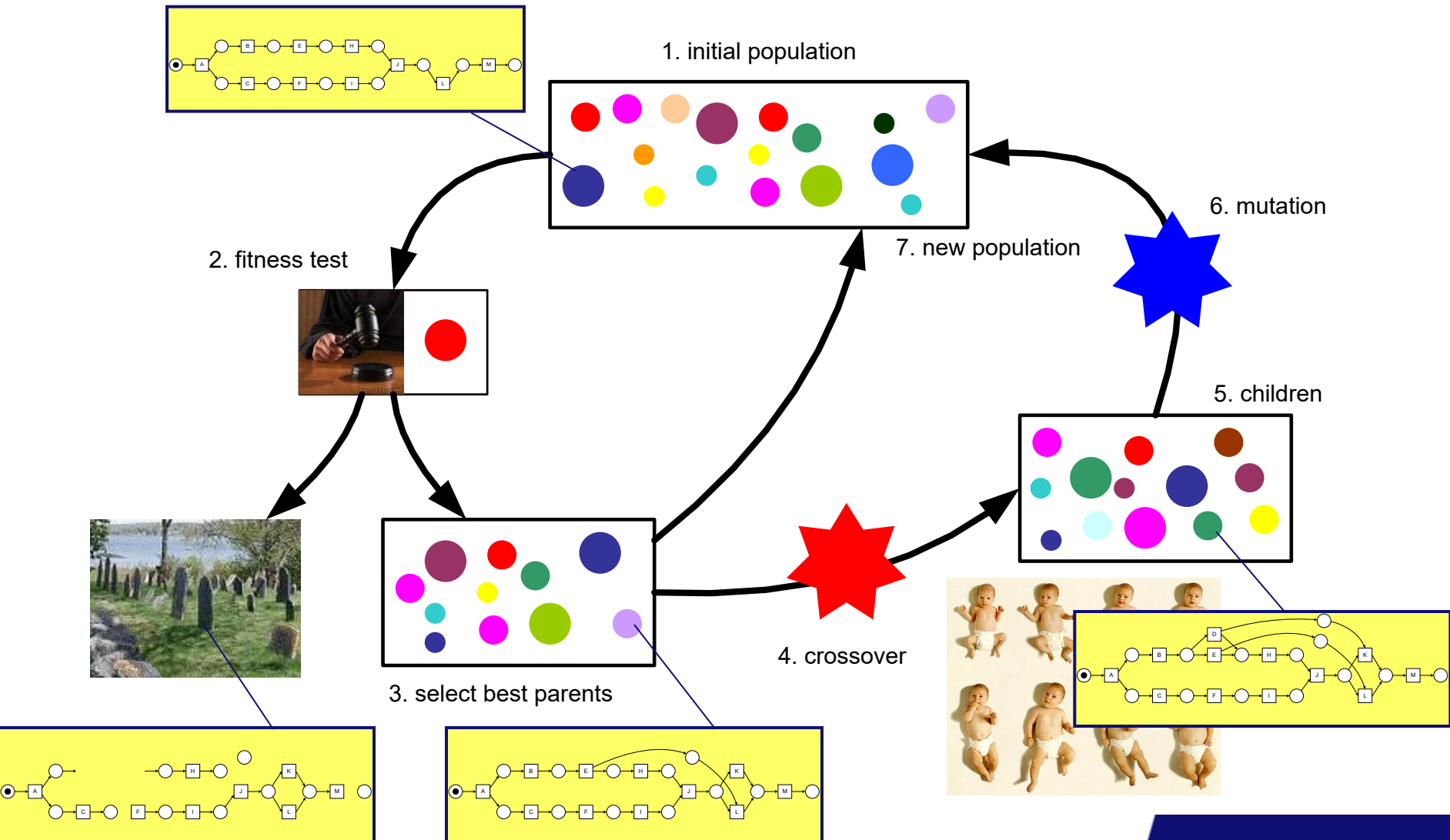
long cases

A bit of theory: Process discovery techniques

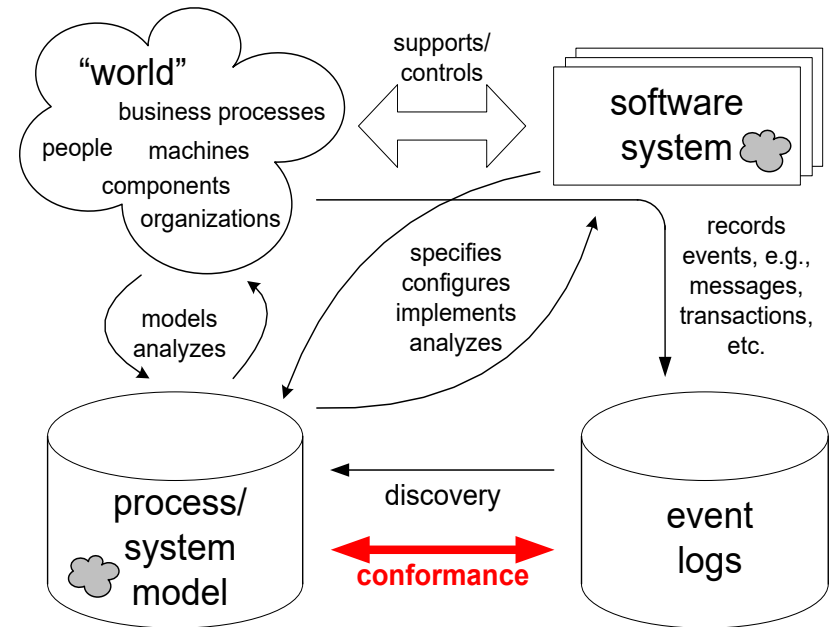
- **Algorithmic techniques**
 - Alpha miner
 - Alpha+, Alpha++, Alpha#
 - Heuristic miner
 - Multi phase miner
 - ...
- **Genetic process mining**
- **Region-based process mining**
 - State-based regions
 - Language based regions



Example: Genetic Mining



Conformance Checking



Conformance Checking

- Compare process model and event log: highlight deviations and measure conformance.
- Compare constraints/business rules and event logs: check e.g. the 4-eyes principle.

The screenshot displays the ProM 4.2 software interface for a conformance checker. The main window shows a Petri net diagram representing a process model. The diagram includes several transitions (circles) and places (rectangles). Transitions are labeled with numbers: 132, 115, 99, 76, and 191. Places are labeled with text: "check if sufficient information is available complete", "register claim complete", "determine threshold of claim complete", "advise claimant on reimbursement complete", "assess claim complete", "initiate payment complete", "close claim complete", and "end complete". There are also several red circles with numbers inside, representing deviations or fitness values: -115, +10, +23, -130, +175, and +191. The diagram is connected to a log of traces on the left side, which lists trace numbers from 1 to 149. The right side of the interface shows a "Model-related Measure" section with a "Fitness" value of 0.7635551. At the bottom, there are several checkboxes for diagnostic perspectives: "Model", "Token Counter", "Failed Tasks", "Remaining Tasks", "Path Coverage", and "Passed Edges". The status bar at the very bottom shows the time "08:50:27 [D] DOT finished on: D:\TEMP\pmt48873.dot".

Tool support



- **Open source initiative started in 2003 after several early prototypes.**
- **Common Public License (CPL).**
- **Current version: 5.0.**
- **ProMimport: to extract MXML from all kinds of applications**
- **Plug-in architecture.**
- **About 250 plug-ins available:**
 - **mining plug-ins: 38 (all mining algorithms presented and many more)**
 - **analysis plug-ins: 71 (e.g., verification, SNA, LTL, conformance checking, etc.)**
 - **import: 21 (for loading EPCs, Petri nets, YAWL, BPMN, etc.)**
 - **export: 44 (for storing EPCs, Petri nets, YAWL, BPMN, BPEL, etc.)**
 - **conversion: 45 (e.g., translating EPCs or BPMN into Petri nets)**
 - **filter: 24 (e.g., removing infrequent activities)**

Screenshot of ProM 5.0



reviewslog_with_fewer_errors.xml

Key data

- Processes: 1
- Cases: 100
- Events: 2297
- Event classes: 20
- Event types: 2
- Originators: 10

Events per case

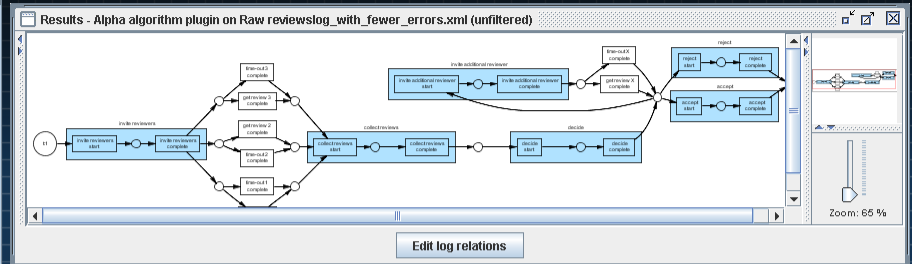
Min 11 Mean 22 Max 50

Event classes per case

Log info

Source: CPN Tools
Source program: CPN Tools
Start date: 2006-01-01 00:00:00
End date: 2008-05-05 01:00:00
Description: Log extracted from CPN Tools

start analyzing this log



Fuzzy Model Animation

Activity: 100 %

Completion: 3 %

case progress

22:33:06 Sun 5 Feb '06

playback speed zoom view

Analysis - Dotted Chart Analysis

Options

- Component type: Instance ID
- Time option: Actual
- Relative Time option: Instance ID
- Time sort (chart): months
- Color By: Task ID
- Shape By: None
- Sort By:

Time sort (metrics): seconds

Component Overall:

- # of components: 100
- Items: values
- time(first): 1/1/06 12:00 AM
- time(end): 5/5/08 1:00 AM
- avg spread: 24662890.0
- min spread: 1036800.0
- max spread: 64281600.0

Component 85:

- # of dots: 35
- Items: values
- time(first): 1/28/07 12:00 AM
- time(end): 1/17/08 12:00 AM

Analysis - Basic Performance Analysis

Basic Performance Analysis

Measure: processing (task) Chart: Meter Chart

1st Dim:	task	accept	collect reviews	decide
2nd Dim:	task			
Time Unit:	hours			
Measure:	Average			
Performance Sort:	Working Time			
Minimum	33.51724137931034			
Upper value (normal)	85.7705329153605			
Upper value (warning)	103.18829676071056			

Update

Business Intelligence Tools?

- **Business Objects (SAP)**
- **Cognos Business Intelligence (IBM)**
- **Oracle Business Intelligence**
- **Hyperion (Oracle)**
- **SAS Business Intelligence**
- **Microsoft Business Intelligence**
- **SAP Business Intelligence (SAP BI)**
- **Jaspersoft (Open Source Business Intelligence)**
- **Pentaho BI Suite (Open Source)**
-



- Dashboards, reports, scorecards, ...
- Slicing and dicing, data mining, ...

Process Mining Software



Futura Reflect



BPM|one



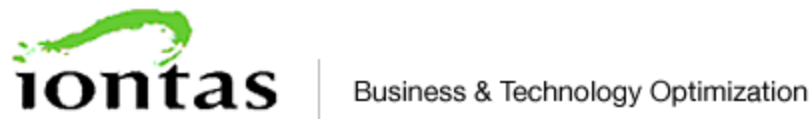
Comprehend



ARIS Process Performance Manager



Interstage Automated Business Process Discovery & Visualization



Process Discovery Focus



Enterprise Visualization Suite

Process Mining: Applications



TU / **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

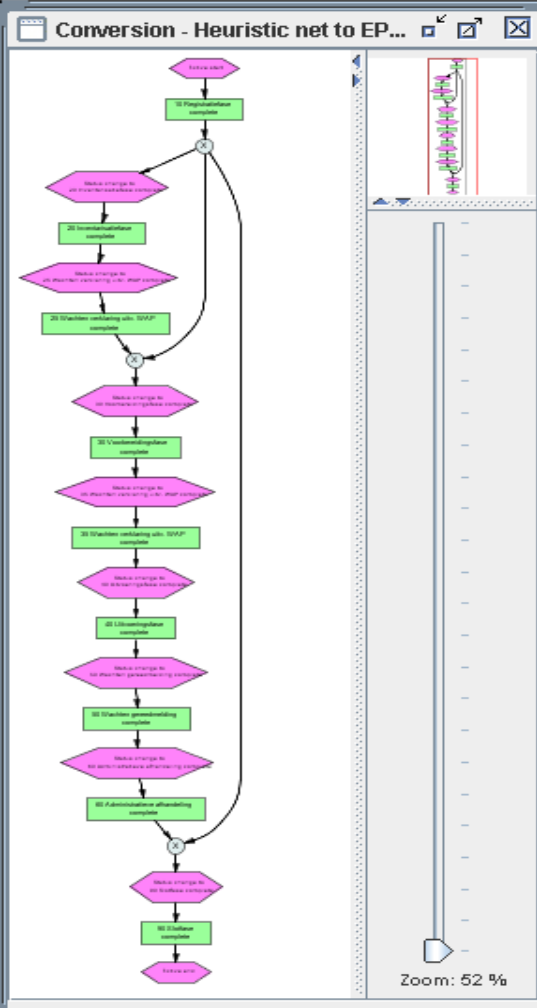
Where did we apply process mining?

- **Municipalities (e.g., Alkmaar, Heusden, Harderwijk, etc.)**
- **Government agencies (e.g., Rijkswaterstaat, Centraal Justitieel Incasso Bureau, Justice department)**
- **Insurance related agencies (e.g., UWV)**
- **Banks (e.g., ING Bank)**
- **Hospitals (e.g., AMC hospital, Catharina hospital)**
- **Multinationals (e.g., DSM, Deloitte)**
- **High-tech system manufacturers and their customers (e.g., Philips Healthcare, ASML, Thales)**
- **Media companies (e.g. Winkwaves)**
- **...**

Example: A Dutch Municipality



ProM [5.0]
File Mining Analysis Conversion Exports Window Help



Results - Fuzzy Miner on Filtered Protos datamining WV proces.mxml.gz (Simple fil

Fuzzy Model Toolkit

Transformer Editor Unary Metrics Binary Metrics Animation

144 cases
1326 events

Conformance check of discovered model

The screenshot displays the ProM 5.0 interface for the Conformance Checker (4). The main window shows a Petri net diagram with the following elements:

- Transitions:** Two red circles labeled $+4$ and $+1 -1$.
- Places:** A green circle labeled 1011 .
- Activities:** A yellow box labeled "20 Inventarisatiefas complete", a green box labeled "25 Wachten verdeling uitv. WAP complete", and a yellow box labeled "30 V...".
- Flow:** Edges with weights 25, 21, 27, 27, and 48.

Annotations in red thought bubbles include:

- "both" (pointing to the $+4$ and $+1 -1$ transitions)
- "performed while not allowed" (pointing to the $+4$ transition)
- "activity is sometimes not performed" (pointing to the 1011 place)
- "good fit 97.9%" (pointing to the fitness value)
- "drill down" (pointing to the log traces list)

On the right, the "Model-related Measures" panel shows:

- Fitness: 0.9793233
- Zoom: 119%

At the bottom, the "Diagnostic Perspective" is set to "Model", and several checkboxes are checked: "Token Counter", "Failed Tasks", "Remaining Tasks", "Path Coverage", and "Passed Edges". Buttons for "Select Fitting", "Invert Selection", and "Update Results" are visible.

Performance analysis

ProM [5.0]

File Mining Analysis Conversion Exports Window Help

Analysis - Performance Analysis with Petri net (2)

Log Traces

10702
12658
13139
10158
10909
11509
10905
12010
11666
12205
10527
10892
12104
11476
12840
11062
10007
12538
10653
12533
11988
10450
10815
10769

Update

Invert Selection

Process information:

Total number selected:
70 cases

Number fitting:
60 cases

Arrival rate:
0.23 cases per day

Throughput time (days)	
avg	84.01
min	0.0
max	258.0
stdev	68.72
fast 25...	0.87
slow 2...	174.53
norma...	80.32

Change Percentages Export Time-Metrics

Performance information of the selected transitions:

Frequency: 20 cases

Time in between (days)	
avg	37.59
min	0.0
max	110.0
stdev	28.13
fast 25.00%...	10.0

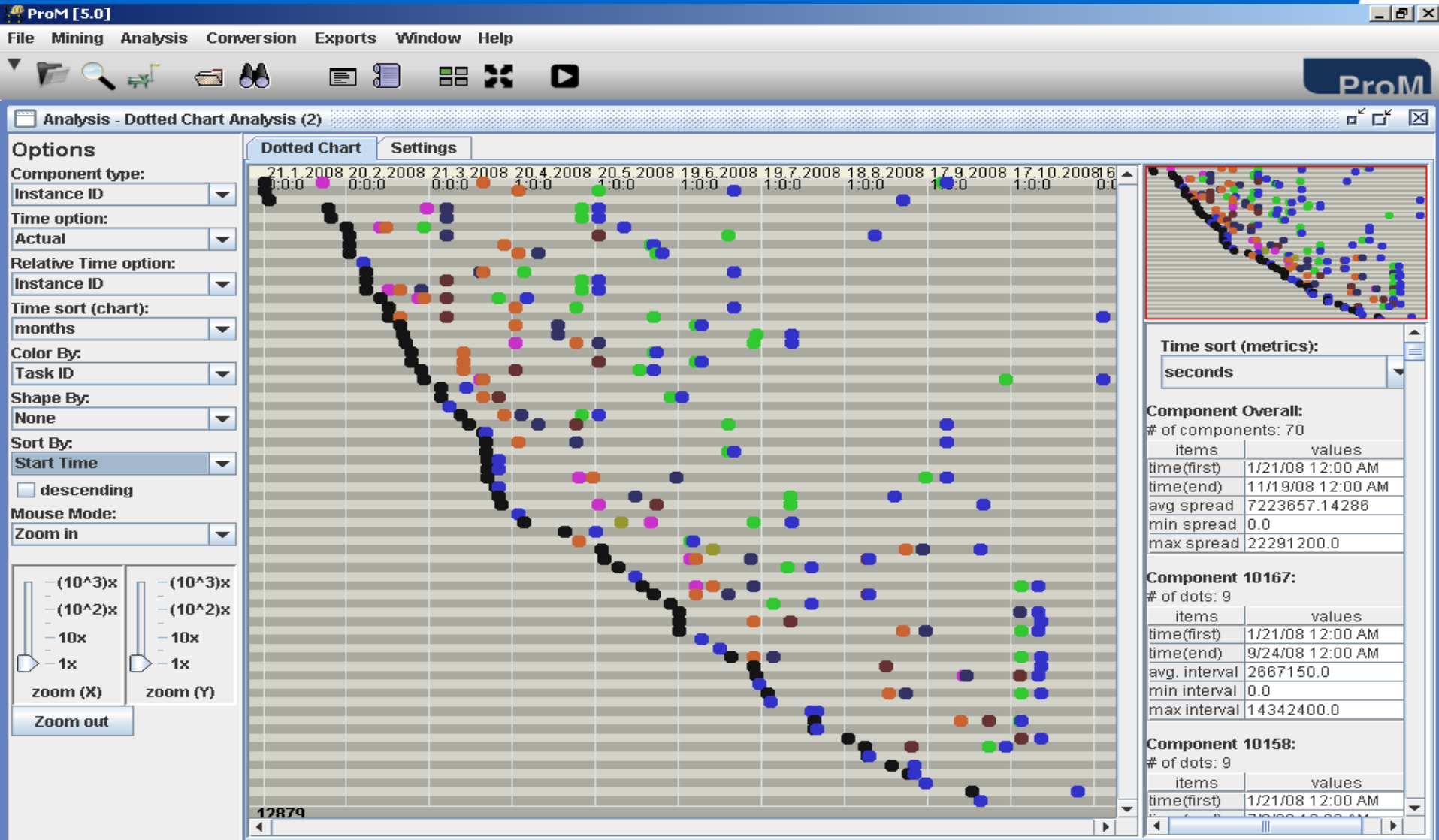
Waiting time:
■ High
■ Medium
■ Low
 Settings

Selected:
 Transition - 10 Registratiefas...
 and:
 Transition - 25 Wachten verkla...

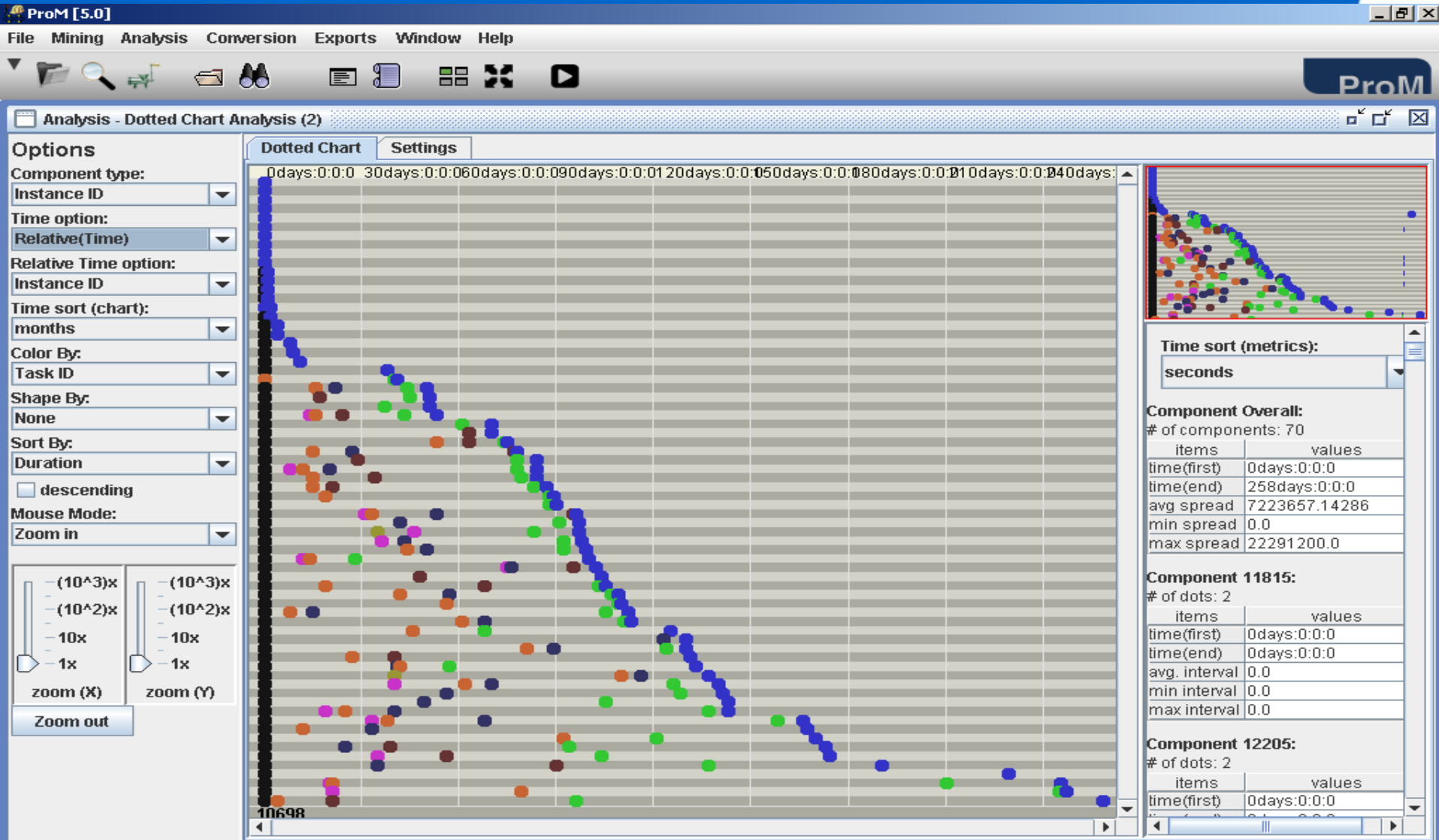
Zoom: 112 %

18:31:19 [D] Buffered log reader created from reader BufferedLogReader: 144 Process Instances and 1326 Audit Trail Entries from "D:\application_data\ProMPallas\WMO-Har...

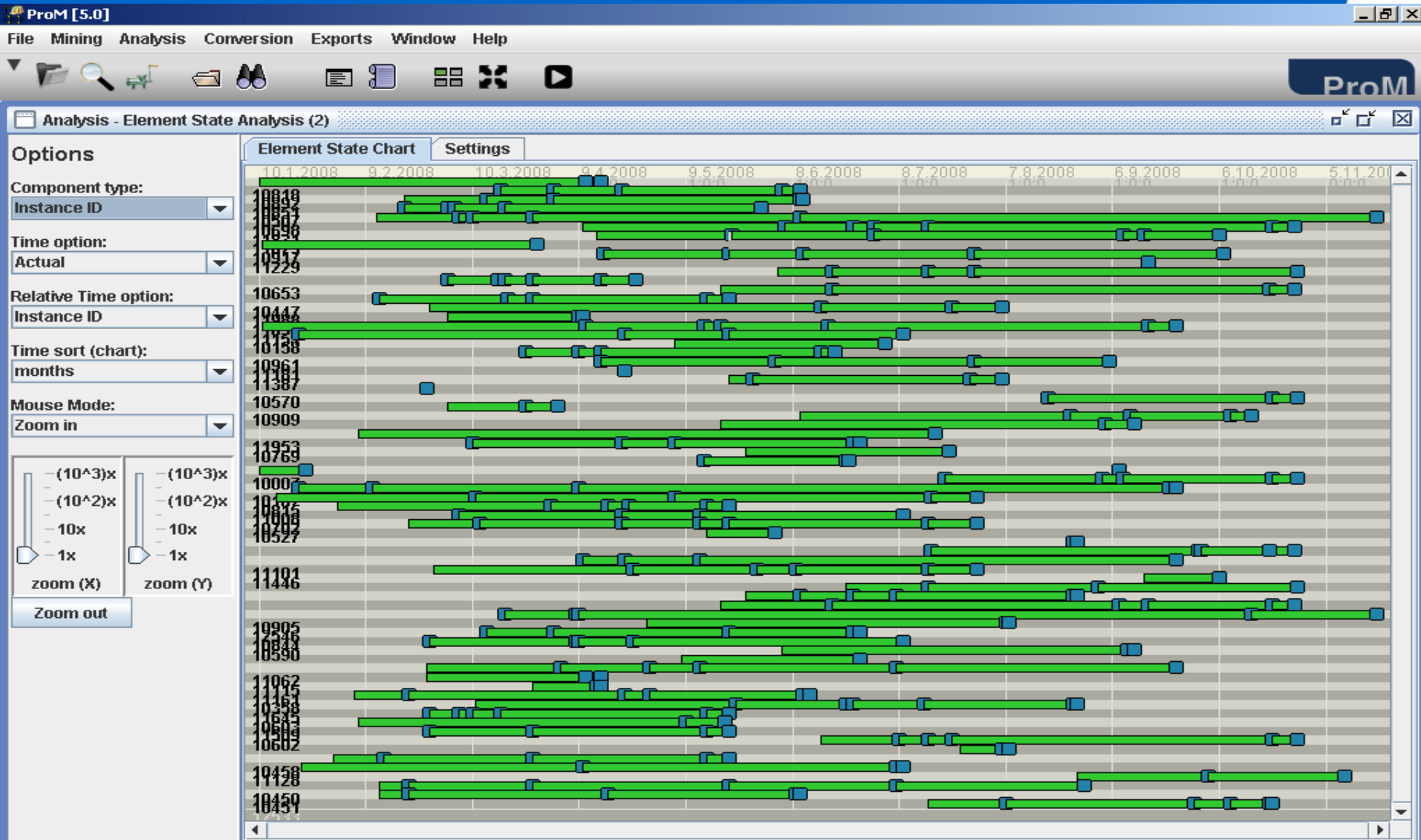
Events sorted by start time of case



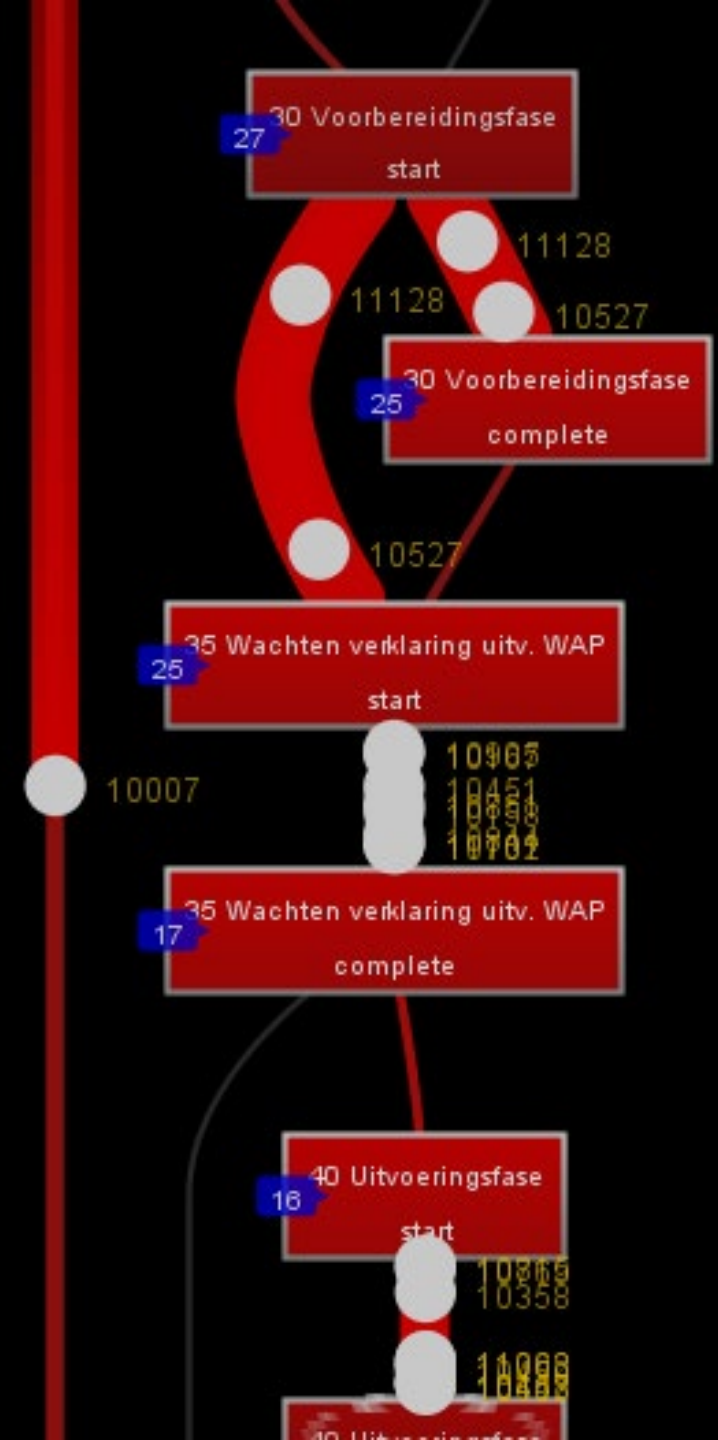
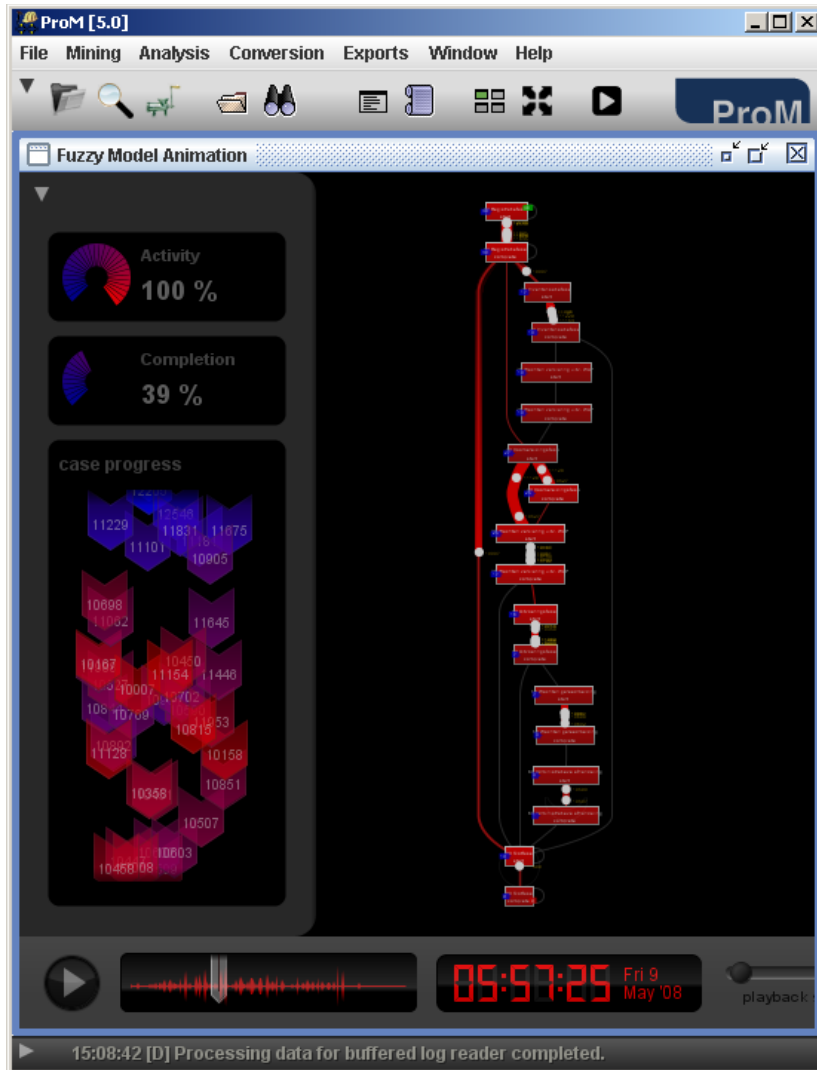
Events sorted by duration



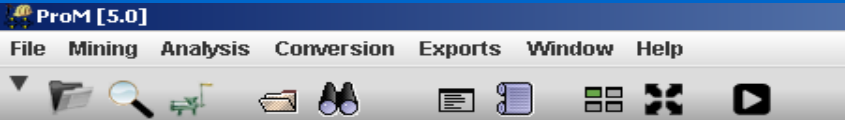
Idle time versus working time



"Real" animation

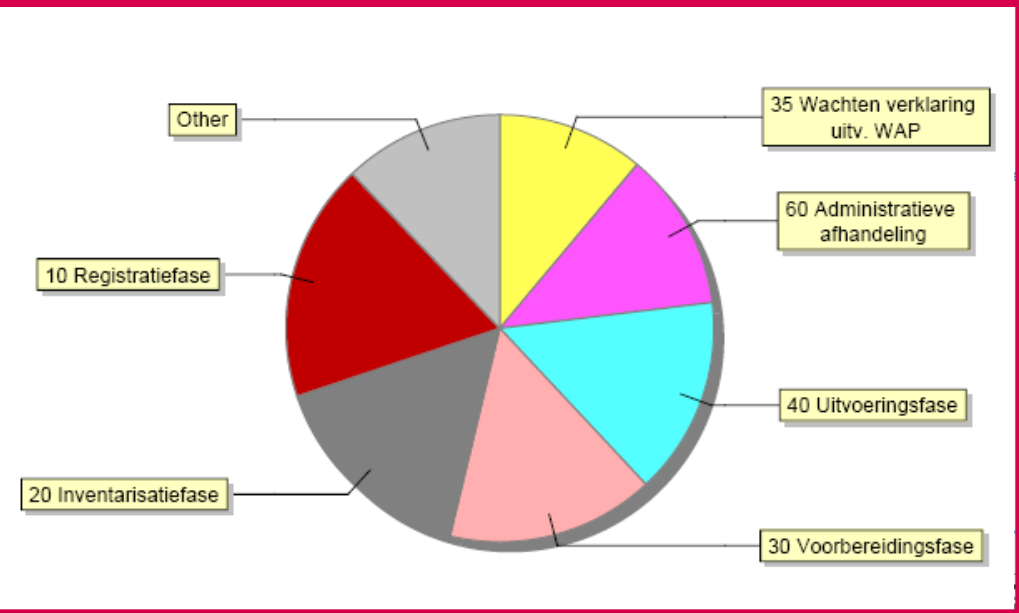
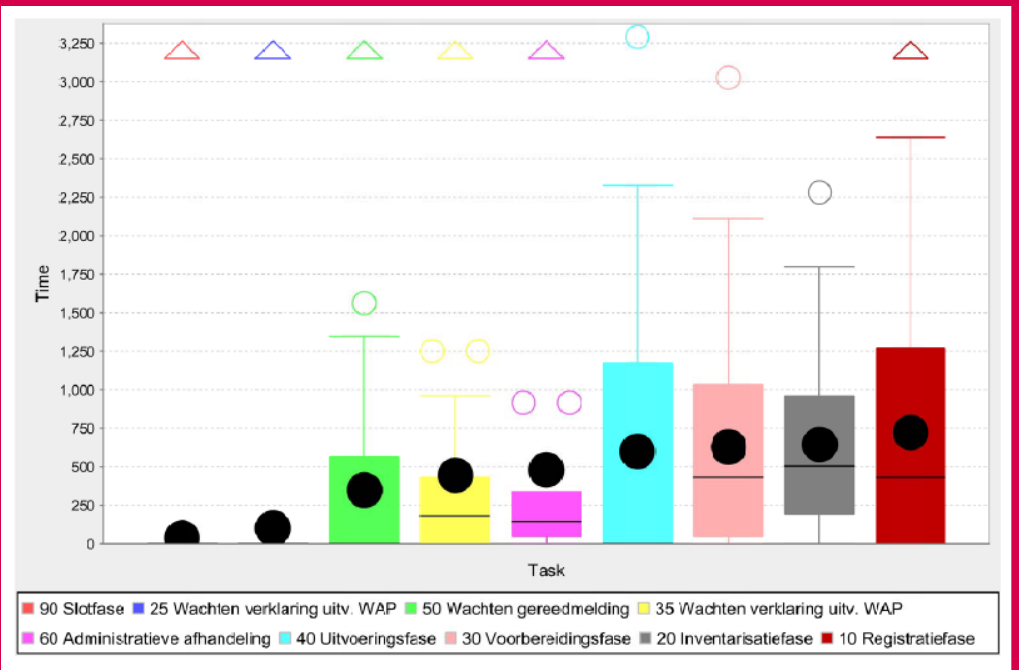
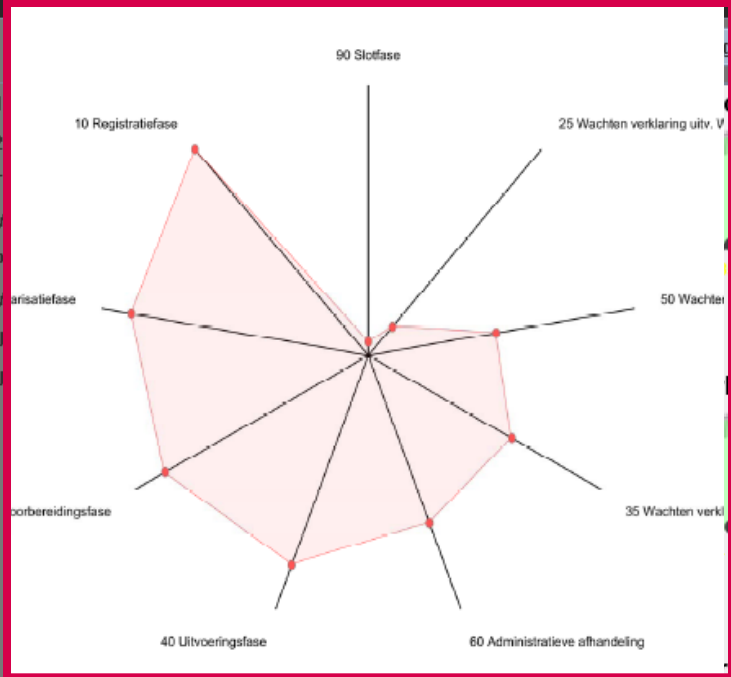


And of course ...



Analysis - Basic Performance Analysis

Basic Performance Analysis



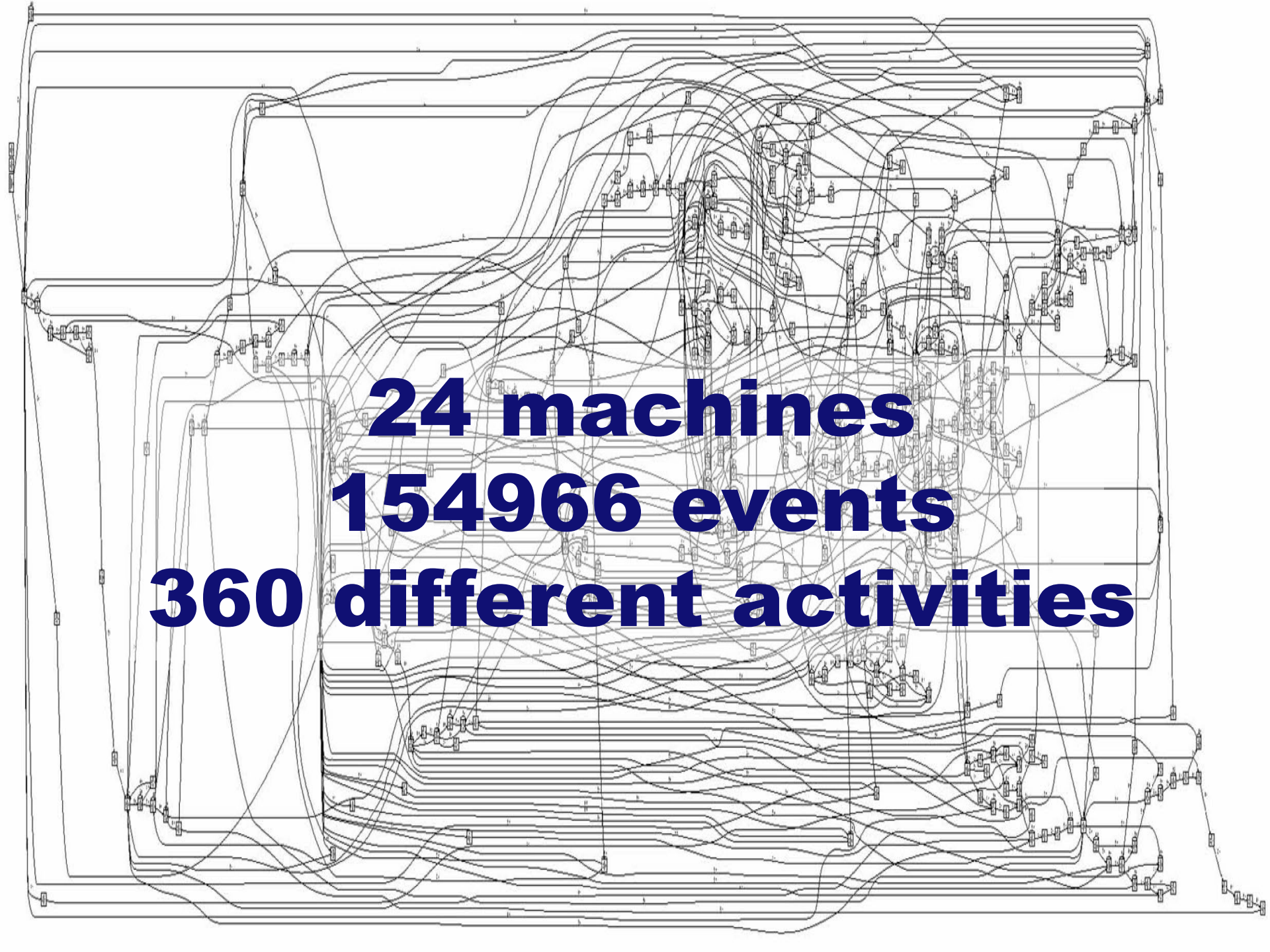
Reality \neq PowerPoint (or Visio)



2712 patients
29258 events
264 different activities



874 patients
10478 events
181 different activities



24 machines
154966 events
360 different activities

#	Log Traces	Fitness	Prec
1	0431		
1	0278		
1	0185		
1	0466		
1	0391		
1	1722		
1	1694		
1	1256		
1	1343		
1	1981		
1	1754		
1	1662		
1	1453		
1	1298		
1	1876		
1	1656		
1	1099		
1	1919		
1	1348		
1	1596		
1	1164		
1	1032		
1	1794		
1	1160		

100

a complete

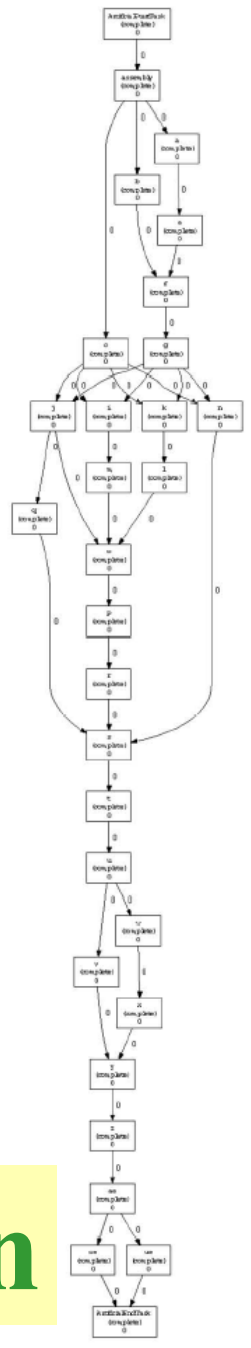
92

89

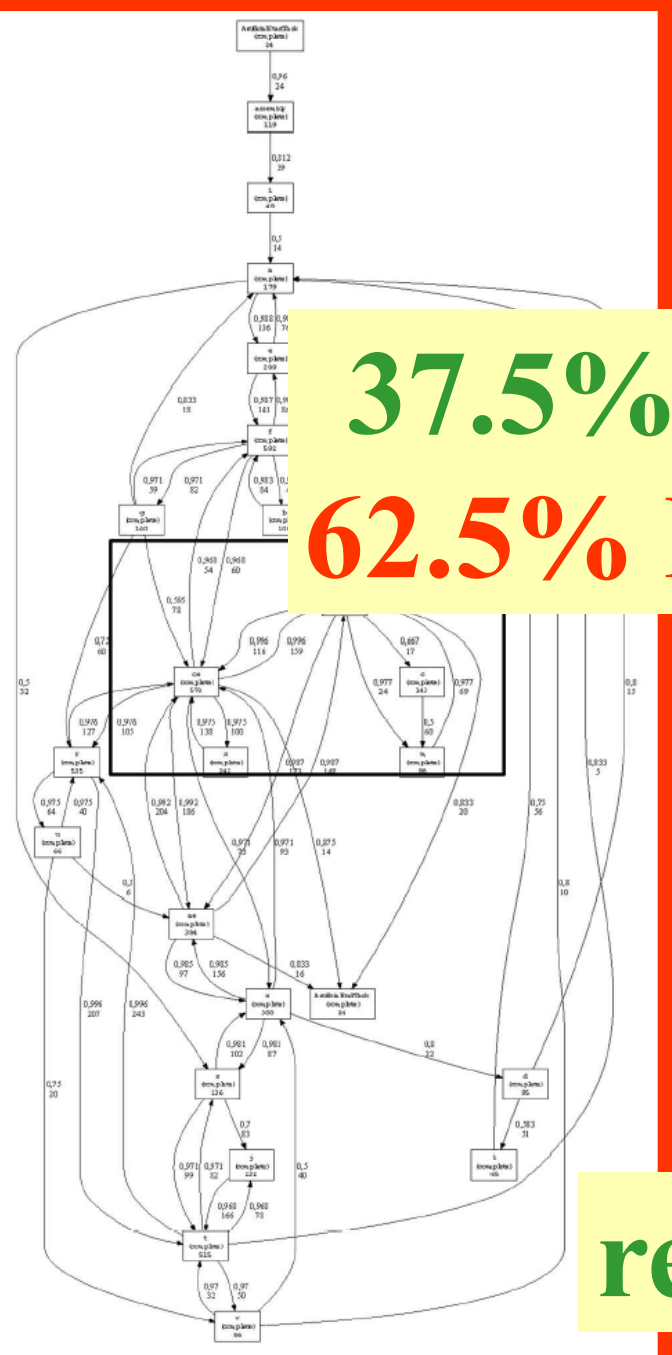
e complete

instances

Diagnostic Perspective



design



37.5% OK
62.5% NOK

Manipulated Measures
Fitness:
0.37501124

with Coverage Passed Edges

100 Update Results

reality

Process Mining: TomTom for Business Processes

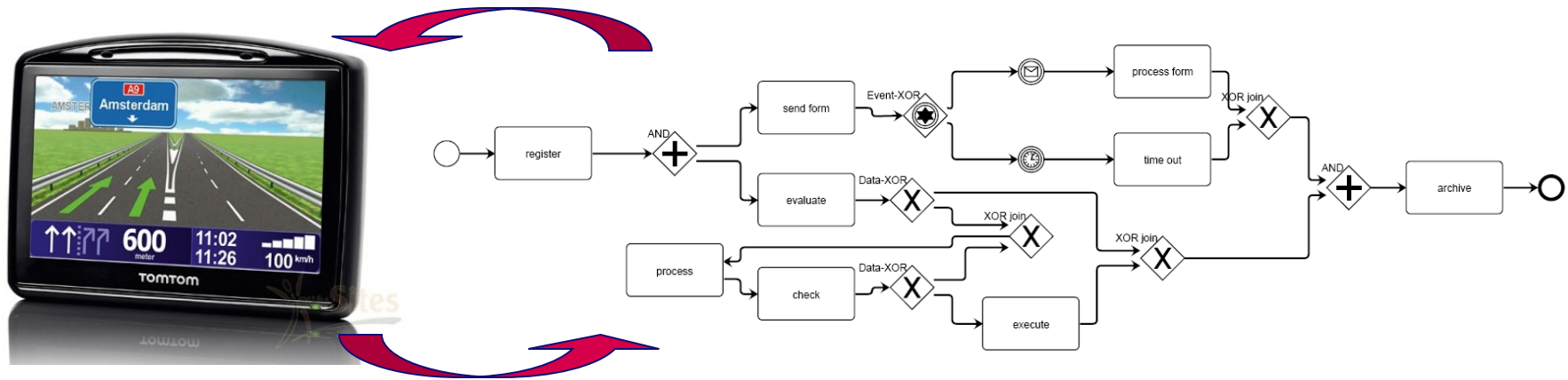


TU / **e**

Technische Universiteit
Eindhoven
University of Technology

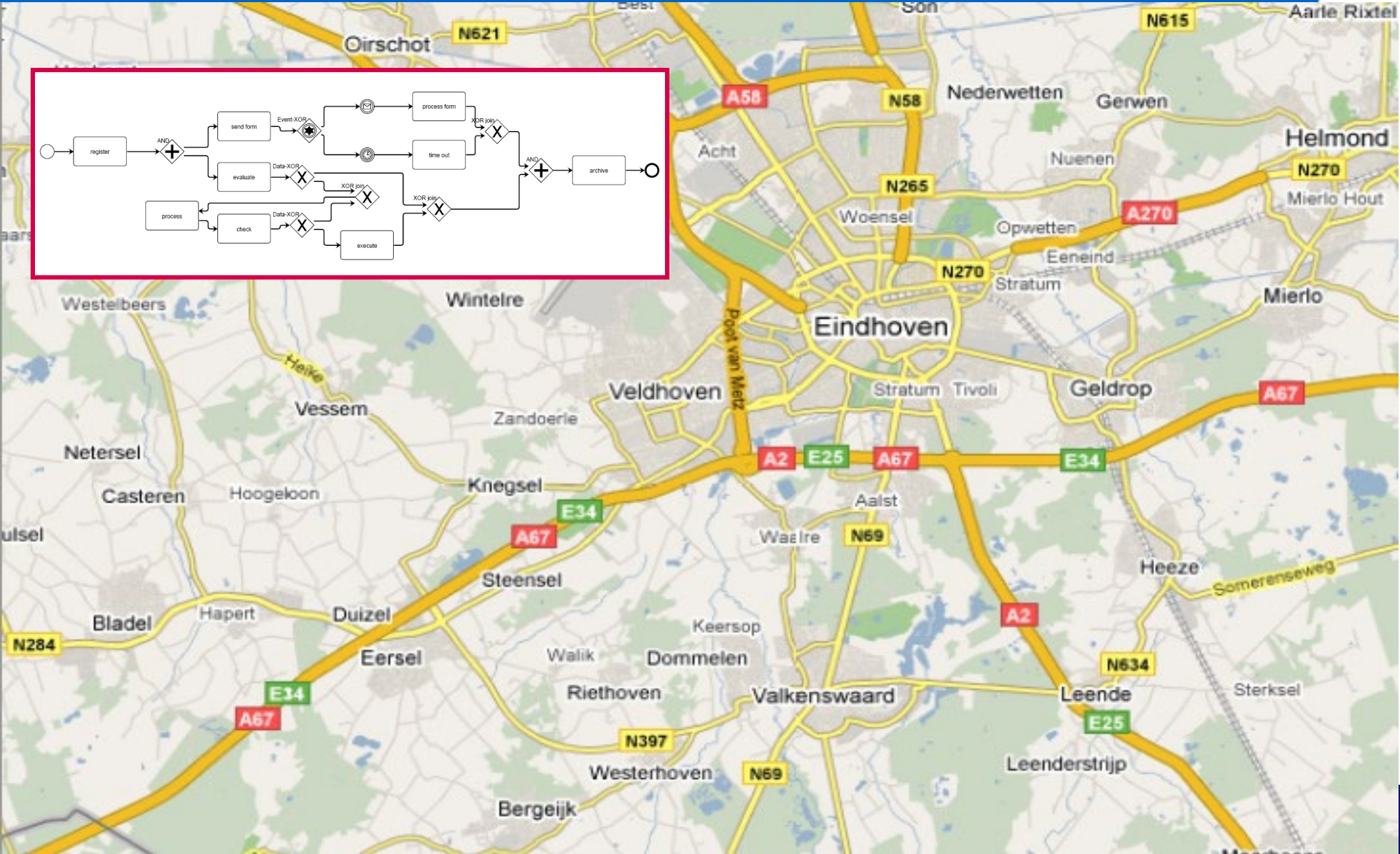
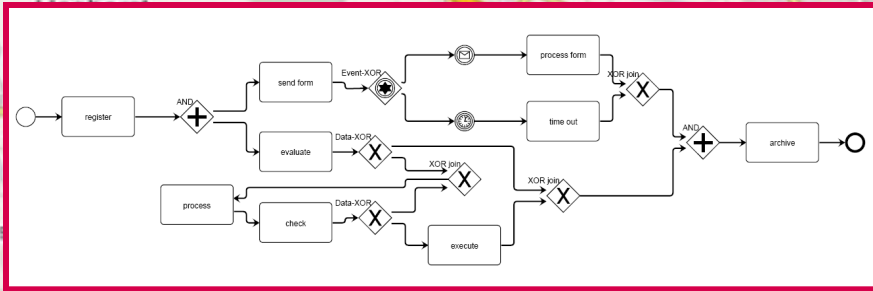
Where innovation starts

Business Process Navigation?



- Often a good process map is missing (incorrect, outdated, no color, ...)
- Process maps inherit the limitations of paper maps (no zoom or views)
- Process maps tend to aim at "controlling the driver"
- Current location unknown
- No traffic information is given
- No recalculation of the route
- No estimated arrival time
- ...

What we can learn from maps ...



Why imitate paper maps?



- **Zoom in - zoom out**
- **Various views (e.g. show hotels and fuel stations at will)**
- **Dynamic content!**
- **Traffic information**
- **Show current location**

ProM's Fuzzy Miner: Seamless zoom

ProM [5.0]

File Mining Analysis Conversion Exports Window Help

Results - Fuzzy Miner on Raw MiningBronbestand_Goed_Ontology.mxml (unfiltered)

Fuzzy Model Toolkit

Transformer Editor Unary Metrics Binary Metrics Animation

The image displays two side-by-side screenshots of the ProM software interface, demonstrating the 'Seamless zoom' feature. Both screenshots show the 'Fuzzy Model Toolkit' window with the same fuzzy model loaded. The left screenshot shows the model at a low zoom level, appearing as a dense, complex network of nodes and edges. The right screenshot shows the same model at a higher zoom level, where individual nodes and edges are clearly visible. The zoomed-in view includes a search bar at the bottom right with the text 'search...', a vertical zoom slider set to 41%, and filter controls for 'Concurrency filter', 'Edge filter', and 'Node filter' with a 'Significance cutoff' slider set to 0.485. The ProM logo is visible in the top right corner of the interface.

ProM's "real animation"

The screenshot displays the ProM 5.0 interface. The main window shows a "Fuzzy Model Animation (2)" with a complex flowchart of nodes and edges. The nodes are labeled with Dutch text and "complete" status. The flowchart is rendered in a 3D style with red and white nodes and edges. On the left side, there are two gauges: "Activity" at 100% and "Completion" at 43%. Below the gauges is a "case progress" section showing a grid of numbers. At the bottom, there is a playback control bar with a play button, a progress bar, a digital clock showing "18:36:05 Fri 9 Nov '07", and sliders for "playback speed" and "zoom view".

Activity
100 %

Completion
43 %

case progress

2210817	204587	203322	39
212090	182871	180950	10
190543	203862011	180950	10
11761809	192887	180950	10
1433611	117315	120119703	10
111182	124524973	1408	10
1177339	14721	58643	10
144788	138417	14721	10
80437	85944	80861582	10
604	49039	124815	10
121511	44	10879388	10
44655	78035	55885	10

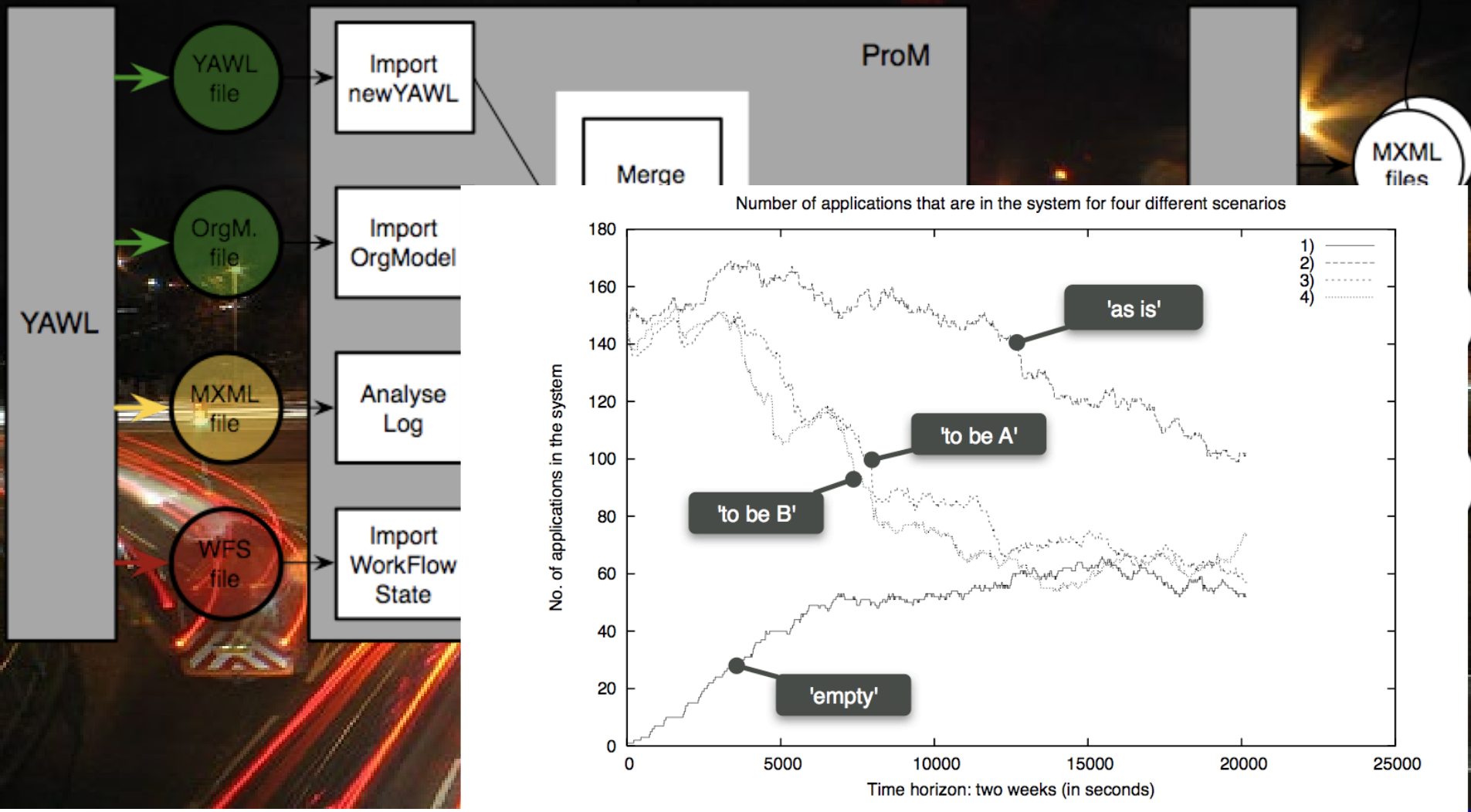
18:36:05 Fri 9 Nov '07

playback speed

zoom view

19:40:35 [D] Processing data for buffered log reader completed.

ProM's "real simulation"

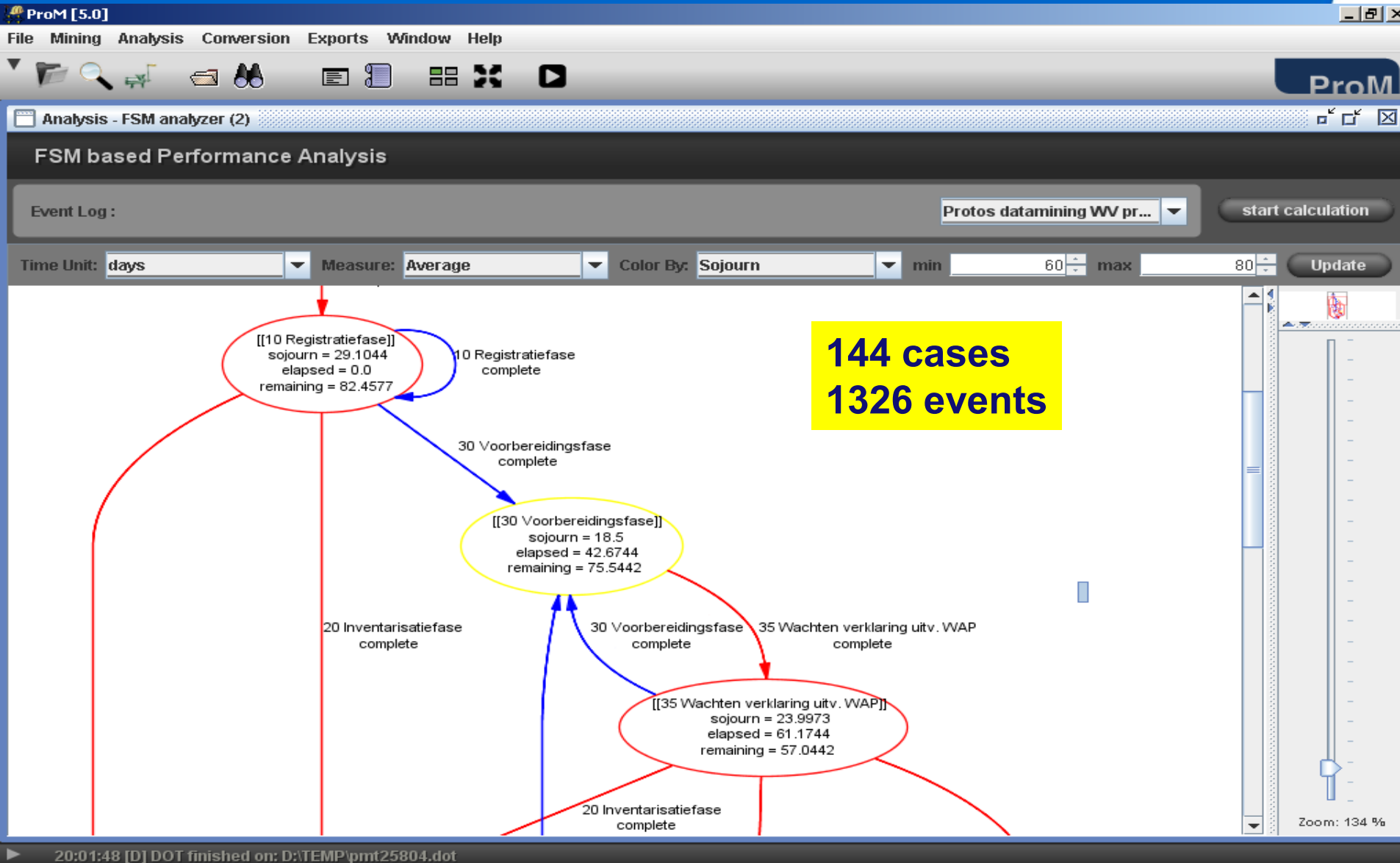


Prediction and recommendation



- **Prediction: When are we home?**
- **Recommendation: What should I do next?**
- **Suggestions without force and the willingness to continuously recalculate the route.**

ProM's Case prediction capabilities



Conclusion



TU / **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Conclusion

- The abundance of event data enables a wide variety of process mining techniques ranging from process discovery to conformance checking.
- This is already possible today!
- Check out ProM with its 250+ plug-ins.
- A reality check for people that are involved in process modeling.
- Demand TomTom functionality!

Thanks!

cf. www.processmining.org

- Wil van der Aalst
- Peter van den Brand
- Boudewijn van Dongen
- Christian Günther
- Eric Verbeek
- Ana Karla Alves de Medeiros
- Anne Rozinat
- Minseok Song
- Ton Weijters
- Remco Dijkman
- Gianluigi Greco
- Antonella Guzzo
- Kristian Bisgaard Lassen
- Ronny Mans
- Jan Mendling
- Vladimir Rubin
- Kenny van Uden
- Irene Vanderfeesten
- Barbara Weber
- Lijie Wen
- Mercy Amiyo
- Carmen Bratosin
- Toon Calders
- Jorge Cardoso
- Ronald Crooy
- Florian Gottschalk
- Monique Jansen-Vullers
- Peter Khisa Wakholi
- Nicolas Knaak
- Sven Lambrechts
- Joyce Nakatumba
- Mariska Netjes
- Mykola Pechenizkiy
- Maja Pesic
- Hajo Reijers
- Stefanie Rinderle
- Domenico Saccà
- Helen Schonenberg
- Marc Voorhoeve
- Jianmin Wang
- Jan Martijn van der Werf
- Martin van Wingerden
- Jianhong Ye
- Huub de Beer
- Elena Casares
- Alina Chipaila
- Walid Gaaloul
- Martijn van Giessel
- Shaifali Gupta
- Thomas Hoffmann
- Peter Hornix
- René Kerstjens
- Ralf Kramer
- Wouter Kunst
- Laura Maruster
- Andriy Nikolov
- Adarsh Ramesh
- Jo Theunissen
- ...

Relevant WWW sites

- <http://www.processmining.org>
- [http:// promimport.sourceforge.net](http://promimport.sourceforge.net)
- <http://prom.sourceforge.net>
- <http://www.workflowpatterns.com>
- <http://www.workflowcourse.com>
- <http://www.vdaalst.com>

