Goal 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?
Getting the data

“world”
- people
- machines
- components
- organizations

business processes

models analyzes

specifies configures implements analyzes

supports/controls

software system

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

(process) model

event logs

discovery

conformance

enhancement
From heterogeneous data sources to process mining results

**Diagram Description**

- **Extract, Transform, and Load (ELT)**:
  - Data source → ELT → Data warehouse
  - Optional

- **Extract**
  - Data warehouse → Coarse-grained scoping → XES, MXML, or similar

- **Filter**
  - Unfiltered event logs → Fine-grained scoping

- **Process Mining**
  - Discovery → Conformance → Enhancement

- **Results**
  - Filtered event logs → (process) models → Answers
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>
</tbody>
</table>

| 2       | 35654483     | 30-12-2010:11.32      | register request    | Mike     | 50   |
|         | 35654485     | 30-12-2010:12.12      | check ticket        | Mike     | 100  |
|         | 35654487     | 30-12-2010:14.16      | examine casually    | Pete     | 400  |
|         | 35654488     | 05-01-2011:11.22      | decide              | Sara     | 200  |
|         | 35654489     | 08-01-2011:12.05      | pay compensation    | Ellen    | 200  |

| 3       | 35654521     | 30-12-2010:14.32      | register request    | Pete     | 50   |
|         | 35654522     | 30-12-2010:15.06      | examine casually    | Mike     | 400  |
|         | 35654524     | 30-12-2010:16.34      | check ticket        | Ellen    | 100  |
|         | 35654525     | 06-01-2011:09.18      | decide              | Sara     | 200  |
|         | 35654526     | 06-01-2011:12.18      | reinitiate request  | Sara     | 200  |
|         | 35654527     | 06-01-2011:13.06      | examine thoroughly  | Sean     | 400  |
|         | 35654530     | 08-01-2011:11.43      | check ticket        | Pete     | 100  |
|         | 35654531     | 09-01-2011:09.55      | decide              | Sara     | 200  |
|         | 35654533     | 15-01-2011:10.45      | pay compensation    | Ellen    | 200  |

| 4       | 35654641     | 06-01-2011:15.02      | register request    | Pete     | 50   |
|         | 35654643     | 07-01-2011:12.06      | check ticket        | Mike     | 100  |
|         | 35654644     | 08-01-2011:14.43      | examine thoroughly  | Sean     | 400  |
|         | 35654645     | 09-01-2011:12.02      | decide              | Sara     | 200  |
|         | 35654647     | 12-01-2011:15.44      | reject request      | Ellen    | 200  |
Another view

<table>
<thead>
<tr>
<th>case id</th>
<th>event id</th>
<th>timestamp</th>
<th>properties</th>
<th>attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>activity</td>
<td>resource</td>
</tr>
<tr>
<td>1</td>
<td>35654423</td>
<td>30-12-2010:11.02</td>
<td>register request</td>
<td>Pete</td>
</tr>
<tr>
<td>1</td>
<td>35654424</td>
<td>31-12-2010:10.06</td>
<td>examine thoroughly</td>
<td>Sue</td>
</tr>
<tr>
<td>1</td>
<td>35654425</td>
<td>05-01-2011:15.12</td>
<td>check ticket</td>
<td>Mike</td>
</tr>
<tr>
<td>1</td>
<td>35654426</td>
<td>06-01-2011:11.18</td>
<td>decide</td>
<td>Sara</td>
</tr>
<tr>
<td>1</td>
<td>35654427</td>
<td>07-01-2011:14.24</td>
<td>reject request</td>
<td>Pete</td>
</tr>
<tr>
<td>2</td>
<td>35654483</td>
<td>30-12-2010:11.32</td>
<td>register request</td>
<td>Mike</td>
</tr>
<tr>
<td>2</td>
<td>35654485</td>
<td>30-12-2010:12.12</td>
<td>check ticket</td>
<td>Mike</td>
</tr>
<tr>
<td>2</td>
<td>35654487</td>
<td>30-12-2010:14.16</td>
<td>examine casually</td>
<td>Pete</td>
</tr>
<tr>
<td>2</td>
<td>35654488</td>
<td>05-01-2011:11.22</td>
<td>decide</td>
<td>Sara</td>
</tr>
<tr>
<td>2</td>
<td>35654489</td>
<td>08-01-2011:12.05</td>
<td>pay compensation</td>
<td>Ellen</td>
</tr>
<tr>
<td>3</td>
<td>3565521</td>
<td>30-12-2010:14.32</td>
<td>register request</td>
<td>Pete</td>
</tr>
<tr>
<td>3</td>
<td>3565522</td>
<td>30-12-2010:15.06</td>
<td>examine casually</td>
<td>Mike</td>
</tr>
<tr>
<td>3</td>
<td>3565524</td>
<td>30-12-2010:16.34</td>
<td>check ticket</td>
<td>Ellen</td>
</tr>
<tr>
<td>3</td>
<td>3565525</td>
<td>06-01-2011:09.18</td>
<td>decide</td>
<td>Sara</td>
</tr>
<tr>
<td>3</td>
<td>3565526</td>
<td>06-01-2011:12.18</td>
<td>reininitiate request</td>
<td>Sara</td>
</tr>
<tr>
<td>3</td>
<td>3565527</td>
<td>06-01-2011:13.06</td>
<td>examine thoroughly</td>
<td>Sean</td>
</tr>
<tr>
<td>3</td>
<td>3565530</td>
<td>08-01-2011:11.43</td>
<td>check ticket</td>
<td>Pete</td>
</tr>
<tr>
<td>3</td>
<td>3565531</td>
<td>09-01-2011:09.55</td>
<td>decide</td>
<td>Sara</td>
</tr>
<tr>
<td>3</td>
<td>3565533</td>
<td>15-01-2011:10.45</td>
<td>pay compensation</td>
<td>Ellen</td>
</tr>
<tr>
<td>4</td>
<td>3565461</td>
<td>06-01-2011:15.02</td>
<td>register request</td>
<td>Pete</td>
</tr>
<tr>
<td>4</td>
<td>3565463</td>
<td>07-01-2011:12.06</td>
<td>check ticket</td>
<td>Mike</td>
</tr>
<tr>
<td>4</td>
<td>3565464</td>
<td>08-01-2011:14.43</td>
<td>examine thoroughly</td>
<td>Sean</td>
</tr>
<tr>
<td>4</td>
<td>3565465</td>
<td>09-01-2011:12.02</td>
<td>decide</td>
<td>Sara</td>
</tr>
<tr>
<td>4</td>
<td>3565467</td>
<td>12-01-2011:15.44</td>
<td>reject request</td>
<td>Ellen</td>
</tr>
</tbody>
</table>
Standard transactional life-cycle model

- **Schedule**
- **Assign**
- **Reassign**
- **Start**
- **Resume**
- **Suspend**

- **Autoskip**
- **Manualskip**
- **Successful**
- **Termination**
- **Complete**

- **Abort_activity**
- **Withdraw**
- **Abort_case**
- **Unsuccessful**
- **Termination**
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

The diagram shows the workflow of activity instances with transitions for schedule, assign, reassign, start, suspend, resume, abort_activity, and complete actions.
Overlapping activity instances

\[ \text{a: start complete complete} \]
\[ \text{5 hours} \]
\[ \text{6 hours} \]

\[ \text{a: start start complete complete} \]
\[ \text{2 hours} \]
\[ \text{9 hours} \]
Using attributes

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), Nitro, 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.
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
Order instance

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

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

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

Attempt
- When: DateTime
- Successful: Bool
<table>
<thead>
<tr>
<th>case id</th>
<th>activity</th>
<th>timestamp</th>
<th>other attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>91245</td>
<td>create order</td>
<td>28-11-2011:08.12</td>
<td>Customer: John, Amount: 100</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.13</td>
<td>OrderLineID: 112245, Product: iPhone 4G, NoItems: 1, TotalWeight: 0.250, DellID: 882345</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.14</td>
<td>OrderLineID: 112246, Product: iPod nano, NoItems: 2, TotalWeight: 0.300, DellID: 882346</td>
</tr>
<tr>
<td>91245</td>
<td>enter order line</td>
<td>28-11-2011:08.15</td>
<td>OrderLineID: 112247, Product: iPod classic, NoItems: 1, TotalWeight: 0.200, DellID: 882345</td>
</tr>
<tr>
<td>91245</td>
<td>secure order line</td>
<td>28-11-2011:08.55</td>
<td>OrderLineID: 112245, Product: iPhone 4G, NoItems: 1, TotalWeight: 0.250, DellID: 882345</td>
</tr>
<tr>
<td>91245</td>
<td>create backorder</td>
<td>28-11-2011:08.55</td>
<td>OrderLineID: 112246, Product: iPod nano, NoItems: 2, TotalWeight: 0.300, DellID: 882346</td>
</tr>
<tr>
<td>91245</td>
<td>pay order</td>
<td>02-12-2011:13.45</td>
<td>Customer: John, Amount: 100</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>05-12-2011:08.43</td>
<td>DellID: 882346, Successful: true, DelAddress: 5513XG-45, Contact: 040-2298761</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>05-12-2011:08.55</td>
<td>DellID: 882345, Successful: false, DelAddress: 5513VJ-22a, Contact: 0497-2553660</td>
</tr>
<tr>
<td>91245</td>
<td>complete order</td>
<td>05-12-2011:11.33</td>
<td>Customer: John, Amount: 100</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>06-12-2011:09.12</td>
<td>DellID: 882345, Successful: false, DelAddress: 5513VJ-22a, Contact: 0497-2553660</td>
</tr>
<tr>
<td>91245</td>
<td>delivery attempt</td>
<td>07-12-2011:08.56</td>
<td>DellID: 882345, Successful: true, DelAddress: 5513VJ-22a, Contact: 0497-2553660</td>
</tr>
<tr>
<td>91561</td>
<td>create order</td>
<td>28-11-2011:12.22</td>
<td>Customer: Mike, Amount: 530</td>
</tr>
<tr>
<td>91561</td>
<td>enter order line</td>
<td>28-11-2011:12.23</td>
<td>OrderLineID: 112248, Product: iPhone 4G, NoItems: 1, TotalWeight: 0.250, DellID: 882345</td>
</tr>
</tbody>
</table>

...
Orderline instance

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

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

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

**Attempt**
- DellID: DellID
- When: DateTime
- Successful: Bool

**Order**
- Case id: 112345
- Activity: create order
  - Timestamp: 28-11-2011:08.12
- Customer: John
- Amount: 100

**Orderline**
- Case id: 112345
- Activity: enter order line
  - Timestamp: 28-11-2011:08.13
- Product: iPhone 4G
- NofItems: 1
- TotalWeight: 0.250

**Delivery**
- Case id: 112345
  - Activity: delivery attempt
    - Timestamp: 05-12-2011:08.56
    - Successful: true
    - DelAddress: 5513VJ-22a
    - Contact: 0497-2553660
Other examples

- The life cycles of reviewers, authors, papers, reviews, PC chairs, etc.
- The life cycles of job applications and vacancies.
- X-ray machine logs: machine, machine day, patient, treatment, routine, etc.?

- Therefore, the selection and scoping of instances is needed.
- Like making deciding on the elements to be put on map; there may be many maps covering partially overlapping areas.
Extracting event logs

• Not just a syntactical issue.
• Different views are possible.
• Important:
  – Selecting the right instance notion.
  – Ordering of events.
  – Selection of events.
• Proclets: the true fabric of real-life processes.