Business Process Intelligence Course
〈 Lecture 4 〉

Event Data and Tooling

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

Part I: Preliminaries
- Chapter 2: Process Modeling and Analysis
- Chapter 3: Data Mining

Part II: From Event Logs to Process Models
- Chapter 4: Getting the Data
- Chapter 5: Process Discovery: An Introduction
- Chapter 6: Advanced Process Discovery Techniques

Part III: Beyond Process Discovery
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

Part IV: Putting Process Mining to Work
- Chapter 10: Tool Support
- Chapter 11: Analyzing "Lasagna Processes"
- Chapter 12: Analyzing "Spaghetti Processes"

Part V: Reflection
- Chapter 13: Cartography and Navigation
- Chapter 14: Epilogue
Overview: wrap-up process discovery, event data, and tools

Software systems model, analyze, and support processes in the 'world' of people, machines, components, and organizations. Event logs record events such as messages and transactions. Process mining tools like Prom, Disco, and Celonis enhance and discover processes, ensuring conformance and improvement.
overview of process discovery approaches
Characteristics of process discovery algorithms

- **Representational bias**
  - Inability to represent concurrency
  - Inability to deal with (arbitrary) loops
  - Inability to represent silent actions
  - Inability to represent duplicate actions
  - Inability to model OR-splits/joins
  - Inability to represent non-free-choice behavior
  - Inability to represent hierarchy

- **Ability to deal with noise**

- **Completeness notion assumed**

- **Approach used** (direct algorithmic approaches, two-phase approaches, computational intelligence approaches, partial approaches, etc.)
Examples

- Algorithmic techniques
  - Alpha miner
  - Alpha+, Alpha++, Alpha#
  - FSM miner
  - Fuzzy miner
  - Heuristic miner
  - Multi phase miner
  - Inductive miner

- Genetic process mining
  - Single/duplicate tasks
  - Distributed GM
  - Evolutionary Tree Miner

- Region-based process mining
  - State-based regions
  - Language based regions

- Classical approaches not dealing with concurrency
  - Inductive inference
  - Sequence mining
  - Hidden Markov models
a more direct approach: language-based regions
Region-based mining is about finding places

- Two types of regions theory:
  - **State-based** regions (seen before)
  - **Language-based** regions (next)
- All about discovering places (like in the $\alpha$ algorithm)!
Remember: conditions used by the \( \alpha \)-algorithm to create places

4. \( X_L = \{ (A,B) \mid A \subseteq T_L \land A \neq \emptyset \land B \subseteq T_L \land B \neq \emptyset \land \forall a \in A \forall b \in B \ a \rightarrow_L b \land \forall a_1,a_2 \in A \ a_1 \#_L a_2 \land \forall b_1,b_2 \in B \ b_1 \#_L b_2 \} \),

5. \( Y_L = \{ (A,B) \in X_L \mid \forall (A',B') \in X_L A \subseteq A' \land B \subseteq B' \Rightarrow (A,B) = (A',B') \} \)
Remember: conditions used by the state-based region miner to create places

region needs to agree on all labels (and be minimal)

For example p3:
- enter: b, e
- leave: d
- do-not-cross: a, c
Quiz Question: How to remove behavior?
Answer:
Add places or remove transitions
Quiz Question:
How to add behavior?
Answer:
Add transitions or remove places!
Places limit behavior

- abcd
- ad
- abed
- abccd
- acbd

- aebcd
- aed
- aad
- caed
- aded
Places limit behavior

- abcd
- ad
- abed
- abcccd
- acbd
- aecccd
- aed
- aad
- caed
- aded
Places limit behavior

- abcd
- ad
- abed
- abccd
- acbd

- aebcd
- aed
- aad
- caed
- aded
Places limit behavior

- abcd
- ad
- abed
- abcccd
- acbd

- aebcd
- aed
- aad
- caed
- aded
Places limit behavior

- abcd
- ad
- abed
- abccd
- acbd
- aebcd
- aed
- aad
- caed
- aded
Region $R = (X,Y,c)$ corresponding to place $p_R$: $X = \{a_1,a_2,c_1\}$ = transitions producing a token for $p_R$, $Y = \{b_1,b_2,c_1\}$ = transitions consuming a token from $p_R$, and $c$ is the initial marking of $p_R$. 
Basic idea: enough tokens should be present when consuming

A place is feasible if it can be added without disabling any of the traces in the event log.

for any $\sigma \in L, k \in \{1, \ldots, |\sigma|\}$, $\sigma_1 = h \Delta^{k-1}(\sigma)$, $a = \sigma(k)$, $\sigma_2 = h \Delta^k(\sigma) = \sigma_1 \oplus a$:

$$c + \sum_{t \in X} \partial_{\text{multiset}}(\sigma_1)(t) - \sum_{t \in Y} \partial_{\text{multiset}}(\sigma_2)(t) \geq 0.$$
Basic idea: enough tokens should be present when consuming

A place is feasible if it can be added without disabling any of the traces in the event log.

for any $\sigma \in L, k \in \{1, \ldots, |\sigma|\}$, $\sigma_1 = hd^{k-1}(\sigma)$, $a = \sigma(k)$, $\sigma_2 = hd^k(\sigma) = \sigma_1 \oplus a$:

$$c + \sum_{t \in X} \partial_{multiset}(\sigma_1)(t) - \sum_{t \in Y} \partial_{multiset}(\sigma_2)(t) \geq 0.$$
Example of language-based region

\[ R = (\{a, b, c\}, \{c, d, e\}, 0) \]

1. accd
2. bd
3. bce
4. ace
5. acd
6. bcce
7. ade

\[ \downarrow \text{accd} : 0 + 0 - 0 \geq 0 \]
\[ \downarrow \text{ccd} : 0 + 1 - 1 \geq 0 \]
\[ \downarrow \text{cd} : 0 + 2 - 2 \geq 0 \]
\[ \downarrow \text{d} : 0 + 3 - 3 \geq 0 \]
\[ \downarrow \text{ade} : 0 + 0 - 0 \geq 0 \]
\[ \downarrow \text{de} : 0 + 1 - 1 \geq 0 \]
\[ \downarrow \text{e} : 0 + 1 - 2 < 0 \]
Example

\[ L_9 = [\langle a, c, d \rangle^{45}, \langle b, c, e \rangle^{42}] \]

Initially "the place" contains \( c \) tokens.

Transition \( a \) produces \( x_a \) tokens for "the place".

Transition \( a \) consumes \( y_a \) tokens from "the place".

\[
\begin{align*}
c - y_a & \geq 0 \\
c + x_a - (y_a + y_b) & \geq 0 \\
c + x_a + x_c - (y_a + y_c + y_b) & \geq 0 \\
c - y_b & \geq 0 \\
c, x_a, \ldots, x_e, y_a, \ldots, y_e & \in \{0, 1\}
\end{align*}
\]
Regions

\[ L_9 = [\langle a, c, d \rangle^{45}, \langle b, c, e \rangle^{42}] \]

\[ R_1 = (\emptyset, \{a, b\}, 1) \]
\[ c = y_a = y_b = 1, \quad x_a = x_b = x_c = x_d = x_e = y_c = y_d = y_e = 0 \]

\[ R_2 = (\{a, b\}, \{c\}, 0) \]
\[ x_a = x_b = y_c = 1, \quad c = x_c = x_d = x_e = y_a = y_b = y_d = y_e = 0 \]

\[ R_3 = (\{c\}, \{d, e\}, 0) \]
\[ x_c = y_d = y_e = 1, \quad c = x_a = x_b = x_d = x_e = y_a = y_b = y_c = 0 \]

\[ R_4 = (\{d, e\}, \emptyset, 0) \]
\[ x_d = x_e = 1, \quad c = x_a = x_b = x_c = y_a = y_b = y_c = y_d = y_e = 0 \]

\[ R_5 = (\{a\}, \{d\}, 0) \]
\[ x_a = y_d = 1, \quad c = x_b = x_c = x_d = x_e = y_a = y_b = y_c = y_e = 0 \]

\[ R_6 = (\{b\}, \{e\}, 0) \]
\[ x_b = y_e = 1, \quad c = x_a = x_c = x_d = x_e = y_a = y_b = y_c = y_d = 0 \]
\[ L_9 = [\langle a, c, d \rangle^{45}, \langle b, c, e \rangle^{42}] \]

\[ R_1 = (\emptyset, \{a, b\}, 1) \]
\[ c = y_a = y_b = 1, \quad x_a = x_b = x_c = x_d = x_e = y_c = y_d = y_e = 0 \]

\[ R_2 = (\{a, b\}, \{c\}, 0) \]
\[ x_a = x_b = y_c = 1, \quad c = x_c = x_d = x_e = y_a = y_b = y_d = y_e = 0 \]

\[ R_3 = (\{c\}, \{d, e\}, 0) \]
\[ x_c = y_d = y_e = 1, \quad c = x_a = x_b = x_d = x_e = y_a = y_b = y_c = 0 \]

\[ R_4 = (\{d, e\}, \emptyset, 0) \]
\[ x_d = x_e = 1, \quad c = x_a = x_b = x_c = y_a = y_b = y_c = y_d = y_e = 0 \]

\[ R_5 = (\{a\}, \{d\}, 0) \]
\[ x_a = y_d = 1, \quad c = x_b = x_c = x_d = x_e = y_a = y_b = y_c = y_e = 0 \]

\[ R_6 = (\{b\}, \{e\}, 0) \]
\[ x_b = y_e = 1, \quad c = x_a = x_c = x_d = x_e = y_a = y_b = y_c = y_d = 0 \]
Characteristics of region-based mining

• Can be used to discover more complex control-flow structures.
• Classical approaches need to be adapted (overfitting!).
• Representational bias can be parameterized (e.g., free-choice nets, label splitting, etc.).
• Problems dealing with noise.
genetic process mining
Genetic process mining

- event log
- create initial population
- compute fitness
- tournament
- select best individual
- next generation
- elitism
- mutation
- children
- crossover
- parents
- “dead” individuals
- tournament
- select best individual
- event log
Design decisions

- Representation of individuals
- Initialization
- Fitness function
- Selection strategy (tournament and elitism)
- Crossover
- Mutation
Example: crossover

- **start**
  - a: register request
  - b: examine thoroughly
  - c: examine casually
  - d: check ticket
  - g: pay compensation

- **pay compensation**
  - h: reject request
  - e: reinitiate request

- **end**
  - f: decision point

- **start**
  - a: register request
  - b: examine thoroughly
  - c: examine casually
  - d: check ticket
  - g: pay compensation

- **pay compensation**
  - h: reject request
  - e: reinitiate request

- **end**
  - f: decision point
Example: mutation

start

register request

examine thoroughly

examine casually

check ticket

decide

pay compensation

reject request

reinitiate request

end

remove place

added arc
Characteristics of genetic process mining

- Requires a lot of computing power.
- Can be distributed easily.
- Can deal with noise, infrequent behavior, duplicate tasks, invisible tasks, etc.
- Allows for incremental improvement and combinations with other approaches (heuristics post-optimization, etc.).
Other approaches, e.g. fuzzy mining
Disco uses the ideas from the fuzzy miner (PhD work Christian Günter)
event data
Goals of process mining

- What really happened in the past?
- Why did it happen?
- What is likely to happen in the future?
- When and why do organizations and people deviate?
- How to control a process better?
- How to redesign a process to improve its performance?
- ....
Garbage-in garbage-out
Getting the "right" data ...

"world"
- people
- machines
- components
- organizations

software system
records events, e.g., messages, transactions, etc.

models analyzes
specifies configures implements analyzes

(process) model
discovery
conformance
enhancement

event logs
Opportunity and challenge: Events are everywhere!

- Minimal requirement: events referring to an activity name and a process instance.
- Good to have: timestamps, resource information, additional data elements.
- Challenges: scoping and sometimes correlation.

Databases, ERP systems (SAP etc.), WFM/BPM logs, message logs, audit trails, etc.
From heterogeneous data sources to process mining results

- Extract, Transform, and Load (ELT)
- Data source
- Data warehouse
- Extract
- Unfiltered event logs
- Filter
- Process mining
- Discovery
- Conformance
- Enhancement
- Fine-grained scoping
- Coarse-grained scoping
- XES, MXML, or similar
- Optional

©Wil van der Aalst TU/e (use only with permission & acknowledgements)
A process consists of cases.

A case consists of events such that each event relates to precisely one case.

Events within a case are ordered.

Events can have attributes.

Examples of typical attribute names are activity, time, costs, and resource.

<table>
<thead>
<tr>
<th>case id</th>
<th>event id</th>
<th>timestamp</th>
<th>properties</th>
<th>resource</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35654423</td>
<td>30-12-2010:11.02</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654424</td>
<td>31-12-2010:10.06</td>
<td>examine thoroughly</td>
<td>Sue</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654425</td>
<td>05-01-2011:15.12</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654426</td>
<td>06-01-2011:11.18</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654427</td>
<td>07-01-2011:14.24</td>
<td>reject request</td>
<td>Pete</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>35654483</td>
<td>30-12-2010:11.32</td>
<td>register request</td>
<td>Mike</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654485</td>
<td>30-12-2010:12.12</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654487</td>
<td>30-12-2010:14.16</td>
<td>examine casually</td>
<td>Pete</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654488</td>
<td>05-01-2011:11.22</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654489</td>
<td>08-01-2011:12.05</td>
<td>pay compensation</td>
<td>Ellen</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>35654521</td>
<td>30-12-2010:14.32</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654522</td>
<td>30-12-2010:15.06</td>
<td>examine casually</td>
<td>Mike</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654524</td>
<td>30-12-2010:16.34</td>
<td>check ticket</td>
<td>Mike</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654525</td>
<td>06-01-2011:09.18</td>
<td>decide</td>
<td>Ellen</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654526</td>
<td>06-01-2011:12.18</td>
<td>reinitiate request</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654527</td>
<td>06-01-2011:13.06</td>
<td>examine thoroughly</td>
<td>Sean</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654530</td>
<td>08-01-2011:11.43</td>
<td>check ticket</td>
<td>Pete</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654531</td>
<td>09-01-2011:09.55</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654533</td>
<td>15-01-2011:10.45</td>
<td>pay compensation</td>
<td>Ellen</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>35654641</td>
<td>06-01-2011:15.02</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654643</td>
<td>07-01-2011:12.06</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654644</td>
<td>08-01-2011:14.43</td>
<td>examine thoroughly</td>
<td>Sean</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654645</td>
<td>09-01-2011:12.02</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654647</td>
<td>12-01-2011:15.44</td>
<td>reject request</td>
<td>Ellen</td>
<td>200</td>
</tr>
</tbody>
</table>
Another view

<table>
<thead>
<tr>
<th>case id</th>
<th>event id</th>
<th>timestamp</th>
<th>properties</th>
<th>resources</th>
<th>costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35654423</td>
<td>30-12-2010:11.02</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654424</td>
<td>31-12-2010:10.06</td>
<td>examine thoroughly</td>
<td>Sue</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654425</td>
<td>05-01-2011:15.12</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654426</td>
<td>06-01-2011:11.18</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654427</td>
<td>07-01-2011:14.24</td>
<td>reject request</td>
<td>Pete</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>35654483</td>
<td>30-12-2010:11.32</td>
<td>register request</td>
<td>Mike</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654485</td>
<td>30-12-2010:12.12</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654487</td>
<td>30-12-2010:14.16</td>
<td>examine casually</td>
<td>Pete</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654488</td>
<td>05-01-2011:11.22</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654489</td>
<td>08-01-2011:12.05</td>
<td>pay compensation</td>
<td>Ellen</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>35654521</td>
<td>30-12-2010:14.32</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654522</td>
<td>30-12-2010:15.06</td>
<td>examine casually</td>
<td>Mike</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654524</td>
<td>30-12-2010:16.34</td>
<td>check ticket</td>
<td>Ellen</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654525</td>
<td>06-01-2011:09.18</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654526</td>
<td>06-01-2011:12.18</td>
<td>reinstate request</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654527</td>
<td>06-01-2011:13.06</td>
<td>examine thoroughly</td>
<td>Sean</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654530</td>
<td>08-01-2011:11.43</td>
<td>check ticket</td>
<td>Pete</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654531</td>
<td>09-01-2011:09.55</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654533</td>
<td>15-01-2011:10.45</td>
<td>pay compensation</td>
<td>Ellen</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>35654641</td>
<td>06-01-2011:15.02</td>
<td>register request</td>
<td>Pete</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>35654643</td>
<td>07-01-2011:12.06</td>
<td>check ticket</td>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>35654644</td>
<td>08-01-2011:14.43</td>
<td>examine thoroughly</td>
<td>Sean</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>35654645</td>
<td>09-01-2011:12.02</td>
<td>decide</td>
<td>Sara</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>35654647</td>
<td>12-01-2011:15.44</td>
<td>reject request</td>
<td>Ellen</td>
<td>200</td>
</tr>
</tbody>
</table>
Standard transactional life-cycle model
Example: low-level Petri net view

assign, start, and complete are recorded for a and c

start and complete, and optionally also suspend and resume are recorded for b
Five activity instances

a: schedule assign start complete

d: start complete

b: schedule assign reassign start suspend resume complete

c: start suspend resume suspend abort_activity

e: complete

©Wil van der Aalst TU/e (use only with permission & acknowledgements)
Overlapping activity instances

```
a: start start complete complete
    6 hours

5 hours

a: start start complete complete
   2 hours

9 hours
```

Successful termination

Unsuccessful termination

Schedule

Assign

Reassign

Start

Resume

Suspend

Abort activity

Abort case

Autoskip

Manualskip

Complete

Withdraw

Successful termination

Unsuccessful termination
Using attributes: Not just control-flow!

Social network showing how work flows from one person to another

Performance indicators per activity

**Activity b**
Frequency: 456  
Waiting time: 15.6 +/- 2.5 hours  
Service time: 1.2 +/- 0.5 hours  
Costs: 412 +/- 55 euros

**Activity g**
Frequency: 311  
Waiting time: 12.4 +/- 2.1 hours  
Service time: 0.5 +/- 0.2 hours  
Costs: 198 +/- 35 euros

**Activity h**
Frequency: 407  
Waiting time: 7.4 +/- 1.8 hours  
Service time: 1.1 +/- 0.3 hours  
Costs: 209 +/- 38 euros
XES (eXtensible Event Stream)

- Adopted by the IEEE Task Force on Process Mining.
- Predecessor: MXML and SA-MXML.
- The format is supported by tools such as ProM (as of version 6), Disco, XESame, and OpenXES.
- ProMimport supports MXML.
Event log consists of:
  + traces (process instances)
  + events

- Standard extensions:
  - concept (for naming)
  - lifecycle (for transactional properties)
  - org (for the organizational perspective)
  - time (for timestamps)
  - semantic (for ontology references)
extensions loaded

every trace has a name

every event has a name and a transition

classifier = name + transition

start of trace (i.e. process instance)

name of trace

resource

timestamp

name of event (activity name)
Challenges when extracting event logs

- **Correlation**: Events in an event log are grouped per case. This simple requirement can be quite challenging as it requires event correlation, i.e., events need to be related to each other.
- **Timestamps**: Events need to be ordered per case. Typical problems: only dates, different clocks, delayed logging.
- **Snapshots**: Cases may have a lifetime extending beyond the recorded period, e.g., a case was started before the beginning of the event log.
- **Scoping**: How to decide which "tables" to incorporate?
- **Granularity**: the events in the event log are at a different level of granularity than the activities relevant for end users.
How to select events?
Disco provides five filter settings
Flattening reality into event logs

Order
- OrderID : OrderID
- Customer : CustID
- Amount : Euro
- Created : DateTime
- Paid : DateTime
- Completed : DateTime

Orderline
- OrderLineID : OrderLineID
- OrderID : OrderID
- Product : ProdID
- NofItems : PosInt
- TotalWeight : Weight
- Entered : DateTime
- BackOrdered : DateTime
- Secured : DateTime
- DelID : DelID

Delivery
- DelID : DelID
- DelAddress : Address
- Contact : PhoneNo

Attempt
- DelID : DelID
- When : DateTime
- Successful : Bool

1 1..* 0..1
0..* 1
### Order

<table>
<thead>
<tr>
<th>OrderID</th>
<th>Customer</th>
<th>Amount</th>
<th>Created</th>
<th>Paid</th>
<th>Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>91245</td>
<td>John</td>
<td>100</td>
<td>28-11-2011:08.12</td>
<td>02-12-2011:13.45</td>
<td>05-12-2011:11.33</td>
</tr>
<tr>
<td>91561</td>
<td>Mike</td>
<td>530</td>
<td>28-11-2011:12.22</td>
<td>03-12-2011:14.34</td>
<td>05-12-2011:09.32</td>
</tr>
<tr>
<td>91812</td>
<td>Mary</td>
<td>234</td>
<td>29-11-2011:09.45</td>
<td>02-12-2011:09.44</td>
<td>04-12-2011:13.33</td>
</tr>
<tr>
<td>92233</td>
<td>Sue</td>
<td>110</td>
<td>29-11-2011:10.12</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>92345</td>
<td>Kirsten</td>
<td>195</td>
<td>29-11-2011:14.45</td>
<td>02-12-2011:13.45</td>
<td>null</td>
</tr>
<tr>
<td>92355</td>
<td>Pete</td>
<td>320</td>
<td>29-11-2011:16.32</td>
<td>null</td>
<td>null</td>
</tr>
</tbody>
</table>

### Orderline

<table>
<thead>
<tr>
<th>OrderLineID</th>
<th>OrderID</th>
<th>Product</th>
<th>NoItems</th>
<th>TotalWeight</th>
<th>Entered</th>
<th>BackOrdered</th>
<th>Secured</th>
<th>DellID</th>
</tr>
</thead>
<tbody>
<tr>
<td>112345</td>
<td>91245</td>
<td>iPhone 4G</td>
<td>1</td>
<td>0.250</td>
<td>28-11-2011:08.13</td>
<td>null</td>
<td>28-11-2011:08.55</td>
<td>882345</td>
</tr>
<tr>
<td>112346</td>
<td>91245</td>
<td>iPod nano</td>
<td>2</td>
<td>0.300</td>
<td>28-11-2011:08.14</td>
<td>28-11-2011:08.55</td>
<td>30-11-2011:09.06</td>
<td>882346</td>
</tr>
<tr>
<td>112347</td>
<td>91245</td>
<td>iPod classic</td>
<td>1</td>
<td>0.200</td>
<td>28-11-2011:08.15</td>
<td>null</td>
<td>29-11-2011:10.06</td>
<td>882345</td>
</tr>
<tr>
<td>112448</td>
<td>91561</td>
<td>iPhone 4G</td>
<td>1</td>
<td>0.250</td>
<td>28-11-2011:12.23</td>
<td>null</td>
<td>28-11-2011:12.59</td>
<td>882345</td>
</tr>
<tr>
<td>112449</td>
<td>91561</td>
<td>iPod classic</td>
<td>1</td>
<td>0.200</td>
<td>28-11-2011:12.24</td>
<td>28-11-2011:16.22</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>112452</td>
<td>91812</td>
<td>iPhone 4G</td>
<td>5</td>
<td>1.250</td>
<td>29-11-2011:09.46</td>
<td>null</td>
<td>29-11-2011:10.58</td>
<td>882346</td>
</tr>
</tbody>
</table>

### Delivery

<table>
<thead>
<tr>
<th>DellID</th>
<th>DelAddress</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>882345</td>
<td>5513VJ-22a</td>
<td>0497-2553660</td>
</tr>
<tr>
<td>882346</td>
<td>5513XG-45</td>
<td>040-2298761</td>
</tr>
</tbody>
</table>

### Attempt

<table>
<thead>
<tr>
<th>DellID</th>
<th>When</th>
<th>Successful</th>
</tr>
</thead>
<tbody>
<tr>
<td>882345</td>
<td>05-12-2011:08.55</td>
<td>false</td>
</tr>
<tr>
<td>882345</td>
<td>06-12-2011:09.12</td>
<td>false</td>
</tr>
<tr>
<td>882345</td>
<td>07-12-2011:08.56</td>
<td>true</td>
</tr>
<tr>
<td>882346</td>
<td>05-12-2011:08.43</td>
<td>true</td>
</tr>
<tr>
<td>case id</td>
<td>activity</td>
<td>timestamp</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>91245</td>
<td>create order</td>
<td>28-11-2011:08.12</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.13</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.14</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.15</td>
</tr>
<tr>
<td>91245</td>
<td>secure order line</td>
<td>28-11-2011:08.55</td>
</tr>
<tr>
<td>91245</td>
<td>secure order line</td>
<td>29-11-2011:10.06</td>
</tr>
<tr>
<td>91245</td>
<td>secure order line</td>
<td>30-11-2011:09.06</td>
</tr>
<tr>
<td>91245</td>
<td>pay order</td>
<td>02-12-2011:13.45</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>05-12-2011:08.43</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>05-12-2011:08.55</td>
</tr>
<tr>
<td>91245</td>
<td>complete order</td>
<td>05-12-2011:11.33</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>06-12-2011:09.12</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>07-12-2011:08.56</td>
</tr>
<tr>
<td>91561</td>
<td>create order</td>
<td>28-11-2011:12.22</td>
</tr>
<tr>
<td>91561</td>
<td>enter order line</td>
<td>28-11-2011:12.23</td>
</tr>
</tbody>
</table>

...
Another example: concert tickets

- What is the process about:
  - concerts?
  - bookings?
  - customers?
  - concert halls?
  - payments?
  - seats?
  - bands?
Other examples

• Organizing the reviewing process of a conference: Is it about reviewers, authors, papers, reviews, PC chairs, … ?
• An organization hiring people: Is it about job applications or vacancies?
• Analyzing the usage and failures of X-ray machines: Is it about machines, machine days, components, patients, treatments, sub-routines, … ?

Therefore, the **selection and scoping of instances** is needed. Like deciding on the elements to be put on map; there may be many maps covering partially overlapping areas.
Creating a 2-D slice of a 3-D reality: the process is viewed from a specific angle, the process is scoped using a frame, and the resolution determines the granularity of the resulting model.
Outside the scope of this course.
In this course we only create conventional (i.e., flat) models.
Data quality issues

- Missing case ID
- Imprecise timestamps
- Granularity of events
- Missing events
- ....
<table>
<thead>
<tr>
<th>case</th>
<th>event</th>
<th>belongs to</th>
<th>c attribute</th>
<th>position</th>
<th>activity name</th>
<th>timestamp</th>
<th>resource</th>
<th>e attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing data</td>
<td>In reality a case has been executed but it has not been recorded in the log</td>
<td>Events are missing within the trace although they occurred in reality.</td>
<td>Association between events and cases is lost (correlation problem)</td>
<td>Case attribute was not recorded.</td>
<td>Ordering of events in the trace is lost.</td>
<td>Activity names of events are missing.</td>
<td>Timestamps of events are missing.</td>
<td>Resources that executed an activity have not been recorded.</td>
</tr>
<tr>
<td>incorrect data</td>
<td>Some cases in the log belong to a different process.</td>
<td>Events that were not actually executed for some cases are logged</td>
<td>Association between events and cases are logged incorrectly.</td>
<td>Values corresponding to case attributes are logged incorrectly.</td>
<td>Order is mixed up.</td>
<td>Wrong activity names are recorded.</td>
<td>Incorrect timestamps.</td>
<td>Incorrect resource assigned to event.</td>
</tr>
<tr>
<td>imprecise data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>irrelevant data</td>
<td>Irrelevant cases are included and cannot be removed easily.</td>
<td>Events may be irrelevant and difficult to remove</td>
<td></td>
<td>Provided value is too coarse, e.g., city but no address.</td>
<td>For example concurrent events may have become been totally ordered.</td>
<td>Activity names are too coarse.</td>
<td>Days rather than minutes or seconds. Hence, precise order cannot be derived.</td>
<td>Just role or department is recorded.</td>
</tr>
</tbody>
</table>

Bose, R.P.J.C.; Mans, R.S.; van der Aalst, W.M.P., "Wanna improve process mining results?," Computational Intelligence and Data Mining (CIDM 2013), doi: 10.1109/CIDM.2013.6597227
tooling
The Broader Business Process Intelligence (BPI) Field

- Process discovery
- Conformance checking
- Process mining
- Predictive analytics
- Data mining / machine learning

BPM
business intelligence
tools
Business Intelligence?

• “BI is a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision making”

• Examples of products:
  IBM Cognos Business Intelligence (IBM), Oracle Business Intelligence (Oracle), SAP BusinessObjects (SAP), WebFOCUS (Information Builders), MS SQL Server (Microsoft), MicroStrategy (MicroStrategy), NovaView (Panorama Software), QlikView (QlikTech), SAS Enterprise Business Intelligence (SAS), TIBCO Spotfire Analytics (TIBCO), Jaspersoft (Jaspersoft), and Pentaho BI Suite (Pentaho).
Typical functionality

- ETL (Extract, Transform, and Load).
- Ad-hoc querying.
- Reporting.
- Interactive dashboards
- Alert generation.

Three dimensional OLAP cube containing sales data. Each cell refers to all sales of a particular product in a particular region and in a particular period. For each cell the BI product can compute metrics such as the number of items sold or the total value.
### Example: Pentaho

www.pentaho.com

#### Territory Sales Analysis

<table>
<thead>
<tr>
<th>Territory</th>
<th>Line</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sales</td>
<td>Sales</td>
<td>Sales</td>
<td></td>
</tr>
<tr>
<td>APAC</td>
<td>Classic Cars</td>
<td>$115,011</td>
<td>$199,372</td>
<td>$97,574</td>
</tr>
<tr>
<td></td>
<td>Vintage Cars</td>
<td>$111,639</td>
<td>$147,212</td>
<td>$35,888</td>
</tr>
<tr>
<td></td>
<td>Motorcycles</td>
<td>$60,789</td>
<td>$63,159</td>
<td>$65,870</td>
</tr>
<tr>
<td></td>
<td>Trucks and Buses</td>
<td>$11,258</td>
<td>$85,634</td>
<td>$55,735</td>
</tr>
<tr>
<td></td>
<td>Planes</td>
<td>$42,663</td>
<td>$67,601</td>
<td>$11,082</td>
</tr>
<tr>
<td></td>
<td>Ships</td>
<td>-</td>
<td>$13,323</td>
<td>$5,070</td>
</tr>
<tr>
<td></td>
<td>Trains</td>
<td>$1,681</td>
<td>$8,226</td>
<td>-</td>
</tr>
<tr>
<td>EMEA</td>
<td>Classic Cars</td>
<td>$611,273</td>
<td>$1,015,730</td>
<td>$384,338</td>
</tr>
<tr>
<td></td>
<td>Vintage Cars</td>
<td>$263,691</td>
<td>$504,062</td>
<td>$83,324</td>
</tr>
<tr>
<td></td>
<td>Motorcycles</td>
<td>$19,015</td>
<td>$201,042</td>
<td>$151,760</td>
</tr>
<tr>
<td></td>
<td>Trucks and Buses</td>
<td>$228,699</td>
<td>$185,421</td>
<td>$86,859</td>
</tr>
<tr>
<td></td>
<td>Planes</td>
<td>$154,519</td>
<td>$224,523</td>
<td>$288,008</td>
</tr>
<tr>
<td></td>
<td>Ships</td>
<td>$172,428</td>
<td>$186,992</td>
<td>$288,944</td>
</tr>
<tr>
<td></td>
<td>Trains</td>
<td>$29,530</td>
<td>$95,973</td>
<td>-</td>
</tr>
<tr>
<td>EMEA Total</td>
<td></td>
<td>$1,681,987</td>
<td>$2,396,408</td>
<td>$929,829</td>
</tr>
<tr>
<td>EMEA Sum</td>
<td>Classic Cars</td>
<td>$611,273</td>
<td>$1,015,730</td>
<td>$384,338</td>
</tr>
<tr>
<td></td>
<td>Vintage Cars</td>
<td>$263,691</td>
<td>$504,062</td>
<td>$83,324</td>
</tr>
<tr>
<td></td>
<td>Motorcycles</td>
<td>$19,015</td>
<td>$201,042</td>
<td>$151,760</td>
</tr>
<tr>
<td></td>
<td>Trucks and Buses</td>
<td>$228,699</td>
<td>$185,421</td>
<td>$86,859</td>
</tr>
<tr>
<td></td>
<td>Planes</td>
<td>$154,519</td>
<td>$224,523</td>
<td>$288,008</td>
</tr>
<tr>
<td></td>
<td>Ships</td>
<td>$172,428</td>
<td>$186,992</td>
<td>$288,944</td>
</tr>
<tr>
<td></td>
<td>Trains</td>
<td>$29,530</td>
<td>$95,973</td>
<td>-</td>
</tr>
<tr>
<td>EMA Total</td>
<td></td>
<td>$1,681,987</td>
<td>$2,396,408</td>
<td>$929,829</td>
</tr>
<tr>
<td>EMA Sum</td>
<td>Classic Cars</td>
<td>$611,273</td>
<td>$1,015,730</td>
<td>$384,338</td>
</tr>
<tr>
<td></td>
<td>Vintage Cars</td>
<td>$263,691</td>
<td>$504,062</td>
<td>$83,324</td>
</tr>
<tr>
<td></td>
<td>Motorcycles</td>
<td>$19,015</td>
<td>$201,042</td>
<td>$151,760</td>
</tr>
<tr>
<td></td>
<td>Trucks and Buses</td>
<td>$228,699</td>
<td>$185,421</td>
<td>$86,859</td>
</tr>
<tr>
<td></td>
<td>Planes</td>
<td>$154,519</td>
<td>$224,523</td>
<td>$288,008</td>
</tr>
<tr>
<td></td>
<td>Ships</td>
<td>$172,428</td>
<td>$186,992</td>
<td>$288,944</td>
</tr>
<tr>
<td></td>
<td>Trains</td>
<td>$29,530</td>
<td>$95,973</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Dimension 1: Territory
- Europe, the Middle East and Africa (EMEA)
- Asia Pacific and Japan (APAC), etc.

#### Dimension 2: Product Line
- Classic Cars
- Vintage Cars
- Motorcycles
- Trucks and Buses
- Planes
- Ships
- Trains

#### Dimension 3: Year
- 2003
- 2004
- 2005
Business *Un*intelligence

- No real process orientation.
- Only simple views on event data.
- Focus on reporting and monitoring of KPIs.

Data mining ≠ process mining

- Data mining tools provide more “intelligent functionality” than BI tools, but are also not process-centric.
- See for example RapidMiner, WEKA (Waikato Environment for Knowledge Analysis, weka.wikispaces.com) and R (www.r-project.org).
iBPMS: "intelligent" BPM suites according to Gartner
IBM Business Process Manager

IBM SPSS products
With IBM SPSS predictive analytics software, you can use statistical analysis, data and text mining, predictive modeling and decision optimization to anticipate change and take action to improve outcomes.

IBM Cognos products
IBM Cognos business intelligence and performance management software provides you with the integrated dashboards, scorecards, reporting, analysis, and planning and budgeting capabilities you need to gain and act on fact-based insights.

IBM OpenPages products
OpenPages GRC software allows your organization to manage enterprise operational risk and compliance initiatives using a single, integrated solution.

IBM Algorithmics products
Algorithmics software helps businesses like yours to gain transparency into financial risks in advance, providing information that is vital to your organization.

IBM Varicent products
IBM Cognos Sales Performance Management (SPM) products deliver measurable improvements in how your organization’s finance, sales, human resources and IT teams operate.
Process mining tools

- ProM (de facto open source standard)
- Disco (Fluxicon)
- Perceptive Process Mining (before Futura Reflect and BPM|one)
- Celonis Discovery
- ARIS Process Performance Manager
- QPR ProcessAnalyzer
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAnalyzer (XMPro)
- ...
ProM
ProM

- **Open source** (core: GNU Public License GPL, plug-ins: often Lesser GNU Public License L-GPL)
- **ProM supports all of the techniques mentioned in book and on slides!**
- **Pluggable architecture.**
- **De facto tool in process mining research.**
Overview of ProM releases

- ProM 1.0 (29 plug-ins: 7 analysis, 3 conversion, 9 export, 4 import, 0 filter, 6 mining) 9/1/2004
- ProM 1.1 (91 plug-ins: 25 analysis, 10 conversion, 22 export, 10 import, 9 log filter, 15 mining) 5/18/2006
- ProM 2.0 7/1/2005
- ProM 2.1 9/1/2005
- ProM 3.0 2/6/2006
- ProM 3.1 (91 plug-ins: 25 analysis, 10 conversion, 22 export, 10 import, 9 log filter, 15 mining) 5/18/2006
- ProM 4.0 (142 plug-ins) 11/23/2006
- ProM 4.1 (149 plug-ins) 4/16/2007
- ProM 4.2 (175 plug-ins) 9/14/2007
- ProM 5.0 (228 plug-ins) 6/17/2008
- ProM 5.1 (254 plug-ins) 10/16/2008
- ProM 5.2 (274 plug-ins: 90 analysis, 44 conversion, 49 export, 22 import, 32 filter, 37 mining) 9/3/2009
- ProM 6.0 (170 plug-ins distributed over 50 packages) 6/17/2008
- ProM 6.1 (322 plug-ins distributed over 57 packages) 9/5/2011
- ProM 6.2 (standard, 468 plug-ins distributed over 68 packages) 8/16/2012
- ProM 6.2 (extended, 676 plug-ins distributed over 99 packages) 5/27/2013

Today: ProM 6.2 (extended, 676 plug-ins distributed over 99 packages) 5/27/2013

©Wil van der Aalst TU/e (use only with permission & acknowledgements)
Dotted chart
Alpha algorithm
Heuristic Miner
Inductive miner
Fuzzy miner
Social network analysis (several plug-ins)
Performance analysis
Conformance analysis
Petri net analysis
(structural and behavioral, incl. soundness)
Disco
(Fluxicon)
Easy import of reviewing data set

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Activity</th>
<th>Resource</th>
<th>Complete Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>invite reviewers</td>
<td>Mike</td>
<td>2006/01/08 00:00:00</td>
</tr>
<tr>
<td>2</td>
<td>get review 2</td>
<td>Carol</td>
<td>2006/01/09 00:00:00</td>
</tr>
<tr>
<td>3</td>
<td>get review 3</td>
<td>Pam</td>
<td>2006/01/10 00:00:00</td>
</tr>
<tr>
<td>4</td>
<td>get review 1</td>
<td>John</td>
<td>2006/01/11 00:00:00</td>
</tr>
<tr>
<td>5</td>
<td>collect reviews</td>
<td>Anne</td>
<td>2006/01/12 00:00:00</td>
</tr>
<tr>
<td>6</td>
<td>decide</td>
<td>Will</td>
<td>2006/01/13 00:00:00</td>
</tr>
<tr>
<td>7</td>
<td>invite additional reviewer</td>
<td>Anne</td>
<td>2006/01/18 00:00:00</td>
</tr>
<tr>
<td>8</td>
<td>get review X</td>
<td>Pam</td>
<td>2006/01/22 00:00:00</td>
</tr>
<tr>
<td>9</td>
<td>decide</td>
<td>Will</td>
<td>2006/01/27 00:00:00</td>
</tr>
<tr>
<td>10</td>
<td>invite additional reviewer</td>
<td>Mike</td>
<td>2006/01/28 00:00:00</td>
</tr>
<tr>
<td>11</td>
<td>get review X</td>
<td>Mary</td>
<td>2006/01/31 00:00:00</td>
</tr>
<tr>
<td>12</td>
<td>decide</td>
<td>Will</td>
<td>2006/02/16 00:00:00</td>
</tr>
<tr>
<td>13</td>
<td>invite additional reviewer</td>
<td>Mike</td>
<td>2006/02/23 00:00:00</td>
</tr>
<tr>
<td>14</td>
<td>get review X</td>
<td>Sara</td>
<td>2006/02/28 00:00:00</td>
</tr>
<tr>
<td>15</td>
<td>get review 3</td>
<td>John</td>
<td>2006/03/07 00:00:00</td>
</tr>
<tr>
<td>16</td>
<td>time-out X</td>
<td>INVALID</td>
<td>2006/03/08 00:00:00</td>
</tr>
<tr>
<td>17</td>
<td>decide</td>
<td>Will</td>
<td>2006/04/13 01:00:00</td>
</tr>
<tr>
<td>18</td>
<td>invite reviewers</td>
<td>Anne</td>
<td>2006/05/06 00:00:00</td>
</tr>
<tr>
<td>19</td>
<td>get review X</td>
<td>Sara</td>
<td>2006/05/09 00:00:00</td>
</tr>
<tr>
<td>20</td>
<td>get review 2</td>
<td>Sara</td>
<td>2006/06/09 00:00:00</td>
</tr>
<tr>
<td>21</td>
<td>time-out X</td>
<td>INVALID</td>
<td>2006/06/10 00:00:00</td>
</tr>
<tr>
<td>22</td>
<td>decide</td>
<td>Will</td>
<td>2006/08/30 00:00:00</td>
</tr>
<tr>
<td>23</td>
<td>invite additional reviewer</td>
<td>Mike</td>
<td>2006/09/02 00:00:00</td>
</tr>
<tr>
<td>24</td>
<td>get review X</td>
<td>Sara</td>
<td>2006/09/27 00:00:00</td>
</tr>
<tr>
<td>25</td>
<td>get review 3</td>
<td>John</td>
<td>2006/10/03 00:00:00</td>
</tr>
<tr>
<td>26</td>
<td>decide</td>
<td>Will</td>
<td>2006/11/10 00:00:00</td>
</tr>
<tr>
<td>27</td>
<td>invite reviewers</td>
<td>Anne</td>
<td>2006/12/02 00:00:00</td>
</tr>
<tr>
<td>28</td>
<td>get review X</td>
<td>Sara</td>
<td>2006/12/29 00:00:00</td>
</tr>
<tr>
<td>29</td>
<td>get review 2</td>
<td>Sara</td>
<td>2007/01/22 00:00:00</td>
</tr>
<tr>
<td>30</td>
<td>time-out X</td>
<td>INVALID</td>
<td>2007/02/22 00:00:00</td>
</tr>
<tr>
<td>31</td>
<td>decide</td>
<td>Will</td>
<td>2007/03/02 00:00:00</td>
</tr>
<tr>
<td>32</td>
<td>invite additional reviewer</td>
<td>Mike</td>
<td>2007/03/08 00:00:00</td>
</tr>
<tr>
<td>33</td>
<td>get review X</td>
<td>Sara</td>
<td>2007/04/13 00:00:00</td>
</tr>
<tr>
<td>34</td>
<td>get review 3</td>
<td>John</td>
<td>2007/05/17 00:00:00</td>
</tr>
<tr>
<td>35</td>
<td>decide</td>
<td>Will</td>
<td>2007/06/28 00:00:00</td>
</tr>
<tr>
<td>36</td>
<td>invite reviewers</td>
<td>Anne</td>
<td>2007/07/20 00:00:00</td>
</tr>
<tr>
<td>37</td>
<td>get review X</td>
<td>Sara</td>
<td>2007/08/03 00:00:00</td>
</tr>
<tr>
<td>38</td>
<td>get review 2</td>
<td>Sara</td>
<td>2007/08/22 00:00:00</td>
</tr>
</tbody>
</table>
Inspecting the 100 cases
Events per case
Resources involved
Concurrent process model:

- Concurrency is not discovered properly.
- Often additional review(s) needed due to timeouts.
Delays
Filter to investigate the accepted cases
Process model for accepted cases
Another data set: Vestia
Different types of process-related statistics
Quick process maps (informal semantics)
Seamless control over level of detail
Various filters
Animation
Celonis
Import data
Convert

loaded data set

mapping (3x)
Workspace

new analysis process

Reviewing222

Analysis Models

Here you can add your new analysis models.

New dashboard  New process analysis  Import analysis
Open analysis model
Process discovery (initial model)
Same model better layout

unable to discover concurrency properly
Frequencies
Time passing along edges
Accepted cases only (select "flows passing here")
Animation by replaying event log on model
Another event log: BPI challenge 2013
Filtering based on frequency
Even more filtering
Performance dashboard
Creating a "new process"
Three steps

1. **Create Process**
   - Create a process that will contain the event data and the analyses based on that data.

2. **Import Data**
   - Import data from your business system into the process. Choose the import format first.

3. **Analyze**
   - Analyze the data. Discover the process model and replay the case flow through the model.
Converting a CSV file into a data set
Initial model using simple miner

concurrency is not discovered properly
Information on frequencies
More refined model

better, but still concurrency is not discovered properly
Replay
Tokens colored based on flowtime
Save movie
Export as a flash file
Filtering based on frequency
Process model for cases handled within the first year
Social network including frequencies
Filter on top-3 resources
Replay on social network
Various types of charts

- Activity count
Another event log: Importing the CSV file

![Image of CSV import in Perceptive Process Mining]

- **Special Columns**
  - **Case ID column**: A
  - **Activity column**: B
  - **Timestamp column**: C
  - **Activity start timestamp**: Do not use
  - **Day Month Year**
  - **Year Month Day**

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Activity</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>130562</td>
<td>016 Registreren huurpape…</td>
<td>08.04.2007 12:24:52</td>
</tr>
<tr>
<td>130562</td>
<td>036 Volledige toekomst…</td>
<td>10.04.2007 12:45:52</td>
</tr>
<tr>
<td>130562</td>
<td>056 Inplannen afstapplaats…</td>
<td>10.04.2007 12:47:56</td>
</tr>
<tr>
<td>130562</td>
<td>000 Aantalen bevestigen…</td>
<td>17.10.2007 14:24:34</td>
</tr>
<tr>
<td>130562</td>
<td>070 Le± te inspectie uitvoeren…</td>
<td>01.11.2007 13:00:58</td>
</tr>
<tr>
<td>130562</td>
<td>106 Gereedheid te klaar…</td>
<td>05.11.2007 09:09:59</td>
</tr>
<tr>
<td>130562</td>
<td>116 Registreren werkwoord…</td>
<td>05.11.2007 09:00:28</td>
</tr>
<tr>
<td>130562</td>
<td>126 Plannen endissectie…</td>
<td>09.11.2007 09:00:28</td>
</tr>
<tr>
<td>130562</td>
<td>136 Is het opleverings…</td>
<td>06.11.2007 11:55:36</td>
</tr>
<tr>
<td>130562</td>
<td>156 Is er sprake van ZAV…</td>
<td>06.11.2007 11:55:54</td>
</tr>
<tr>
<td>130562</td>
<td>176 Aanpassen plaatplaat…</td>
<td>05.11.2007 12:05:20</td>
</tr>
<tr>
<td>130562</td>
<td>186 Aanpassen woning…</td>
<td>05.11.2007 19:50:27</td>
</tr>
<tr>
<td>130562</td>
<td>196 Actualiseren huurpape…</td>
<td>09.11.2007 11:20:20</td>
</tr>
<tr>
<td>130562</td>
<td>230 Toewijzen woning…</td>
<td>09.11.2007 11:50:41</td>
</tr>
<tr>
<td>130562</td>
<td>216 Registreren voet…</td>
<td>05.02.2008 09:41:45</td>
</tr>
<tr>
<td>130562</td>
<td>226 Is contract geënd…</td>
<td>06.02.2008 12:59:55</td>
</tr>
<tr>
<td>130562</td>
<td>240 Definitief maken huur…</td>
<td>05.03.2018 01:51:08</td>
</tr>
<tr>
<td>130562</td>
<td>250 Aanpassen faktuur…</td>
<td>09.02.2018 11:30:43</td>
</tr>
<tr>
<td>130562</td>
<td>290 After sales…</td>
<td>11.03.2007 11:50:26</td>
</tr>
<tr>
<td>130562</td>
<td>270 Archiefen nieuwe…</td>
<td>11.03.2007 09:51:45</td>
</tr>
<tr>
<td>130562</td>
<td>336 Is endissectie uitvoeren…</td>
<td>09.11.2007 13:00:09</td>
</tr>
<tr>
<td>130562</td>
<td>349 Zijn er nieuwe of niet…</td>
<td>09.11.2007 11:50:26</td>
</tr>
<tr>
<td>130562</td>
<td>400 Beoordelingenlijst…</td>
<td>09.11.2007 11:50:58</td>
</tr>
<tr>
<td>130562</td>
<td>416 Le± opleveringsformule…</td>
<td>09.11.2007 11:57:02</td>
</tr>
<tr>
<td>130562</td>
<td>436 Aanmelden werknemer…</td>
<td>09.11.2007 11:57:15</td>
</tr>
<tr>
<td>130562</td>
<td>446 Wijsten er beurs…</td>
<td>09.11.2007 11:57:20</td>
</tr>
<tr>
<td>130562</td>
<td>480 Vastleggen bruikbaar…</td>
<td>12.11.2007 12:21:40</td>
</tr>
<tr>
<td>130562</td>
<td>495 Vanleggen bruikbaar…</td>
<td>12.11.2007 12:21:40</td>
</tr>
<tr>
<td>130562</td>
<td>606 Oostende aardza…</td>
<td>03.12.2007 11:18:31</td>
</tr>
</tbody>
</table>
Model with 100% fitness
If you set this parameter to 40%, the mining process focuses on finding the simplest model that represents 40 percent of the cases.
The x% metric is not optimal as it looks at fitness at the case level rather than the event level. Later we will consider fitness metrics at the event level.
Process model for 5 selected activities
conclusion and outlook
Data and tools

“world” includes: business processes, people, machines, components, organizations.

Software system supports/controls the "world".

Models and analyzes: (process) model.

Event logs: records events, e.g., messages, transactions, etc.

Discovery and conformance: software system enhances the "world".

Data and tool models:

- Order
  - OrderID, Customer, Amount, Created, Paid, Completed

- Orderline
  - Product, NofItems, TotalWeight, Entered, BackOrdered, Secured

- Delivery
  - DelID, DelAddress, Contact

- Attempt
  - DelID, When, Successful

Data Quality
### Outlook

<table>
<thead>
<tr>
<th><strong>Part I: Preliminaries</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2</td>
</tr>
<tr>
<td>Chapter 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Part II: From Event Logs to Process Models</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 4</td>
</tr>
<tr>
<td>Chapter 5</td>
</tr>
<tr>
<td>Chapter 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Part III: Beyond Process Discovery</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 7</td>
</tr>
<tr>
<td>Chapter 8</td>
</tr>
<tr>
<td>Chapter 9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Part IV: Putting Process Mining to Work</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 10</td>
</tr>
<tr>
<td>Chapter 11</td>
</tr>
<tr>
<td>Chapter 12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Part V: Reflection</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 13</td>
</tr>
<tr>
<td>Chapter 14</td>
</tr>
</tbody>
</table>