

Event Logs

What kind of data does
process mining require?

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

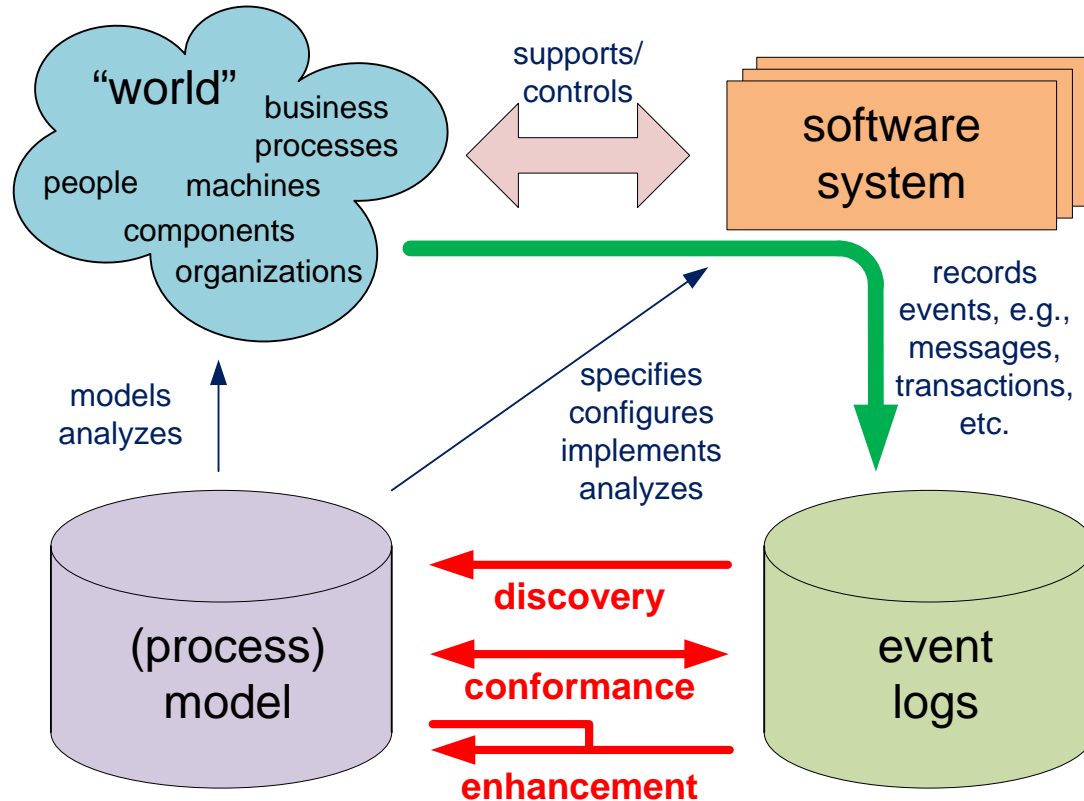


TU/e

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Getting the "right" data ...



Event log

- We assume the existence of an **event log** where each **event** refers to a **case**, an **activity**, and a point in **time**.
- An **event log** can be seen as a **collection of cases**.
- A **case** can be seen as a **trace/sequence of events**.

event data can be found everywhere!

Event data may come from ...

- a database system (e.g., patient data in a hospital),
- a comma-separated values (CSV) file or spreadsheet,
- a transaction log (e.g., a trading system),
- a business suite/ERP system (SAP, Oracle, etc.),
- a message log (e.g., from IBM middleware),
- an open API providing data from websites or social media, ...

An example log

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8
...

case id

activity name

timestamp

other data

Another event log: order handling

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
...

case id

activity name

timestamp

resource

other data

Another event log: patient treatment

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	radiation therapy	23-1-2014@13.57	Dr. Jones	61	140.00
5541	radiation therapy	23-1-2014@13.08	Dr. Jones	61	140.00
...

case id

activity name

timestamp

resource

other data

Minimal requirements in terms of a CSV file

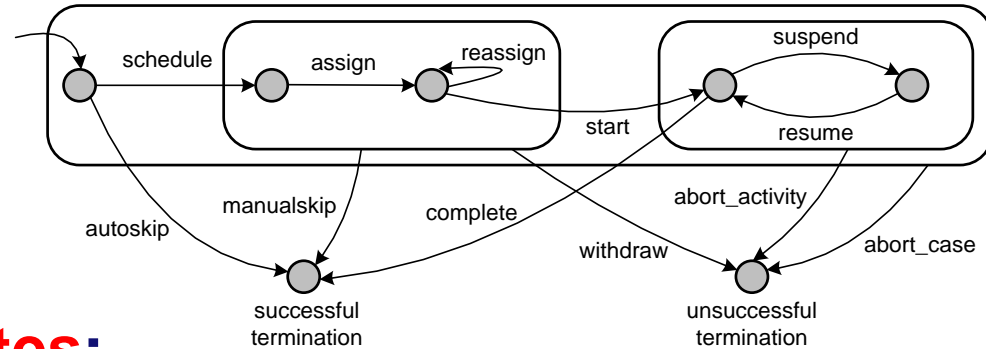
- Each row corresponds to an event.
- There are at least three columns:
 - Case id (patient id, order number, claim number, ...)
 - Activity name (approve, reject, request, send, ...)
 - Timestamp (2015-08-18T06:36:40, ...)
- There may be many other (optional) columns: resource, transaction type, age, costs, etc.

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2

Extensions

- **Transactional information on activity instances:**

An event can represent a **start, complete, suspend, resume, abort, etc.**



- **Case versus event attributes:**

- case attributes do not change, e.g., the birth date or gender of a patient,
- event attributes are related to a particular step in the process.

XES (eXtensible Event Stream)

- Adopted by the **IEEE Task Force on Process Mining**.
- The format is supported by tools such as **ProM** and **Disco** (used in this course).
- Predecessors: **MXML** and **SA-MXML**.
- Conversion from other formats (**CSV**) is easy if the right data are available.
- **XML syntax** and **OpenXES library** available.
- See **www.xes-standard.org**.



Extensible Event Stream

Simplistic view on event data

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
...

case id

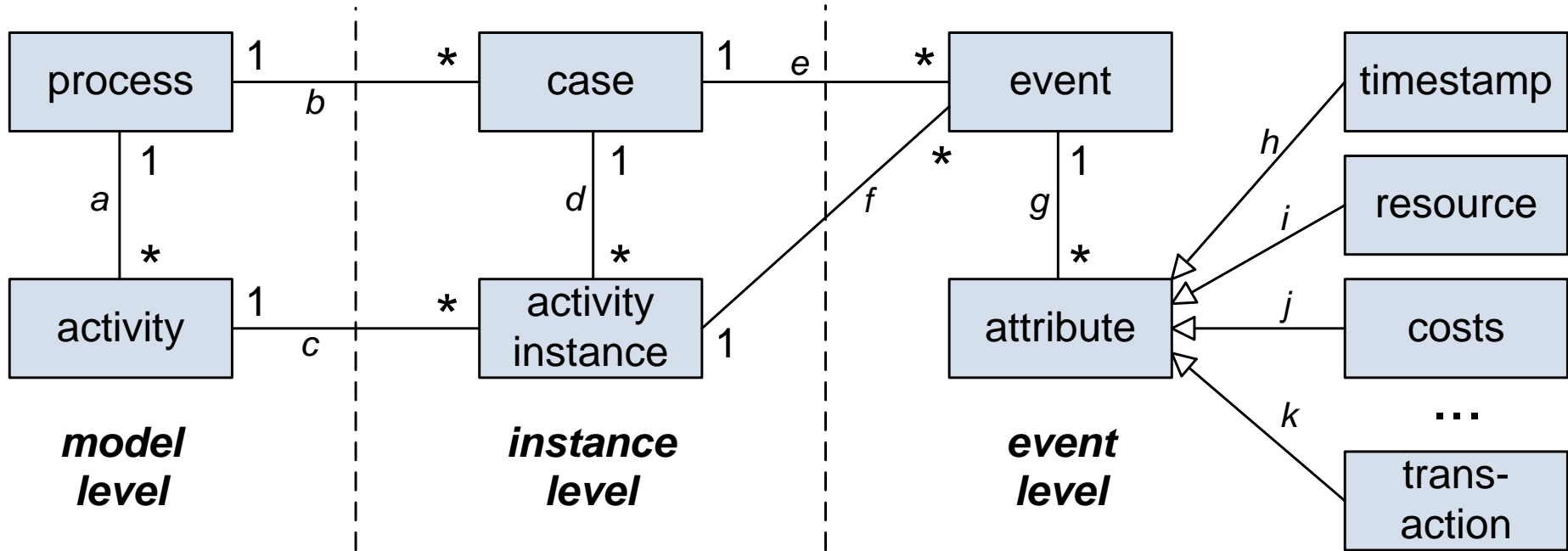
activity name

timestamp

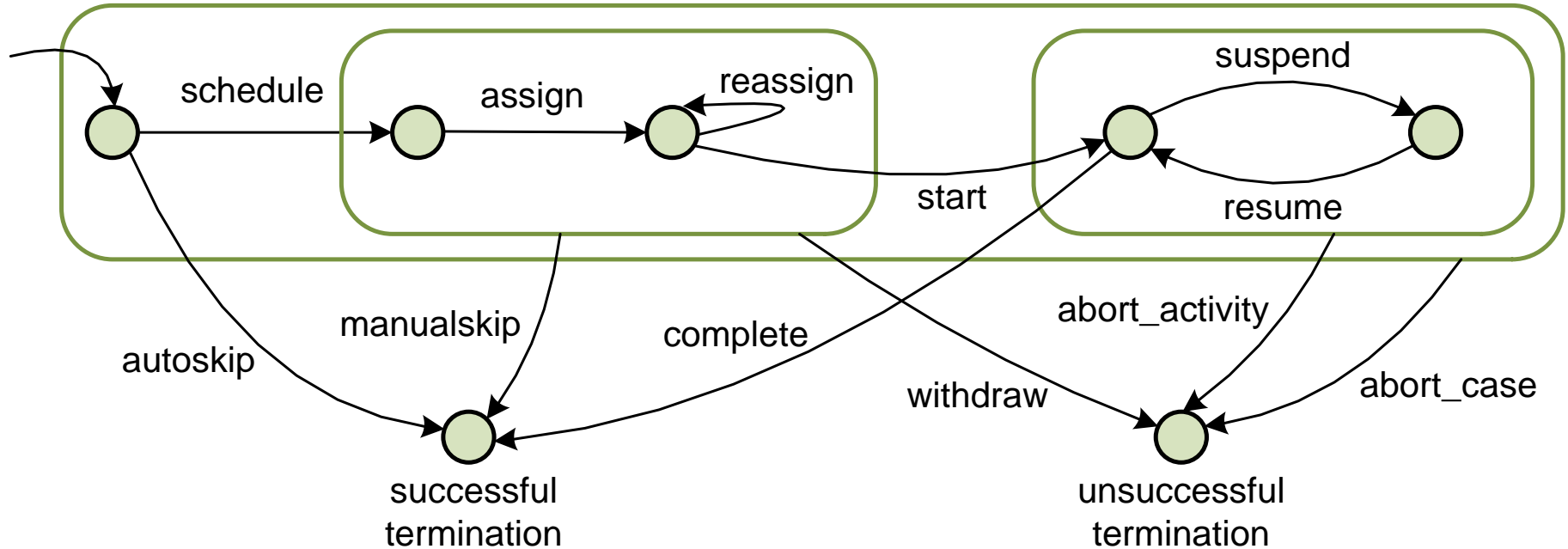
resource

other data

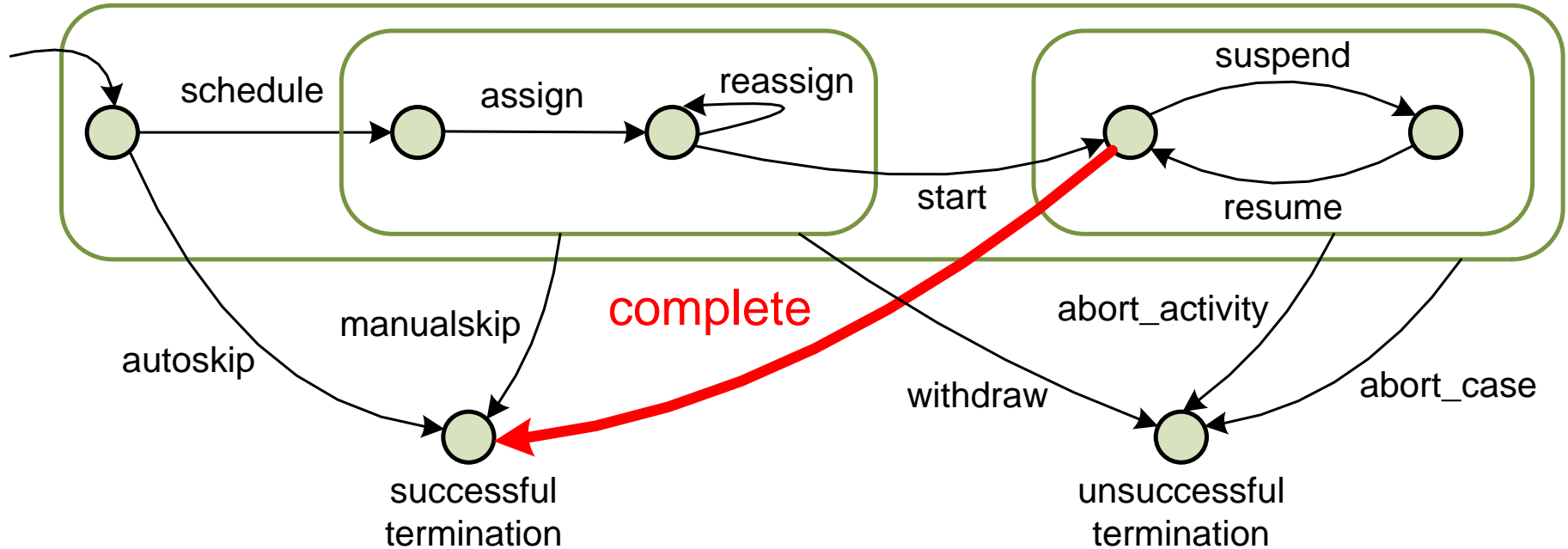
A more refined view



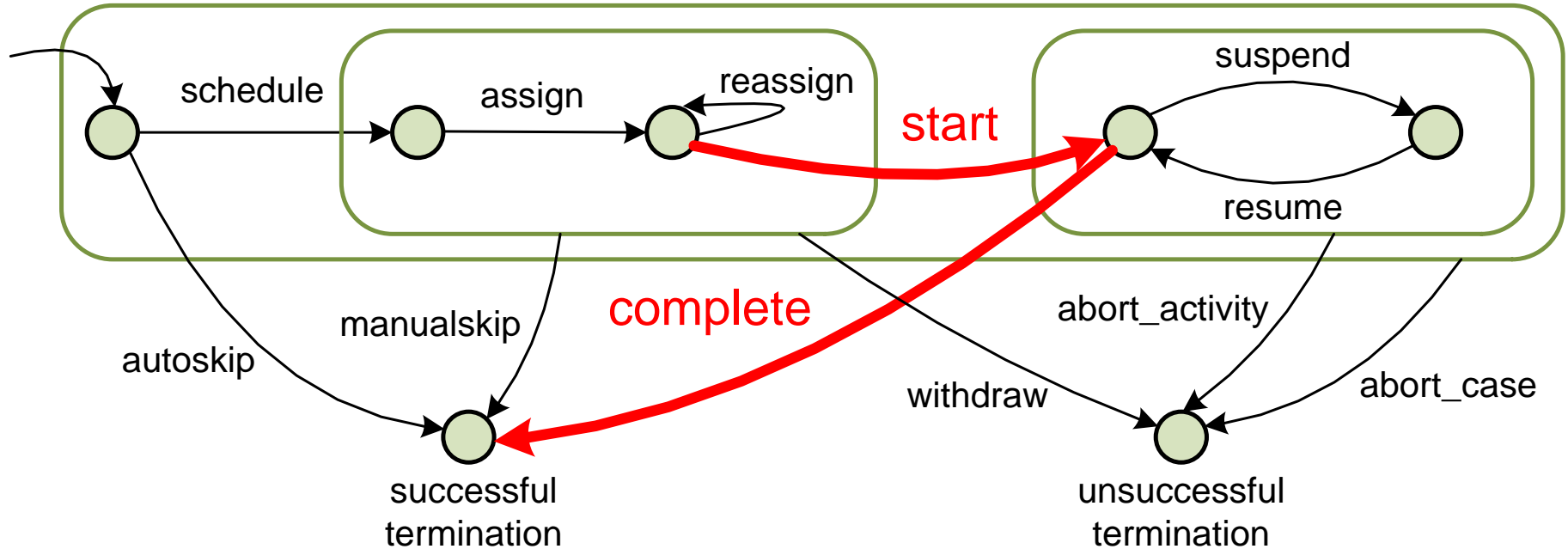
Transactional model for activities



Atomic activities

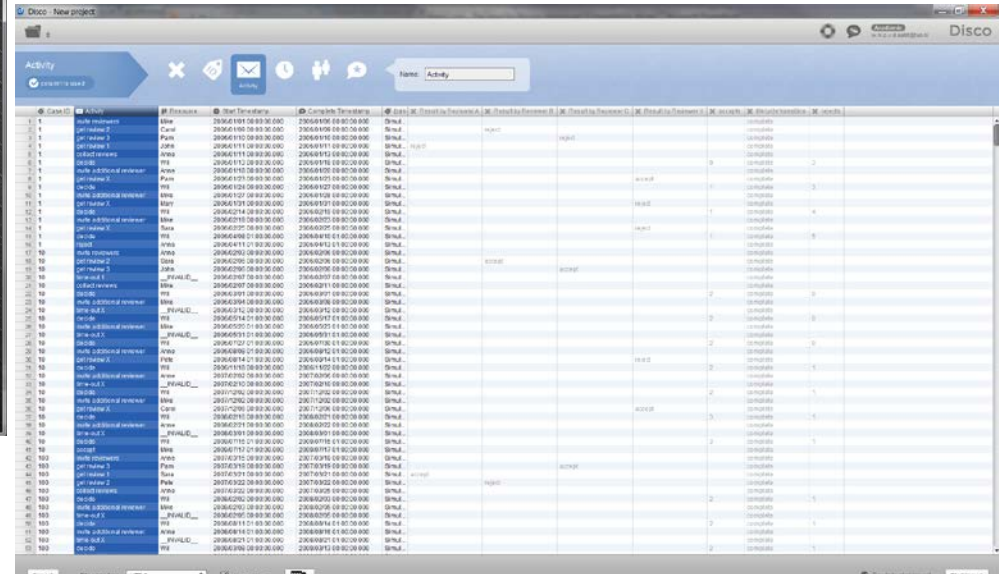
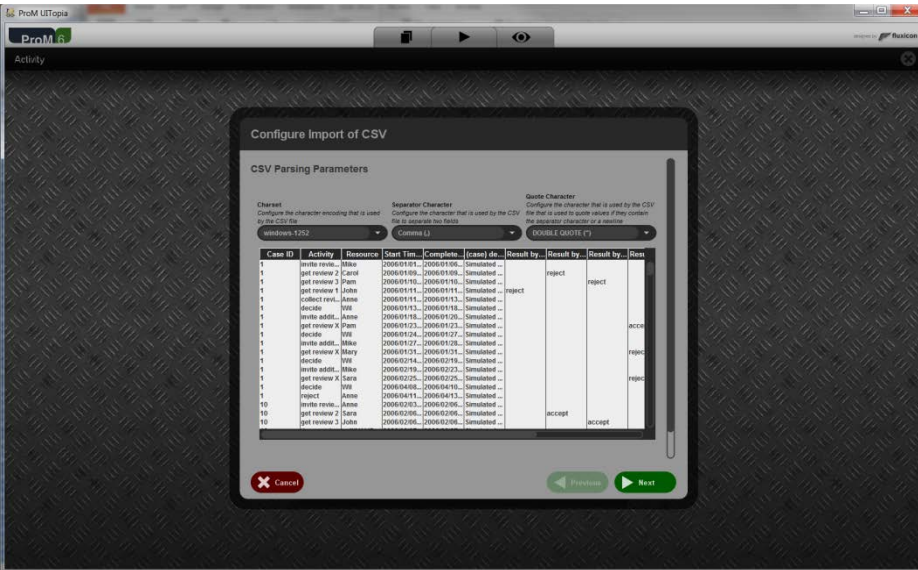


Activities that have a duration



Converting Event Data to XES

- Easy with tools like ProM and Disco!



Import CSV file, then apply
“Convert CSV to XES” plug-in.

Learn more about event logs

- **Extensive introduction:** Chapter 4 of Process Mining: Data Science in Action by W.M.P. van der Aalst, Springer Verlag, 2016 (ISBN 978-3-662-49850-7)
- **XES standard:** www.xes-standard.org
- **More advanced:** W.M.P. van der Aalst. Extracting Event Data from Databases to Unleash Process Mining. In BPM - Driving Innovation in a Digital World, Springer-Verlag, Berlin, 2015. http://dx.doi.org/10.1007/978-3-319-14430-6_8

Example logs files (XES)

- http://data.3tu.nl/repository/collection:event_logs_real
- http://data.3tu.nl/repository/collection:event_logs_synthetic
- http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:process_mining_logs
- <http://www.processmining.org/logs/start>

Event data are everywhere!
What is your excuse?