Process Mining

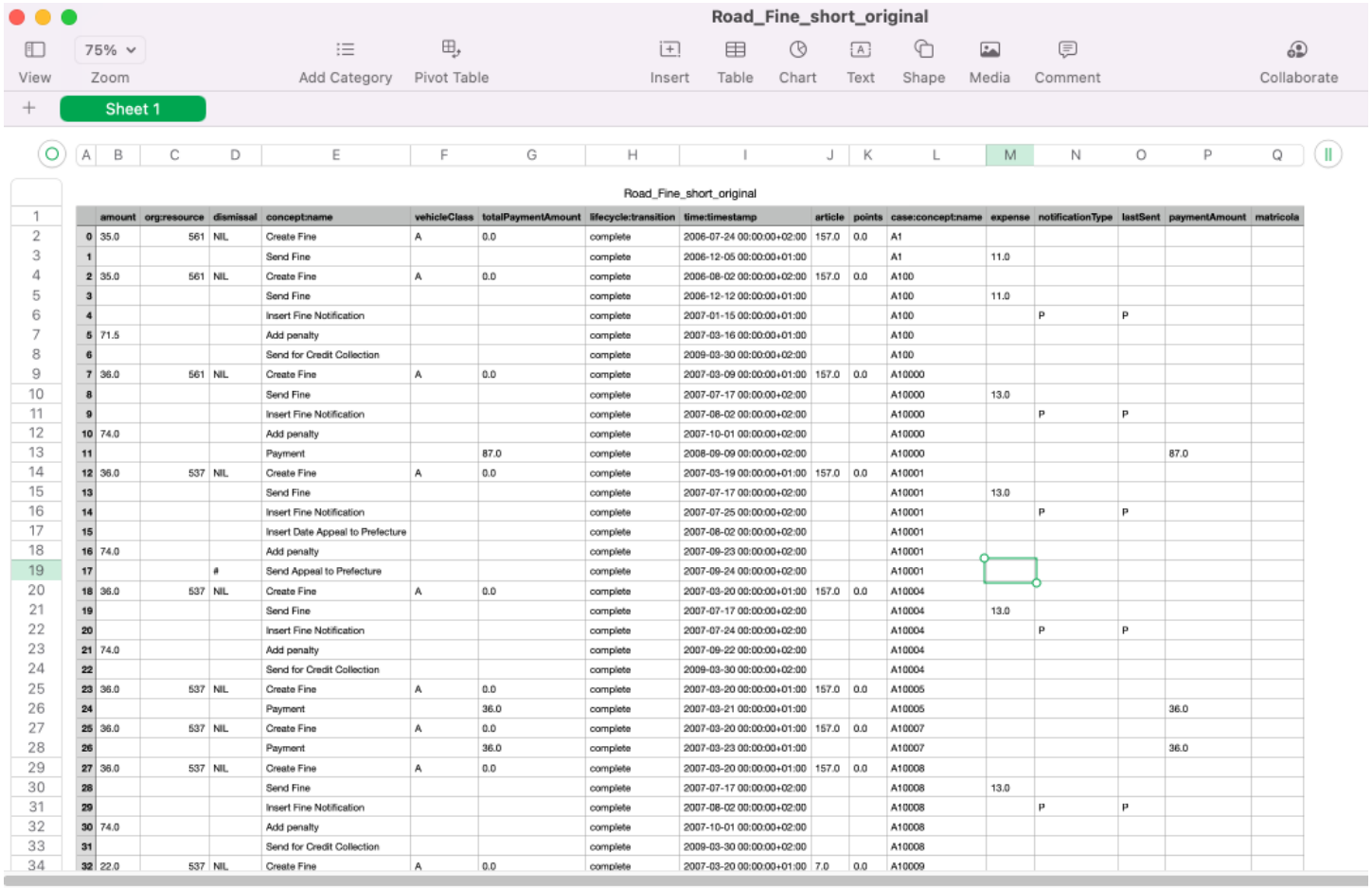
Tools and Techniques

Elisa Marengo

elisa.marengo@unito.it

University of Turin
Computer Science Department

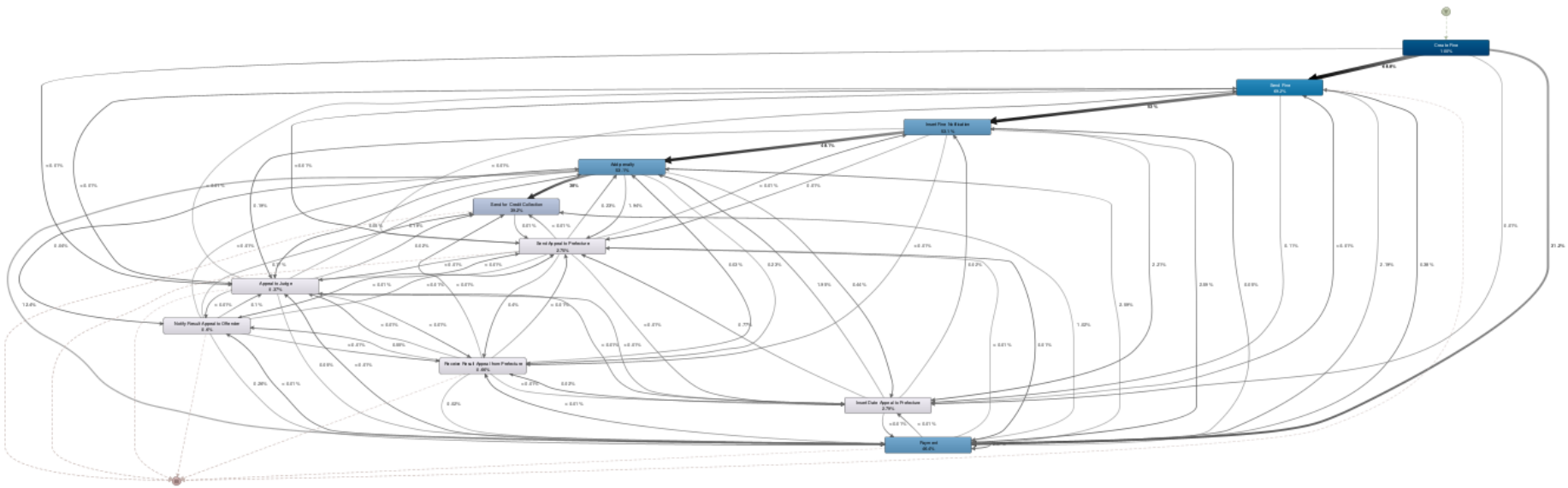
What do you think this is?



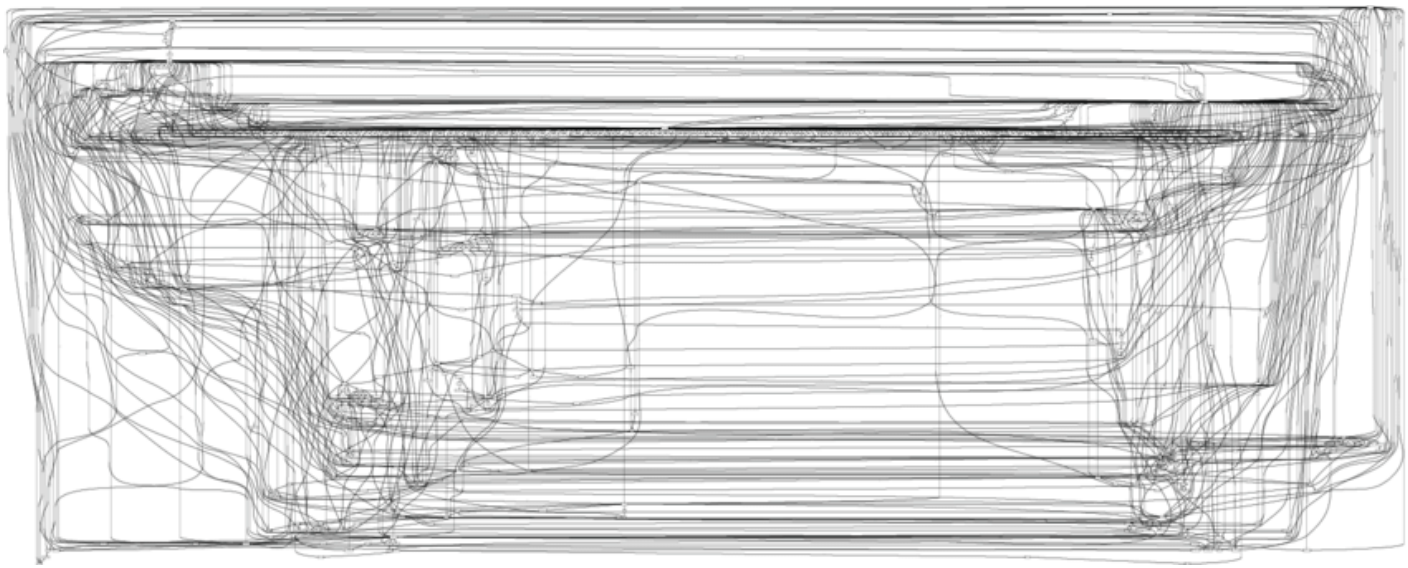
Road_Fine_short_original

	amount	org:resource	dismissal	concept:name	vehicleClass	totalPaymentAmount	lifecycle:transition	time:timestamp	article	points	case:concept:name	expense	notificationType	lastSent	paymentAmount	matricola
0	35.0	561	NIL	Create Fine	A	0.0	complete	2006-07-24 00:00:00+02:00	157.0	0.0	A1					
1				Send Fine			complete	2006-12-05 00:00:00+01:00			A1	11.0				
2	35.0	561	NIL	Create Fine	A	0.0	complete	2006-08-02 00:00:00+02:00	157.0	0.0	A100					
3				Send Fine			complete	2006-12-12 00:00:00+01:00			A100	11.0				
4				Insert Fine Notification			complete	2007-01-15 00:00:00+01:00			A100		P	P		
5	71.5			Add penalty			complete	2007-03-16 00:00:00+01:00			A100					
6				Send for Credit Collection			complete	2009-03-30 00:00:00+02:00			A100					
7	36.0	561	NIL	Create Fine	A	0.0	complete	2007-03-09 00:00:00+01:00	157.0	0.0	A10000					
8				Send Fine			complete	2007-07-17 00:00:00+02:00			A10000	13.0				
9				Insert Fine Notification			complete	2007-08-02 00:00:00+02:00			A10000		P	P		
10	74.0			Add penalty			complete	2007-10-01 00:00:00+02:00			A10000					
11				Payment		87.0	complete	2008-09-09 00:00:00+02:00			A10000				87.0	
12	36.0	537	NIL	Create Fine	A	0.0	complete	2007-03-19 00:00:00+01:00	157.0	0.0	A10001					
13				Send Fine			complete	2007-07-17 00:00:00+02:00			A10001	13.0				
14				Insert Fine Notification			complete	2007-07-25 00:00:00+02:00			A10001		P	P		
15				Insert Date Appeal to Prefecture			complete	2007-08-02 00:00:00+02:00			A10001					
16	74.0			Add penalty			complete	2007-09-23 00:00:00+02:00			A10001					
17			#	Send Appeal to Prefecture			complete	2007-09-24 00:00:00+02:00			A10001					
18	36.0	537	NIL	Create Fine	A	0.0	complete	2007-03-20 00:00:00+01:00	157.0	0.0	A10004					
19				Send Fine			complete	2007-07-17 00:00:00+02:00			A10004	13.0				
20				Insert Fine Notification			complete	2007-07-24 00:00:00+02:00			A10004		P	P		
21	74.0			Add penalty			complete	2007-09-22 00:00:00+02:00			A10004					
22				Send for Credit Collection			complete	2009-03-30 00:00:00+02:00			A10004					
23	36.0	537	NIL	Create Fine	A	0.0	complete	2007-03-20 00:00:00+01:00	157.0	0.0	A10005					
24				Payment		36.0	complete	2007-03-21 00:00:00+01:00			A10005				36.0	
25	36.0	537	NIL	Create Fine	A	0.0	complete	2007-03-20 00:00:00+01:00	157.0	0.0	A10007					
26				Payment		36.0	complete	2007-03-23 00:00:00+01:00			A10007				36.0	
27	36.0	537	NIL	Create Fine	A	0.0	complete	2007-03-20 00:00:00+01:00	157.0	0.0	A10008					
28				Send Fine			complete	2007-07-17 00:00:00+02:00			A10008	13.0				
29				Insert Fine Notification			complete	2007-08-02 00:00:00+02:00			A10008		P	P		
30	74.0			Add penalty			complete	2007-10-01 00:00:00+02:00			A10008					
31				Send for Credit Collection			complete	2009-03-30 00:00:00+02:00			A10008					
32	22.0	537	NIL	Create Fine	A	0.0	complete	2007-03-20 00:00:00+01:00	7.0	0.0	A10009					

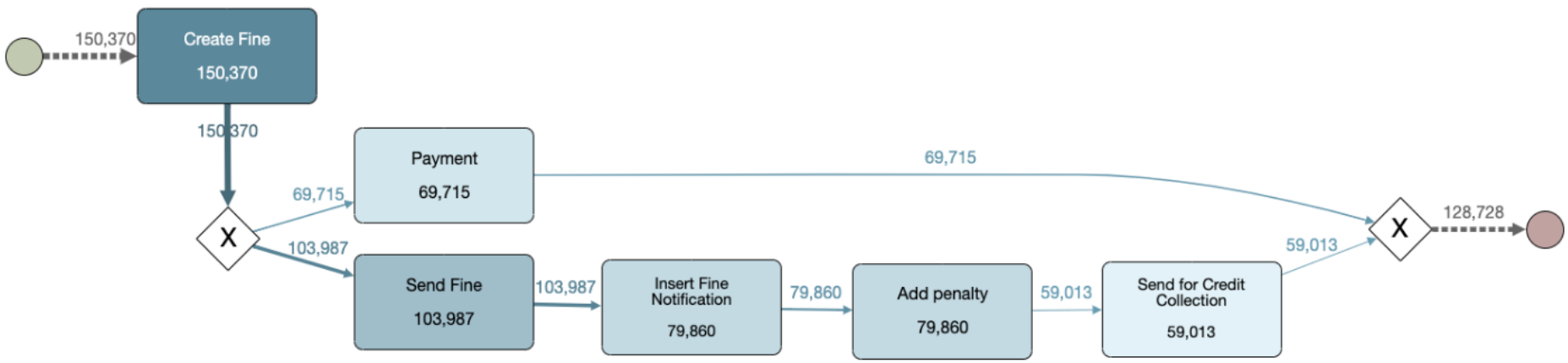
Is this also a process?



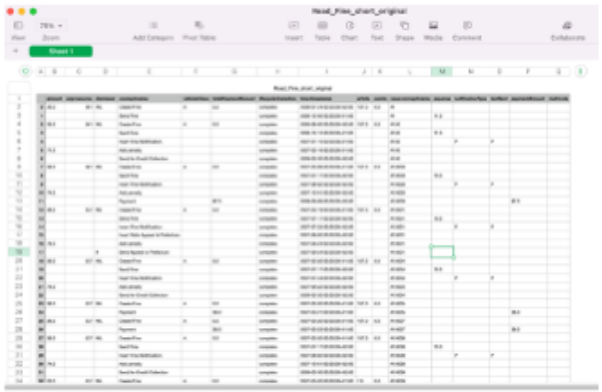
Which information can I get?

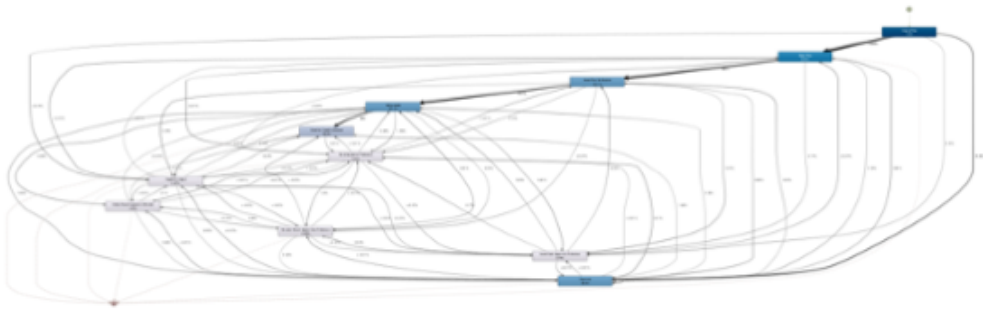


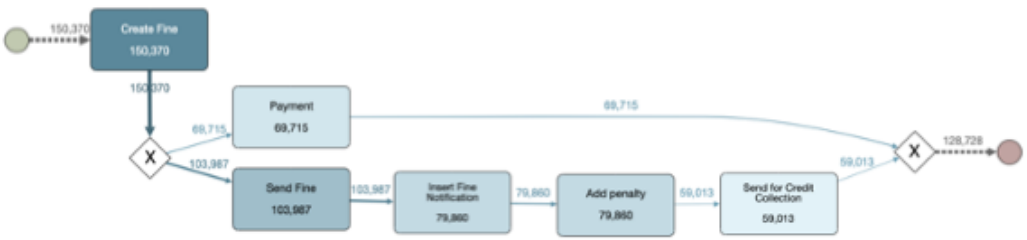
Wasn't a process represented like this??



So, what is Process Mining about?

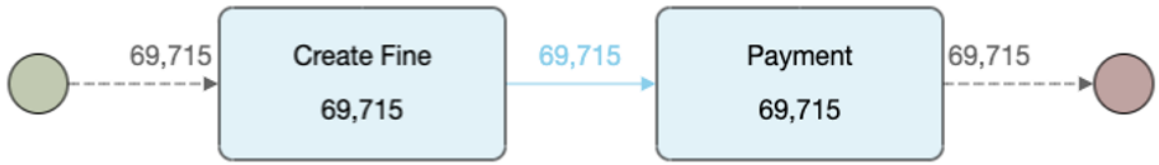




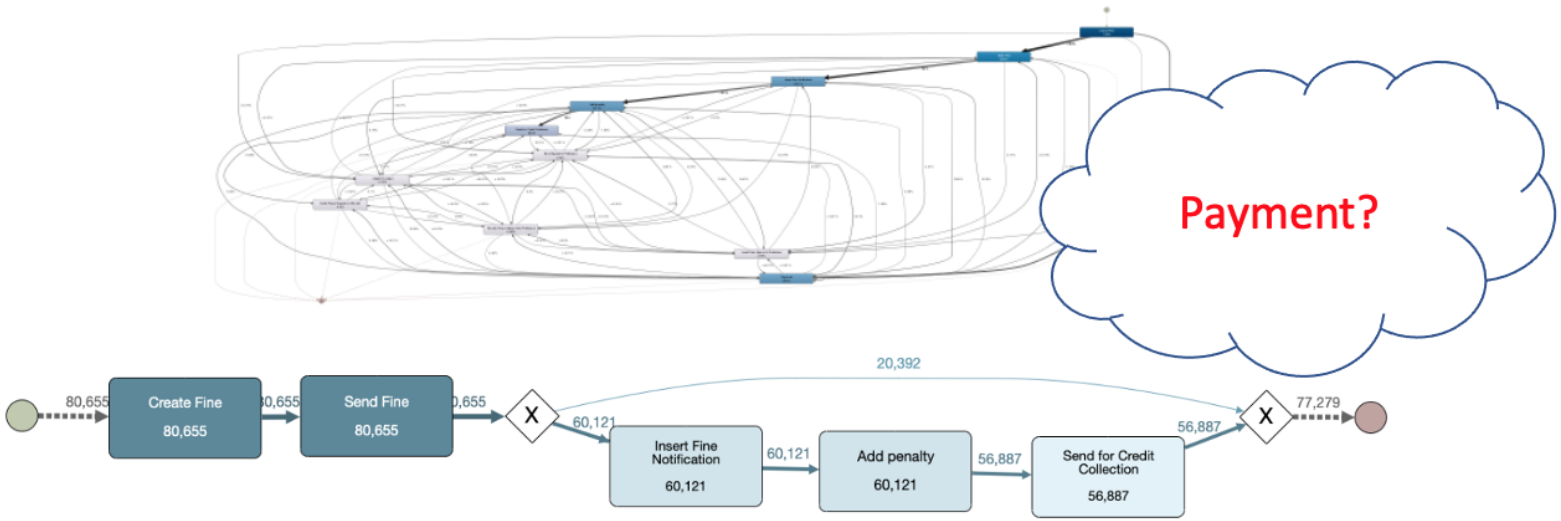


Why to start from real data?

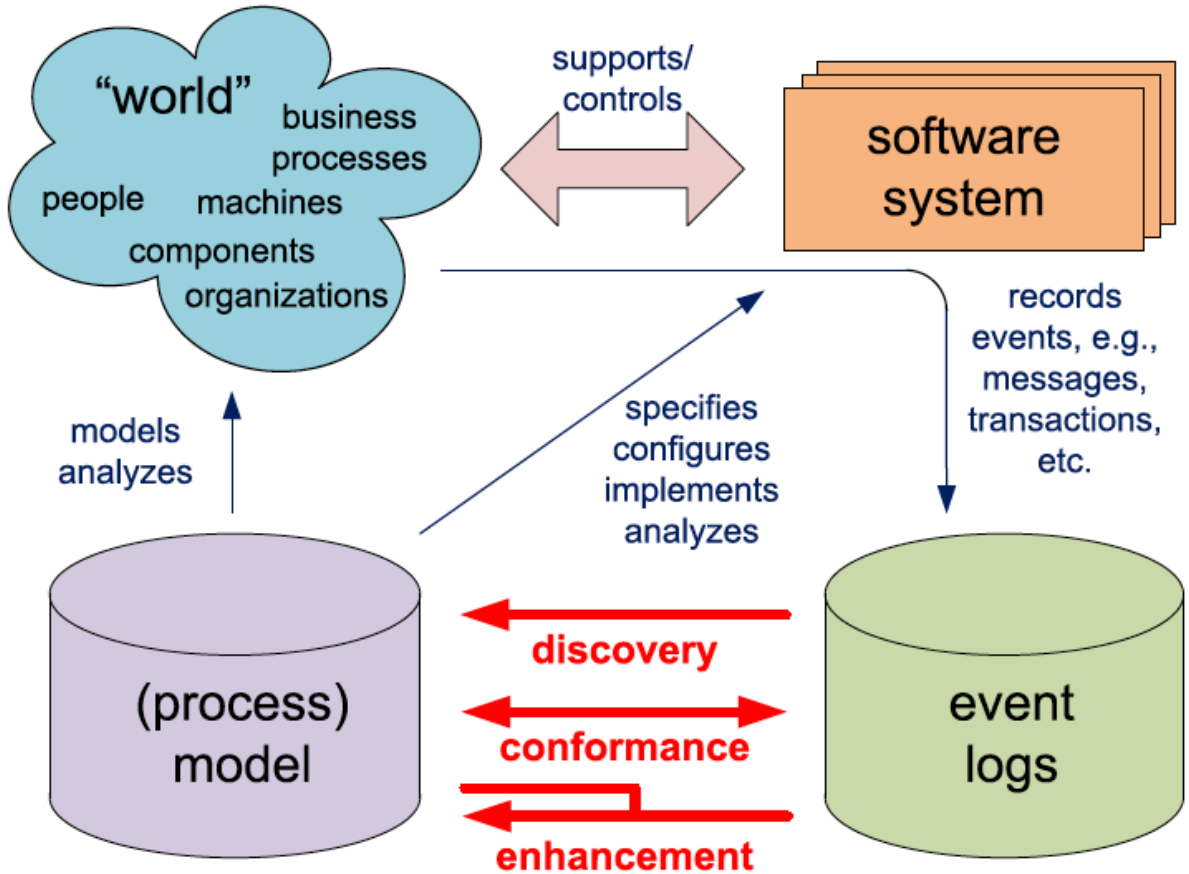
Ideal Process



Reality

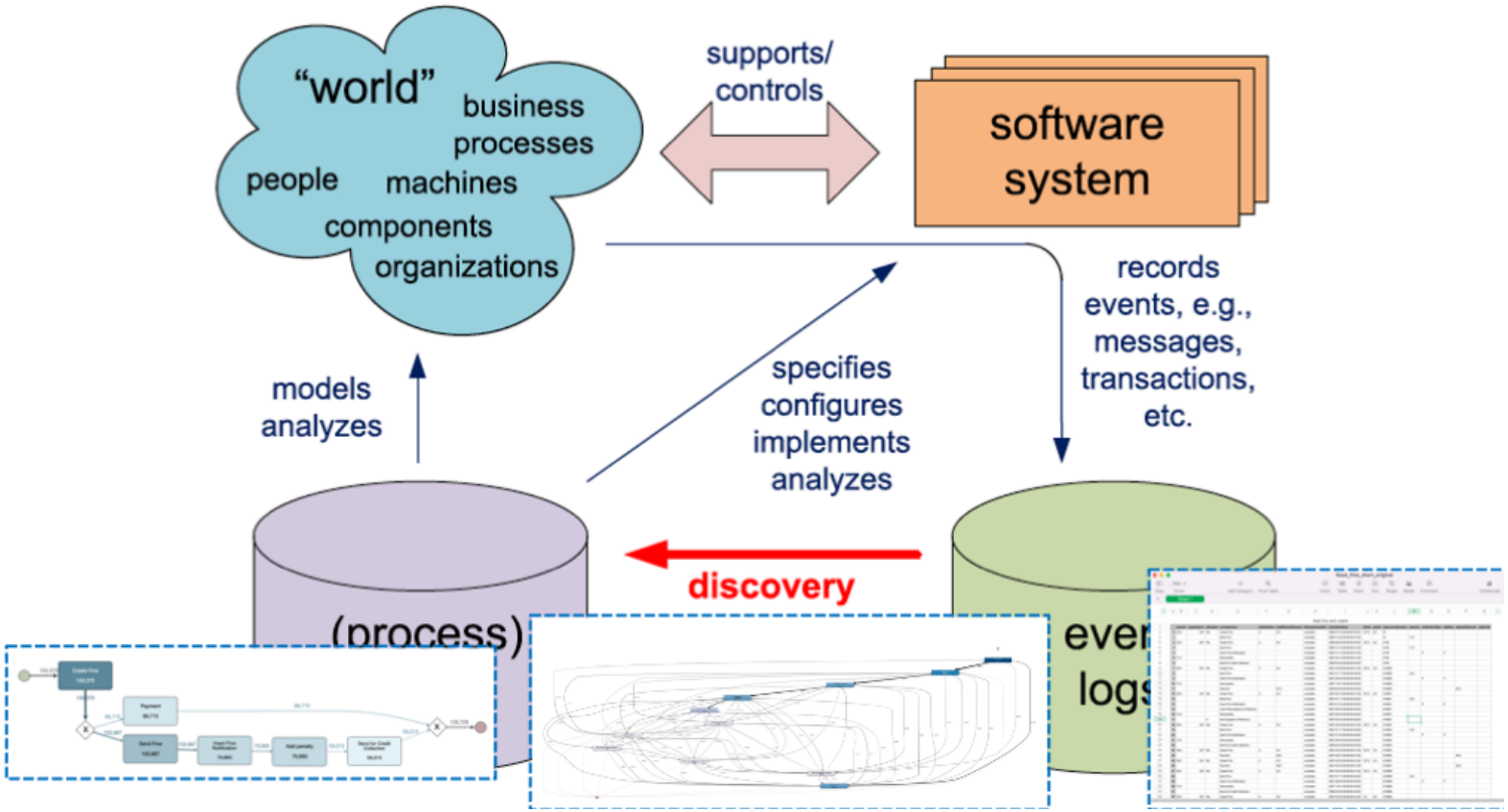


Process Mining[1]



[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Process Mining[1]



[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

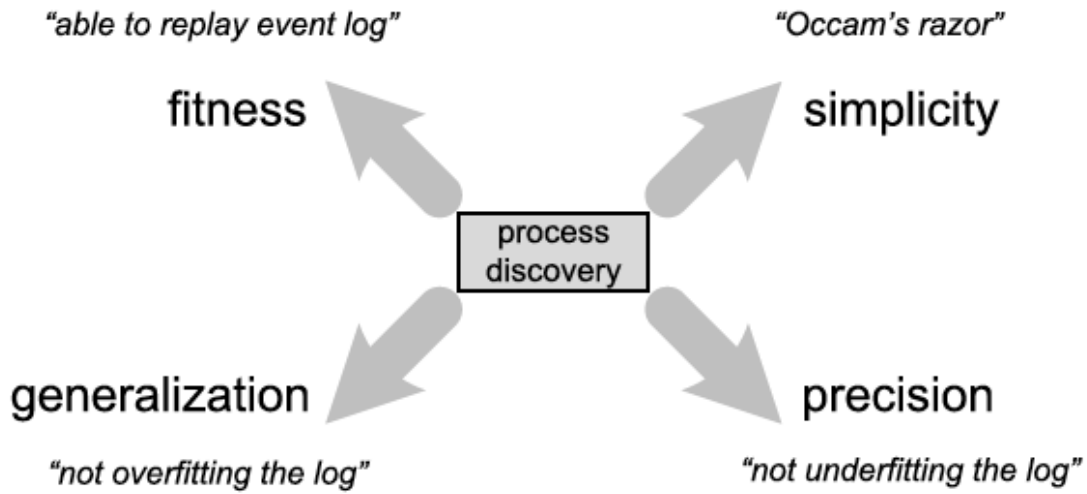
Recap on Process Mining[1]

The idea of process mining is to...

- discover, monitor and improve real processes (i.e., not assumed processes)
- by extracting knowledge from event logs readily available in today's systems

[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Process Mining Challenges[1]

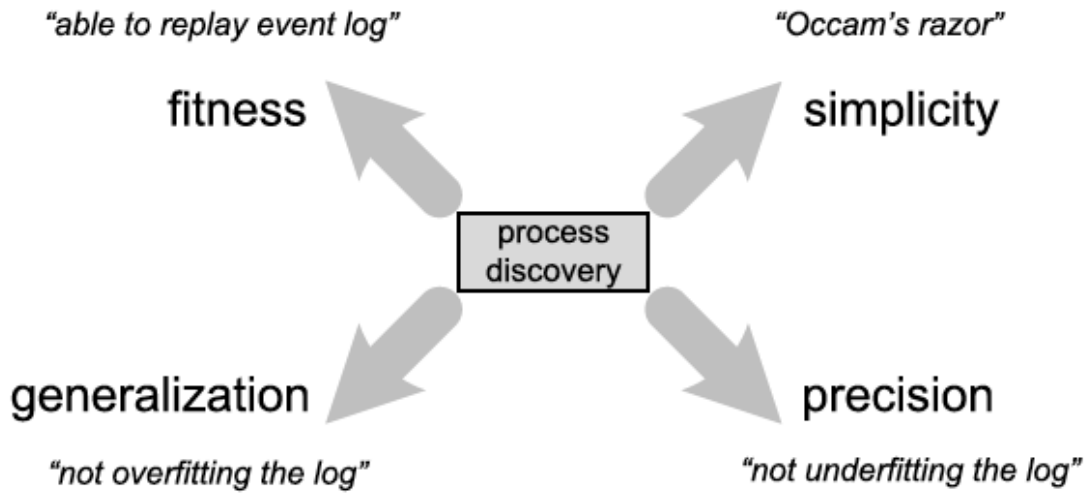


Fitness

The fraction of traces in the log that can be fully replayed.
The fraction of events in the log that are indeed possible according to the model.

[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Process Mining Challenges[1]

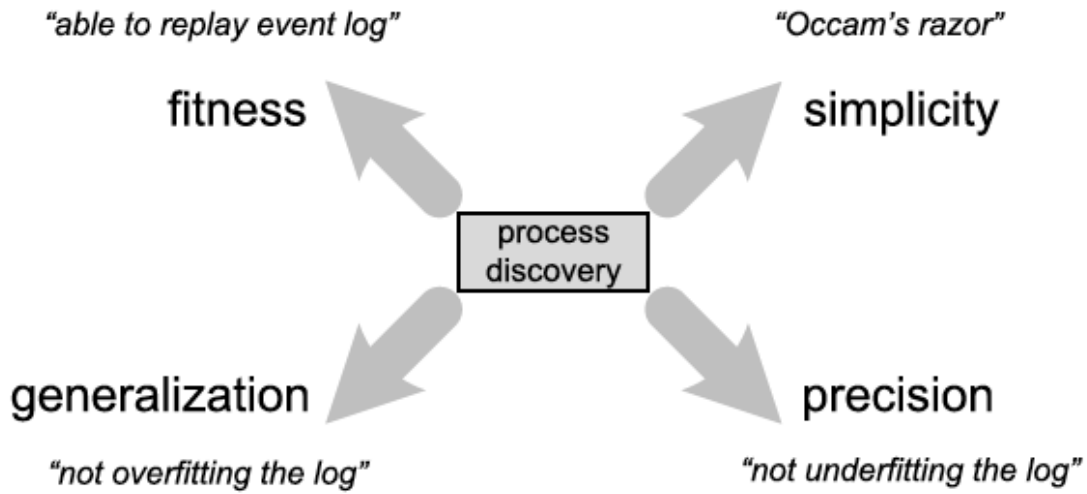


Simplicity

The simplest model that can explain the behavior seen in the log, is the best model. (number of nodes and arcs)

[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Process Mining Challenges[1]

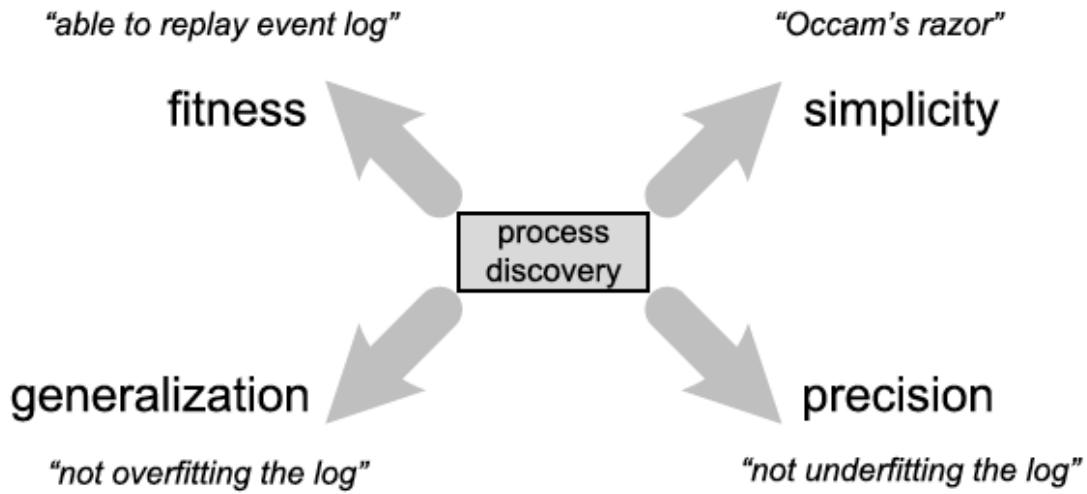


Precision

A model is precise if it does not allow for "too much" behavior. A model that is not precise is "underfitting", i.e., the model allows for behaviors very different from what was seen in the log.

[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Process Mining Challenges[1]



Generalization

A model that does not generalize is "overfitting": If the sequence is not in the event log, it is not possible!.

[1] Wil van der Aalst. Process Mining. Data Science in Action. Springer 2016

Why should this be important?

Celonis: company providing support to execution management

We pioneered the process mining category 10 years ago when we first developed the ability to automatically X-ray processes and find inefficiencies.[a]

[a] `celonis.com`

Perhaps higher performing companies aren't just using better technology – they're using technology better.

Structural issues are holding businesses back

Order Management departments say a **lack of automation** is the least of their worries. They already have the technology in place. But it isn't necessarily making the business less complex.

Most companies simply don't have the visibility into how their processes run across their ecosystem, even if they know something is going wrong. But if they don't know where the problem is and they don't know why it's happening, then they're not empowered to fix it.

Give a brain to your processes

You're the business expert, but you can't possibly know how every single process works, every single time. Afterall, Batman had Alfred. Sherlock had Watson. Cher had Sonny. And Skywalker had R2D2. Now you have Celonis.

Get a 360 view of your business execution

See how your business *really* works. Gain a precise, 360-degree-view of the processes within your enterprise. And find out if they are money-makers or money-takers.

Why should this be important?

Apromore

A pure process mining focus means we combine the most advanced process mining capabilities with intuitive design to unlock business value.[a]

[a] apromore.com

Tools

Academic License

- Disco[1]
- Apromore[2]
- Celonis[3]

Free – Research Oriented

- Nirdizati[4]
- ProM[5]

---

[1] https://fluxicon.com/disco/
[2] https://apromore.com/join-academic-alliance/
[3] https://www.celonis.com/
[4] http://research.nirdizati.org
[5] https://www.promtools.org/doku.php

DEMO

Dataset: Road Traffic Fine Management[1]

What we will see in the demo

- Load from csv
- Show process model
  - Frequences
  - Durations
    - Bottlenecks
    - Replay
  - The *Payment* Activity
    - Strange numbers on Payment occurrence
    - How to understand it: increase paths
    - Isolate and understand variants
  - Filters

[1] https://data.4tu.nl/articles/dataset/Road_Traffic_Fine_Management_Process/12683249

Conclusions

What was this lessons for?

To...

- come to know about Process Mining
- what it is used for
- what challenges it has to deal with
- which tools you can have a look at