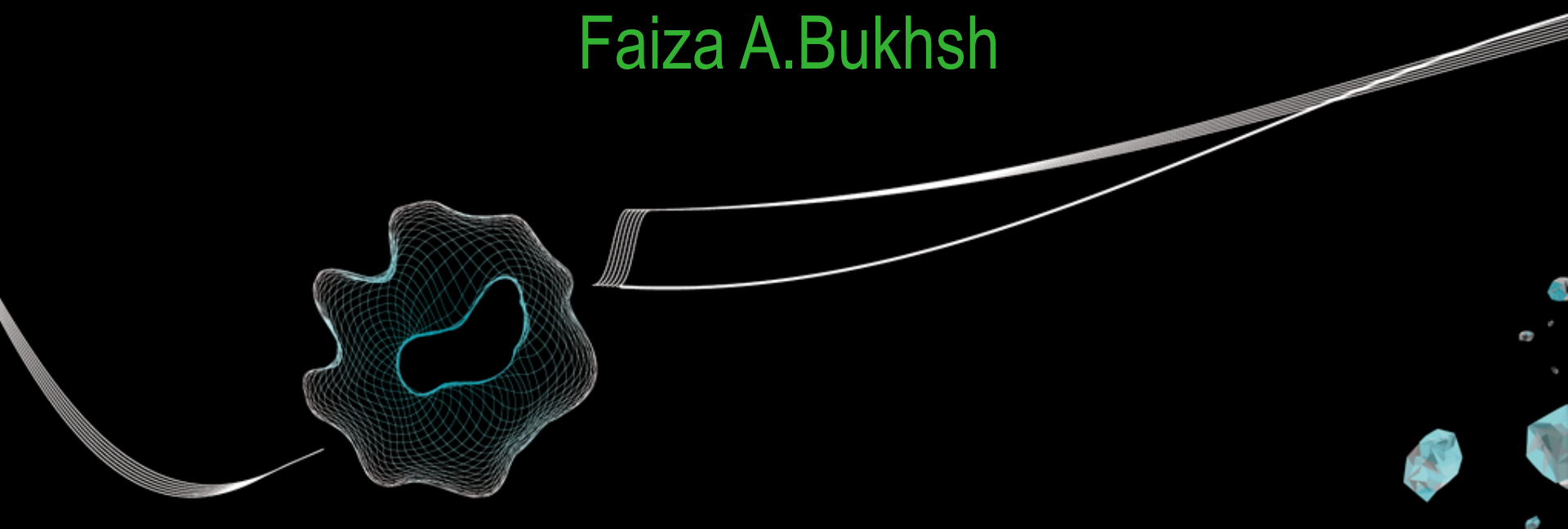
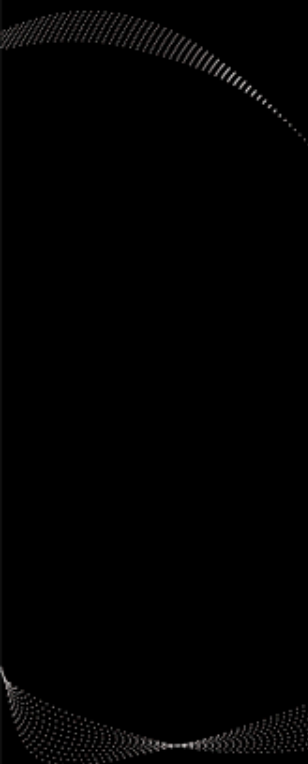


# Process Mining

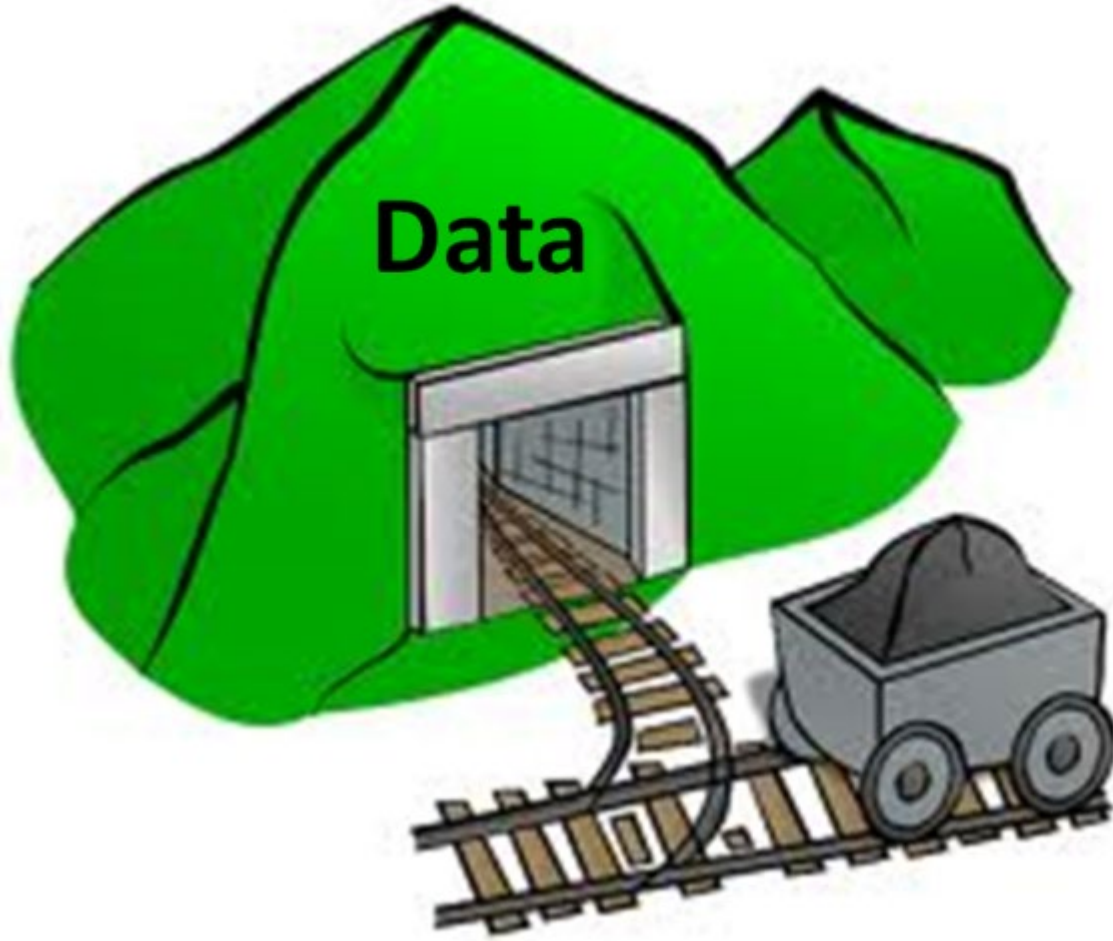
Faiza A.Bukhsh



# Mining



# Data mining



**Useful  
knowledge  
(patterns)**

# Business process



# Business process



# Why we need business process model

- *Insight:*
- *Discussion*
- *Documentation*
- *Verification e.g., potential deadlocks*
- *Performance analysis*
- *Prediction --play out*

# Limitations in business process model

- *Does not provide a complete picture*
- *Not aligned with reality*
- *Provide only an idealized view on the processes at hand: “paper tigers”*

*Given*

*(a) the interest in process models,*

*(b) the abundance of event data*

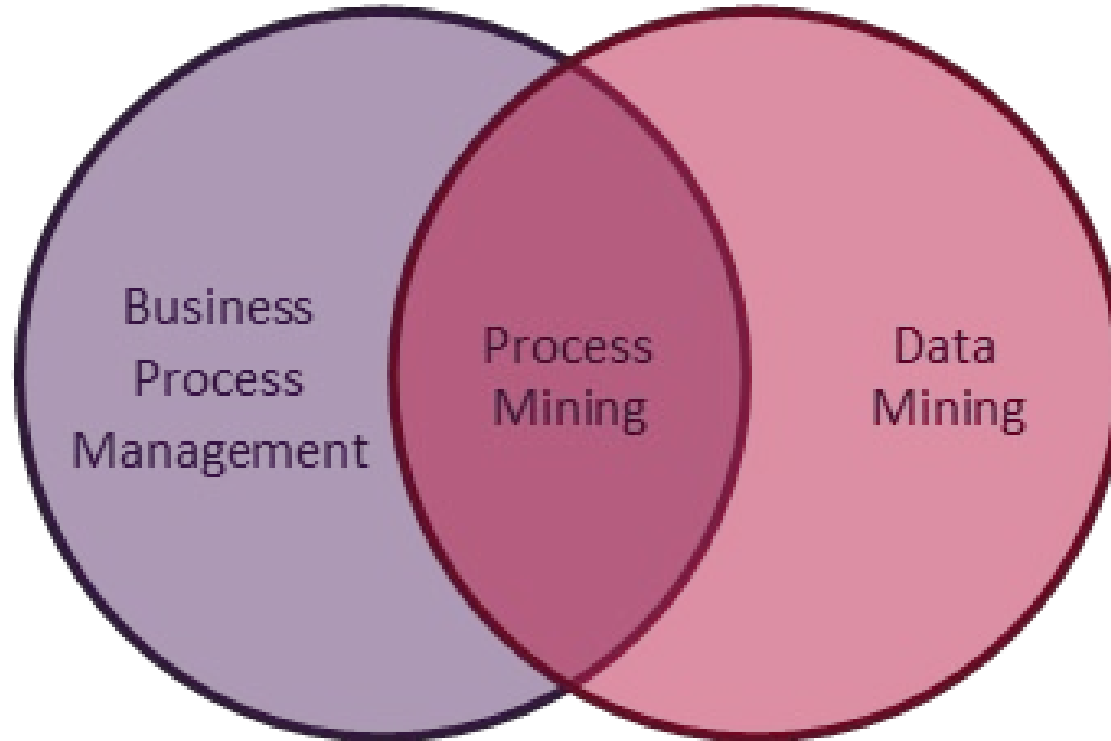
*(c) the limited quality of hand-made models,*

*Necessary to find **relation** between event data to process models*

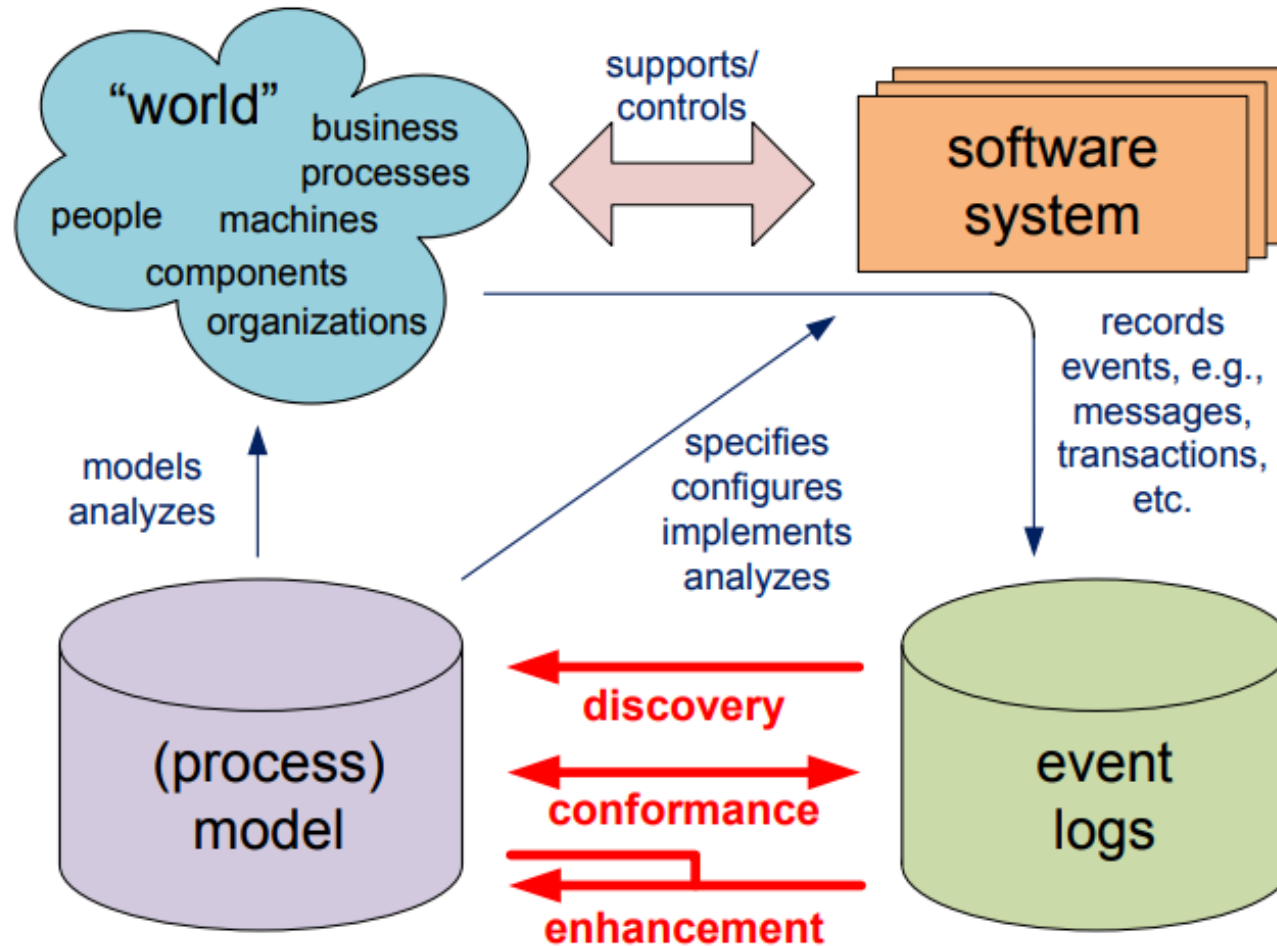
# Processes in everyday life



# Process mining



# Types of process mining



# Orthogonal perspectives

- The *control-flow perspective*:
  - focuses on the control flow, i.e., the ordering of activities.
- The *organizational perspective*
  - focuses on information about resources hidden in the log, i.e., which actors (e.g., people, systems, roles, and departments) are involved and how are they related.
- The *case perspective*
  - focuses on properties of cases, e.g., cases can also be characterized by the values of the corresponding data elements.
- The *time perspective*
  - Focuses on timing and frequency of events.

# Starting point: Event log

case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
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	...

*XES, MXML, SA-MXML, CSV, etc.*

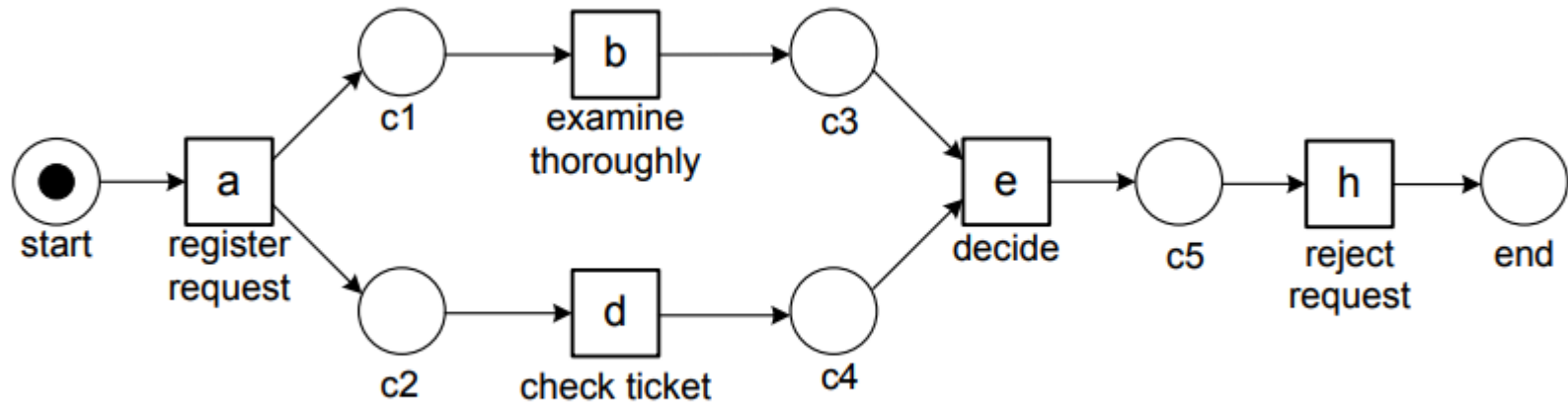
# Simplified event log

case id	event id	properties		
		timestamp	activity	resource
1	35654423	30-12-2010:11.02	register request	Pete
	35654424	31-12-2010:10.06	examine thoroughly	Sue
	35654425	05-01-2011:15.12	check ticket	Mike
	35654426	06-01-2011:11.18	decide	Sara
	35654427	07-01-2011:14.24	reject request	Pete
2	35654483	30-12-2010:11.32	register request	Mike
	35654485	30-12-2010:12.12	check ticket	Mike
	35654487	30-12-2010:14.16	examine casually	Pete
	35654488	05-01-2011:11.22	decide	Sara
	35654489	08-01-2011:12.05	pay compensation	Ellen
3	35654521	30-12-2010:14.32	register request	Pete
	35654522	30-12-2010:15.06	examine casually	Mike
	35654524	30-12-2010:16.34	check ticket	Ellen
	35654525	06-01-2011:09.18	decide	Sara
	35654526	06-01-2011:12.18	reinitiate request	Sara
	35654527	06-01-2011:13.06	examine thoroughly	Sean
	35654530	08-01-2011:11.43	check ticket	Pete
	35654531	09-01-2011:09.55	decide	Sara
35654533	15-01-2011:10.45	pay compensation	Ellen	
4	35654641	06-01-2011:15.02	register request	Pete
	35654643	07-01-2011:12.06	check ticket	Mike
	35654644	08-01-2011:14.43	examine thoroughly	Sean
	35654645	09-01-2011:12.02	decide	Sara
	35654647	12-01-2011:15.44	reject request	Ellen
5	35654711	06-01-2011:09.02	register request	Ellen
	35654712	07-01-2011:10.16	examine casually	Mike
	35654714	08-01-2011:11.22	check ticket	Pete
	35654715	10-01-2011:13.28	decide	Sara
	35654716	11-01-2011:16.18	reinitiate request	Sara
	35654718	14-01-2011:14.33	check ticket	Ellen
	35654719	16-01-2011:15.50	examine casually	Mike
	35654720	19-01-2011:11.18	decide	Sara
	35654721	20-01-2011:12.48	reinitiate request	Sara
	35654722	21-01-2011:09.06	examine casually	Sue
	35654724	21-01-2011:11.34	check ticket	Pete
	35654725	23-01-2011:13.12	decide	Sara
35654726	24-01-2011:14.56	reject request	Mike	
6	35654871	06-01-2011:15.02	register request	Mike
	35654873	06-01-2011:16.06	examine casually	Ellen
	35654874	07-01-2011:16.22	check ticket	Mike
	35654875	07-01-2011:16.52	decide	Sara

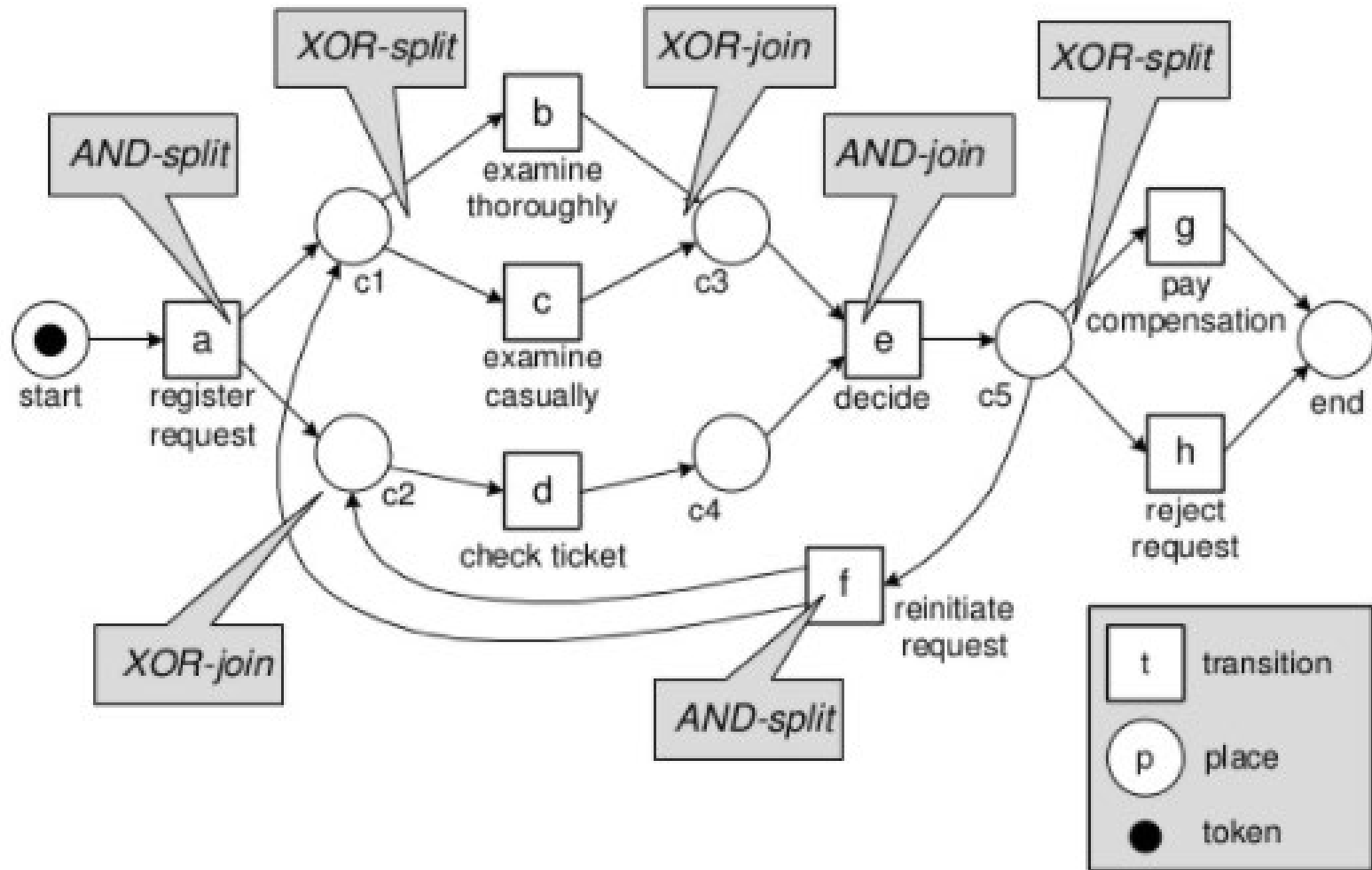
case id	trace
1	$\langle a, b, d, e, h \rangle$
2	$\langle a, d, c, e, g \rangle$
3	$\langle a, c, d, e, f, b, d, e, g \rangle$
4	$\langle a, d, b, e, h \rangle$
5	$\langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle$
6	$\langle a, c, d, e, g \rangle$
...	...

**a = register request,**  
**b = examine thoroughly,**  
**c = examine casually,**  
**d = check ticket,**  
**e = decide,**  
**f = reinitiate request,**  
**g = pay compensation,**  
**and h = reject request**

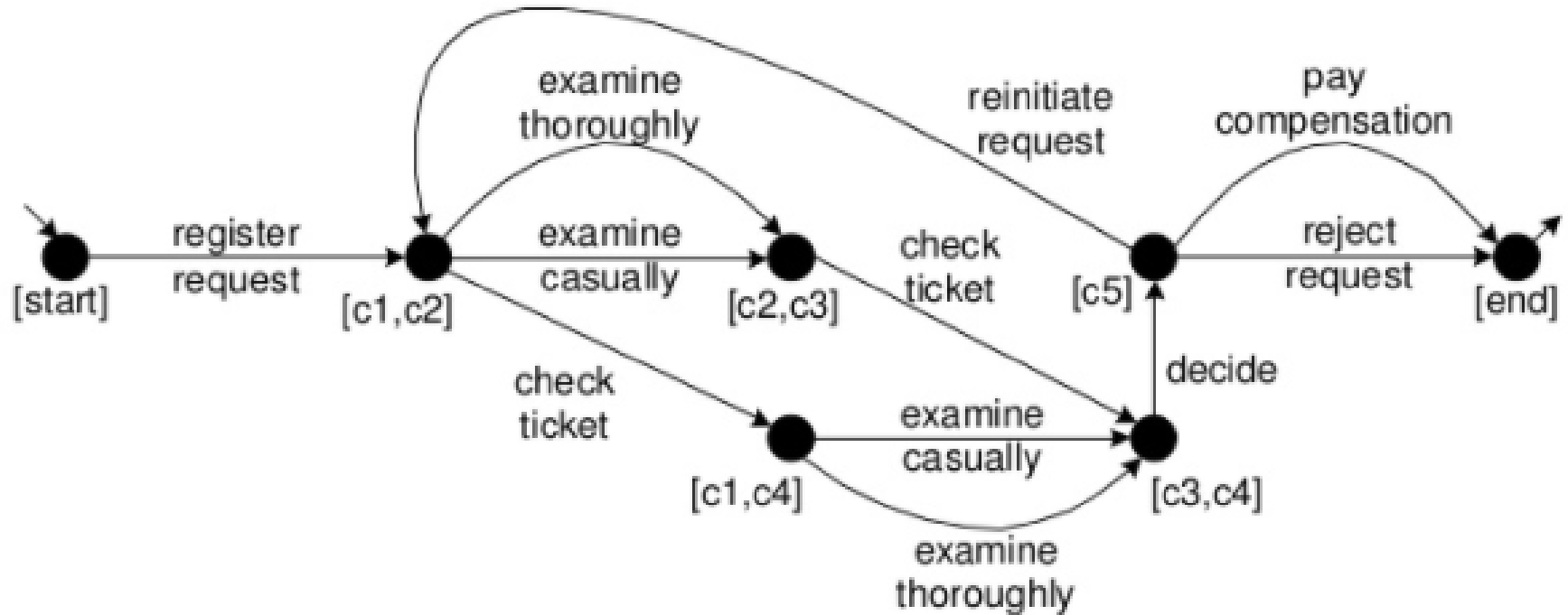
# Process discovery: Petrinet



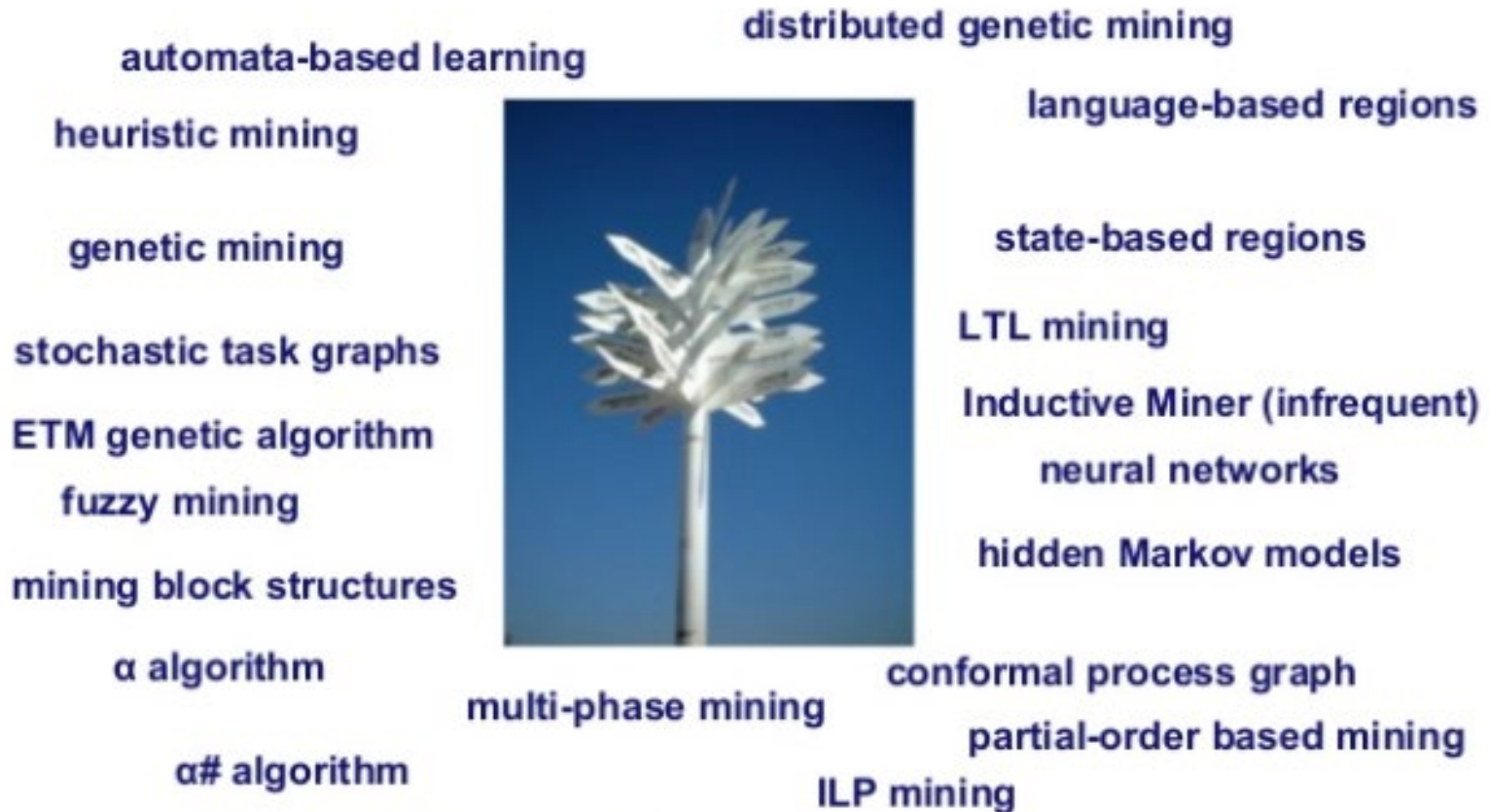
# Process discovery: Petrinet



# Process discovery: Transition system



# Process discovery algorithms



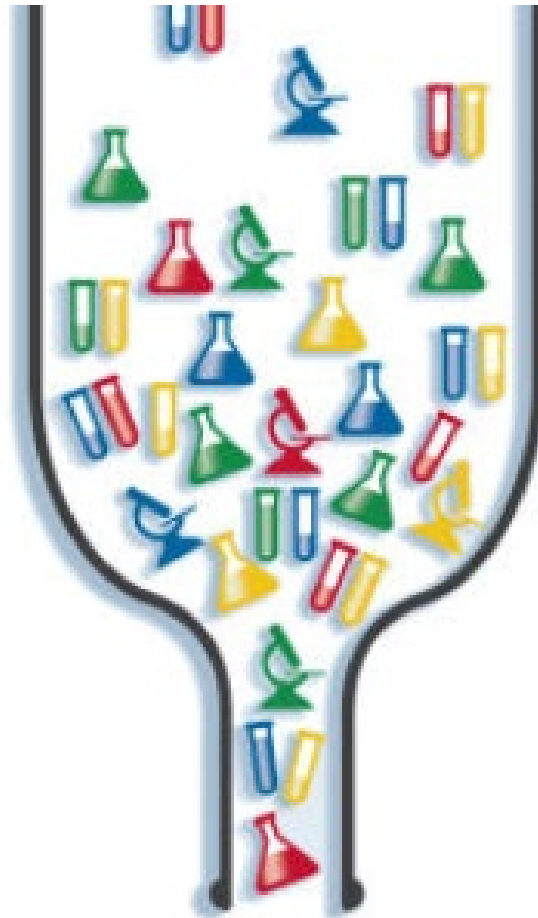
# Process discovery

010011001101010101010  
010011010101010

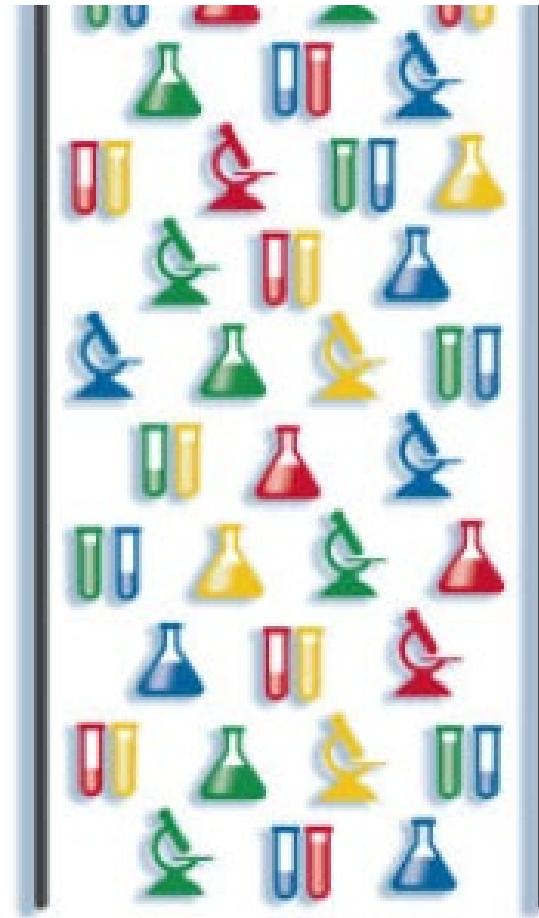
A photograph of a brick path with two cyclists, overlaid with binary code. The path is made of reddish-brown bricks and curves to the right. Two people are riding bicycles on the path: one in a red shirt on the left and one in a blue shirt on the right. The background shows a grassy area, trees, a road with cars, and a white building. The text '010011001101010101010' is written in white, slanted font along the left side of the path, and '010011010101010' is written in white, slanted font along the right side of the path.

[www.olifantenpaadjes.nl](http://www.olifantenpaadjes.nl)

# Business Process Example: Bottleneck

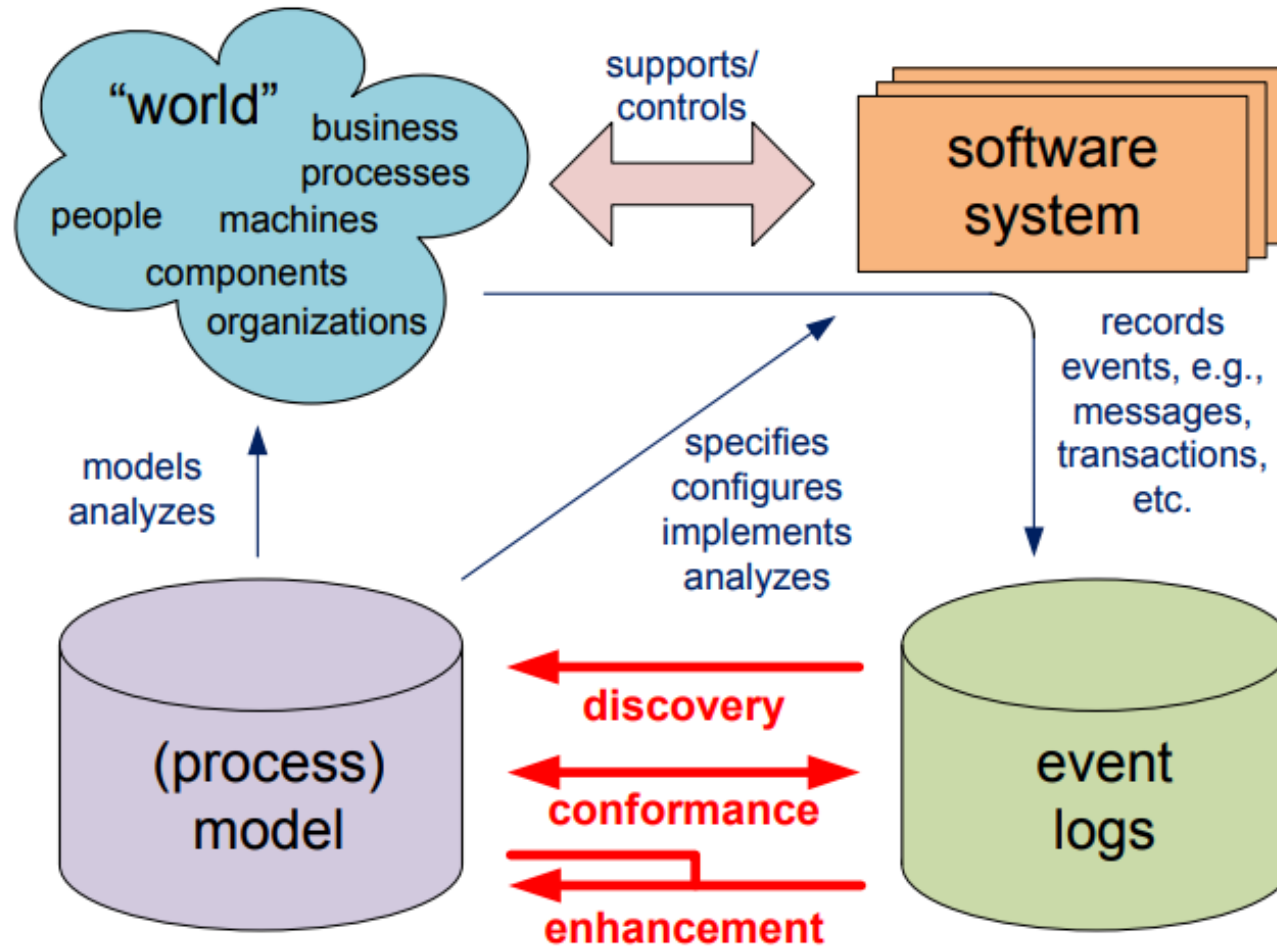


BEFORE

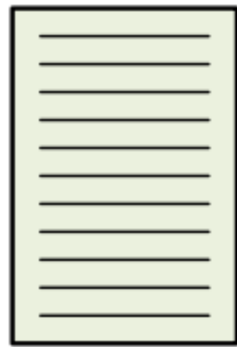


AFTER

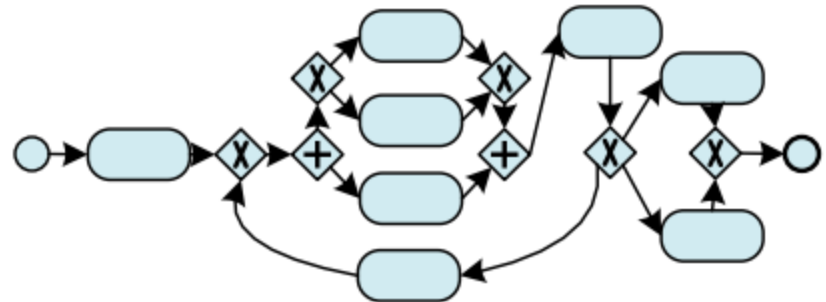
# Beyond Discovery



# Play-in

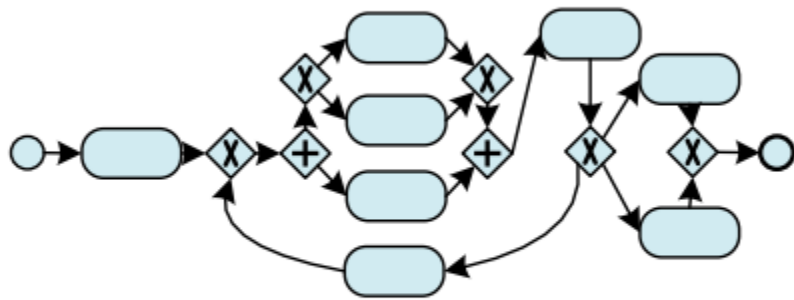


event log

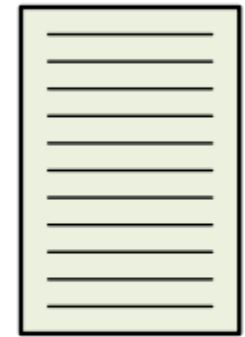


process model

# Play-out

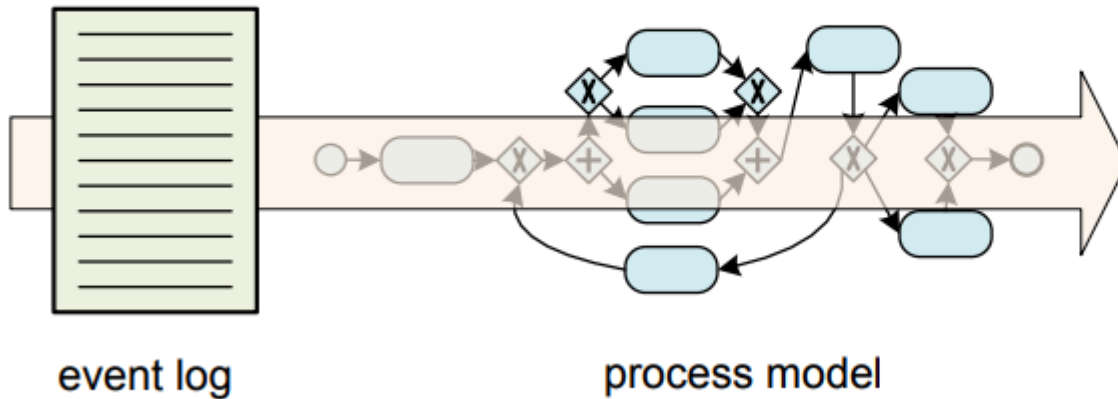


process model



event log

# Replay



- extended model showing times, frequencies, etc.
- diagnostics
- predictions
- recommendations

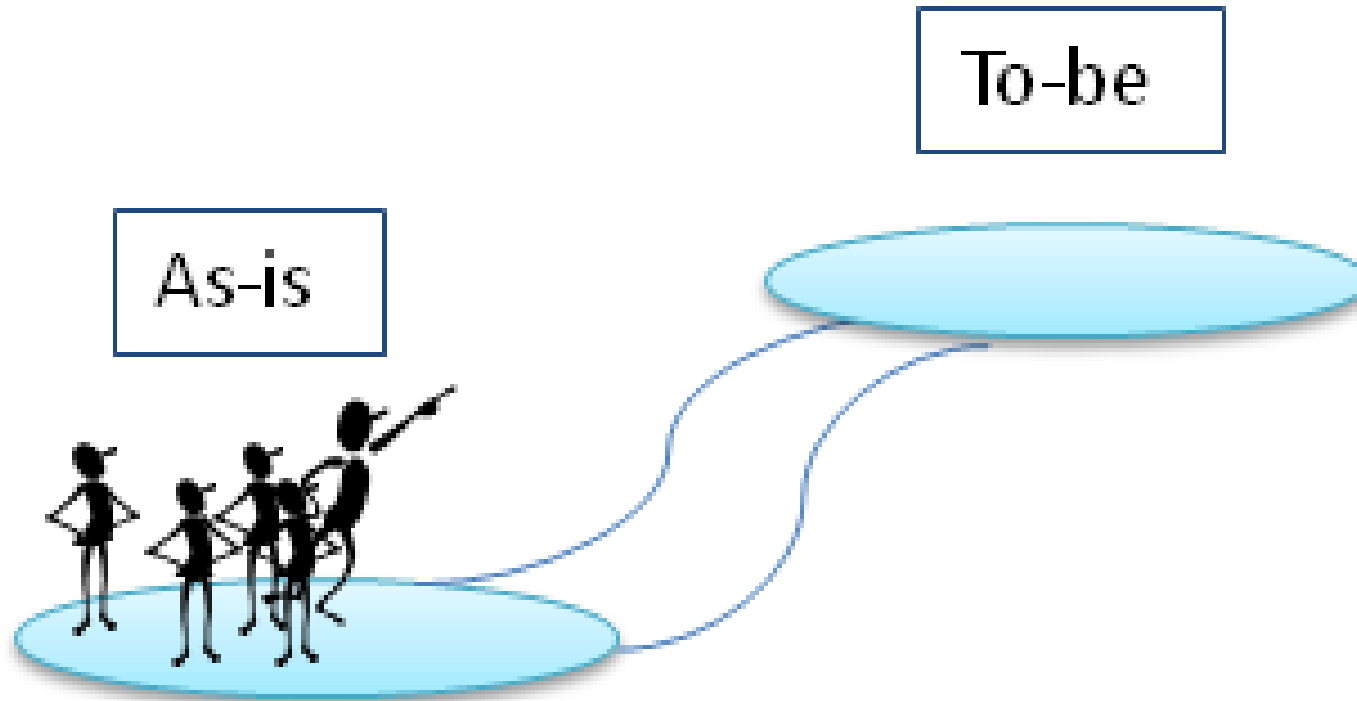
# Replay

- *Connecting models to real events is crucial!*
- *Possible uses:*
  - *Conformance checking*
  - *Repairing models*
  - *Extending the model with frequencies and temporal information*
  - *Constructing predictive models*
  - *Operational support (prediction, recommendation, etc.)*

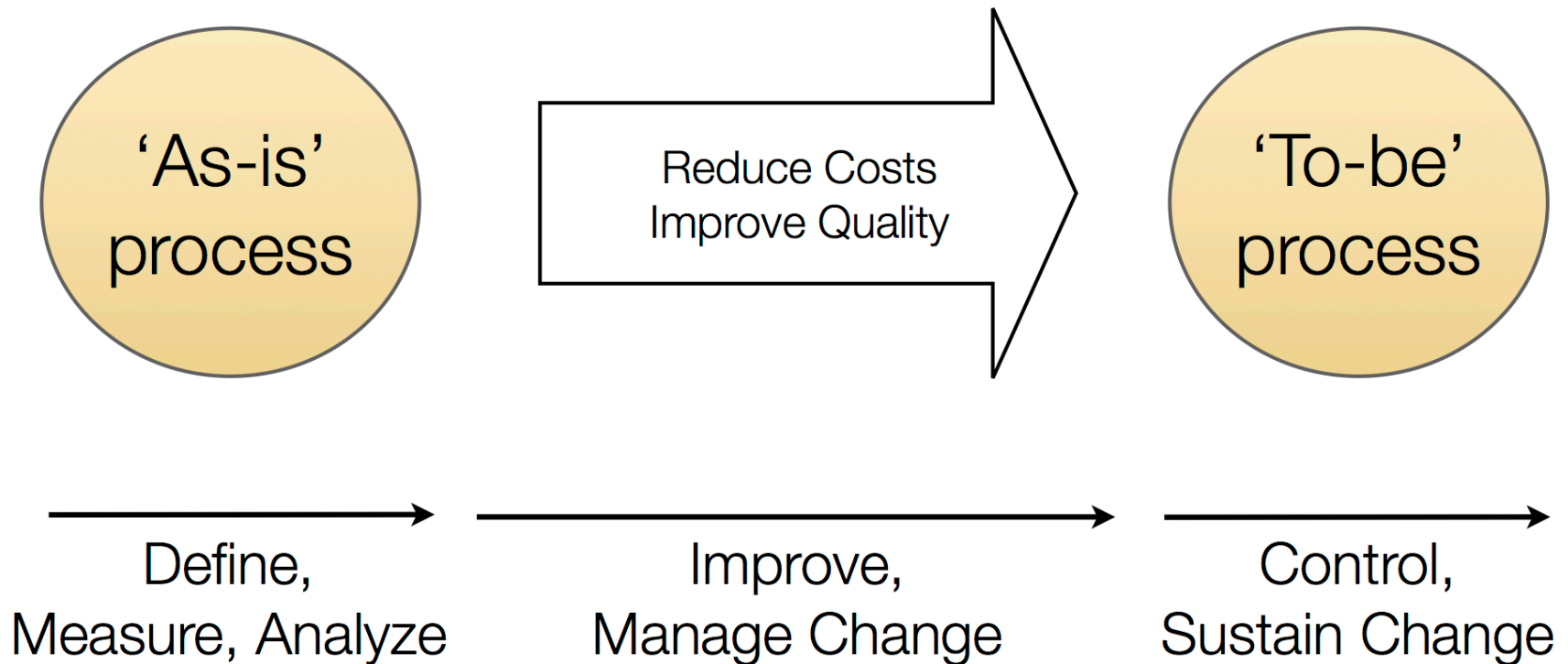
# Performance improvement versus compliance

- Organizations are also putting more emphasis on *corporate governance, risk, and compliance*.
- *Scandals* have fueled interest in more rigorous auditing practices.
- *New legislation* such as the Sarbanes-Oxley Act (SOX) of 2002 and the Basel II Accord of 2004 emerged as a result.
- Importance of verifying whether organizations operate “*within their boundaries*” is increasing.

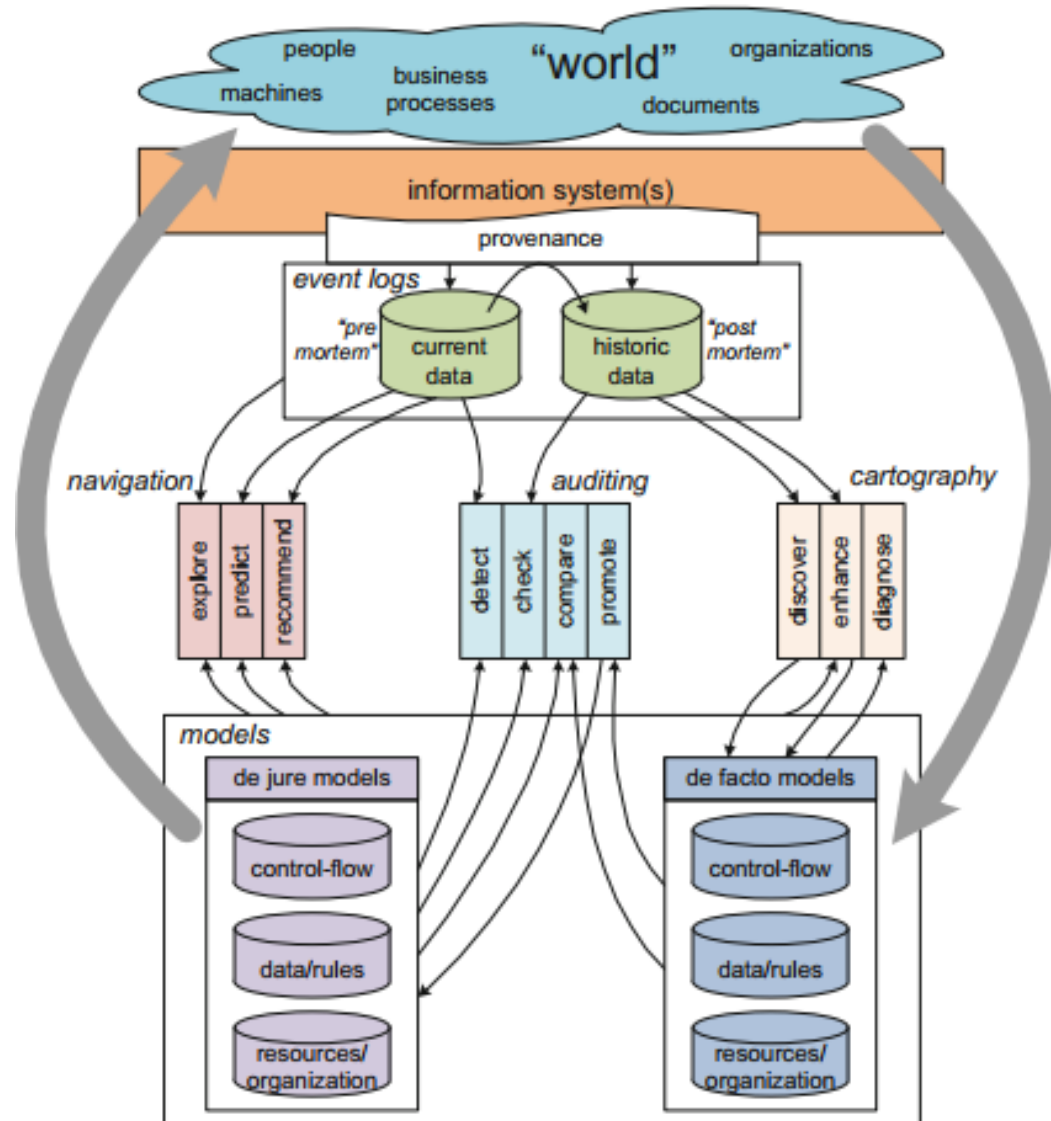
# Auditing 2.0: Basis of process mining



# Auditing 2.0: Basis of process mining



# Process mining spectrum



# Process mining tool



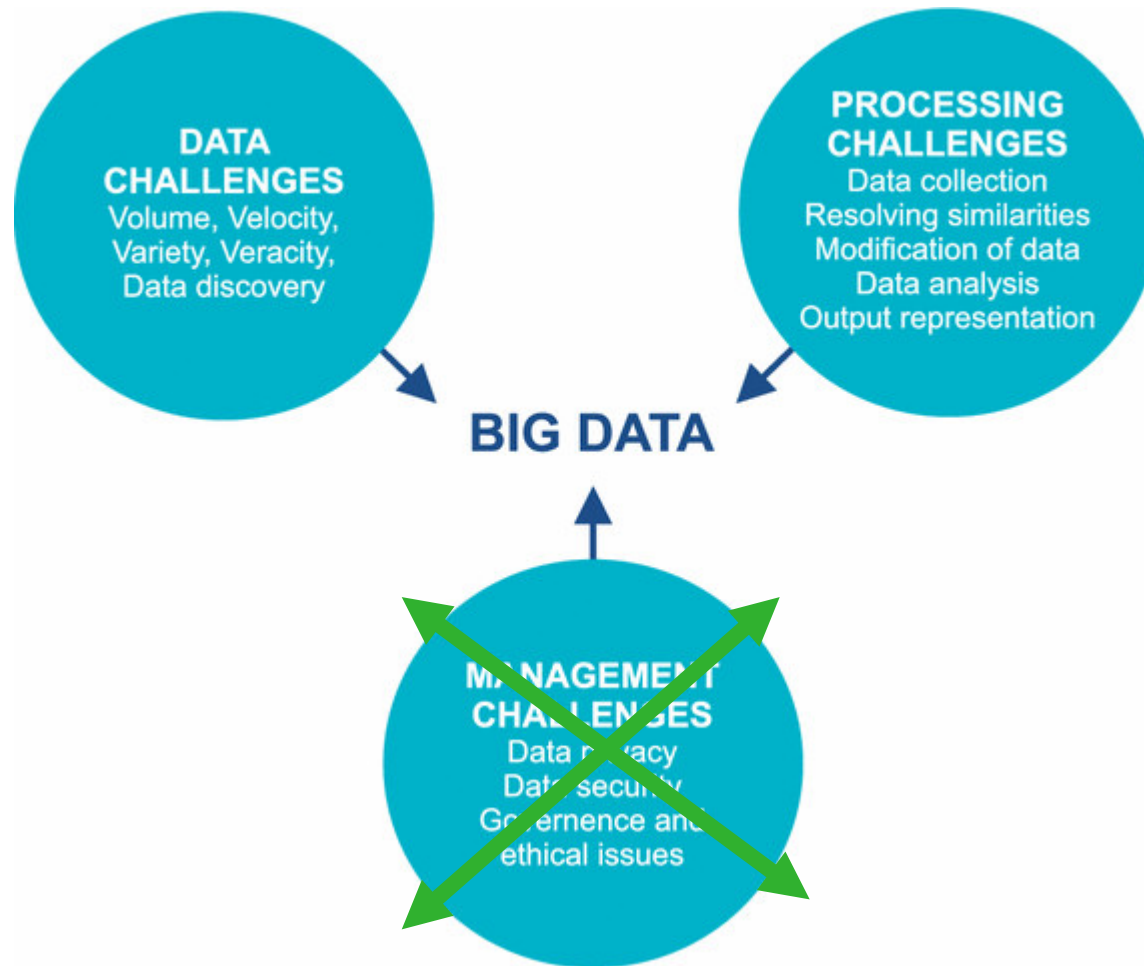
600+ plug-ins available covering the whole process mining spectrum



# PROM application >100 organization

- **Municipalities:** e.g., Alkmaar, Heusden, Harderwijk
- **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:** e.g., Philips Healthcare, ASML, Ricoh, Thales
- **Media companies:** e.g., Winkwaves

# Big Data Challenges



# Big Data Challenges and Process Mining



# Big Data Challenges and Process Mining

*Divide and conquer approach*

*Vertical and horizontal division*

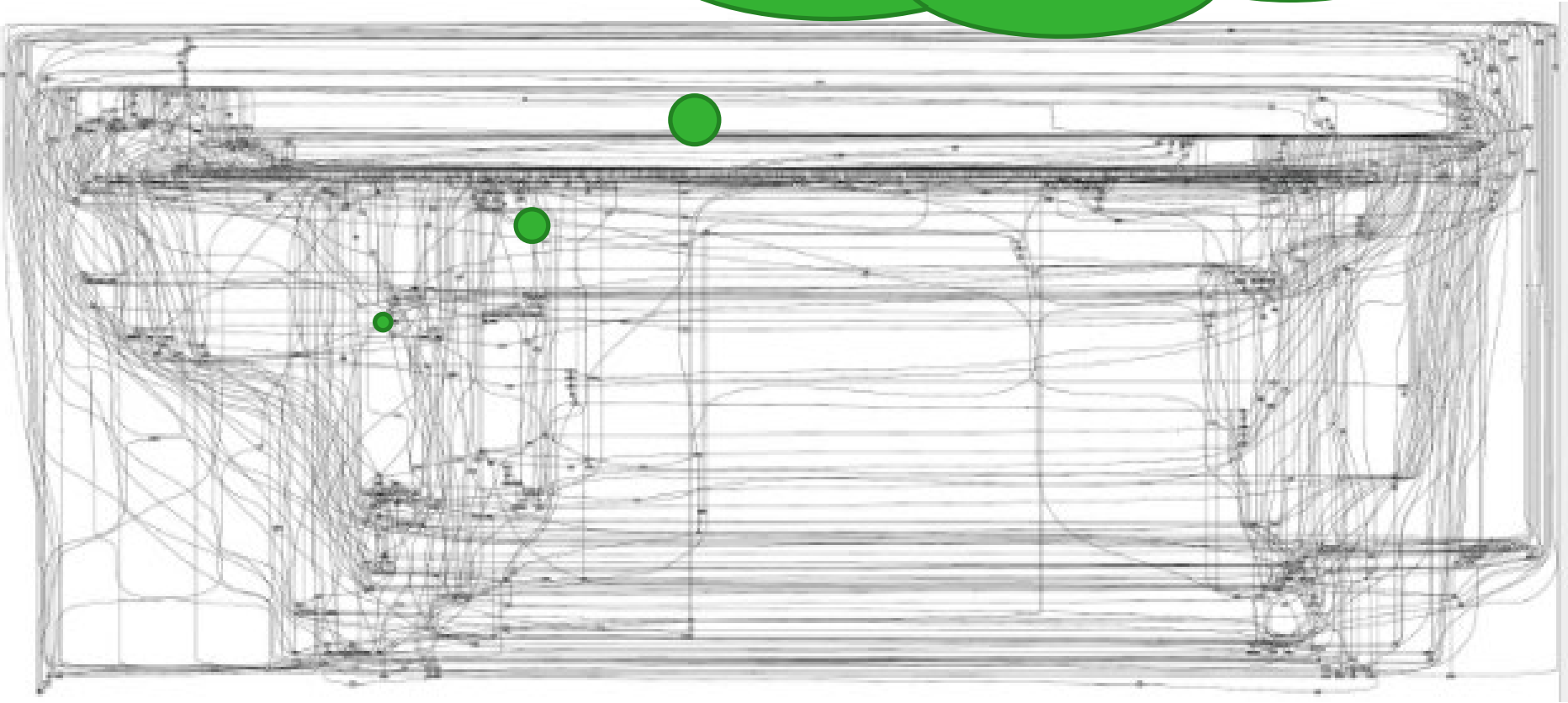
*Process discovery*

*Conformance checking*



# Spaghetti process

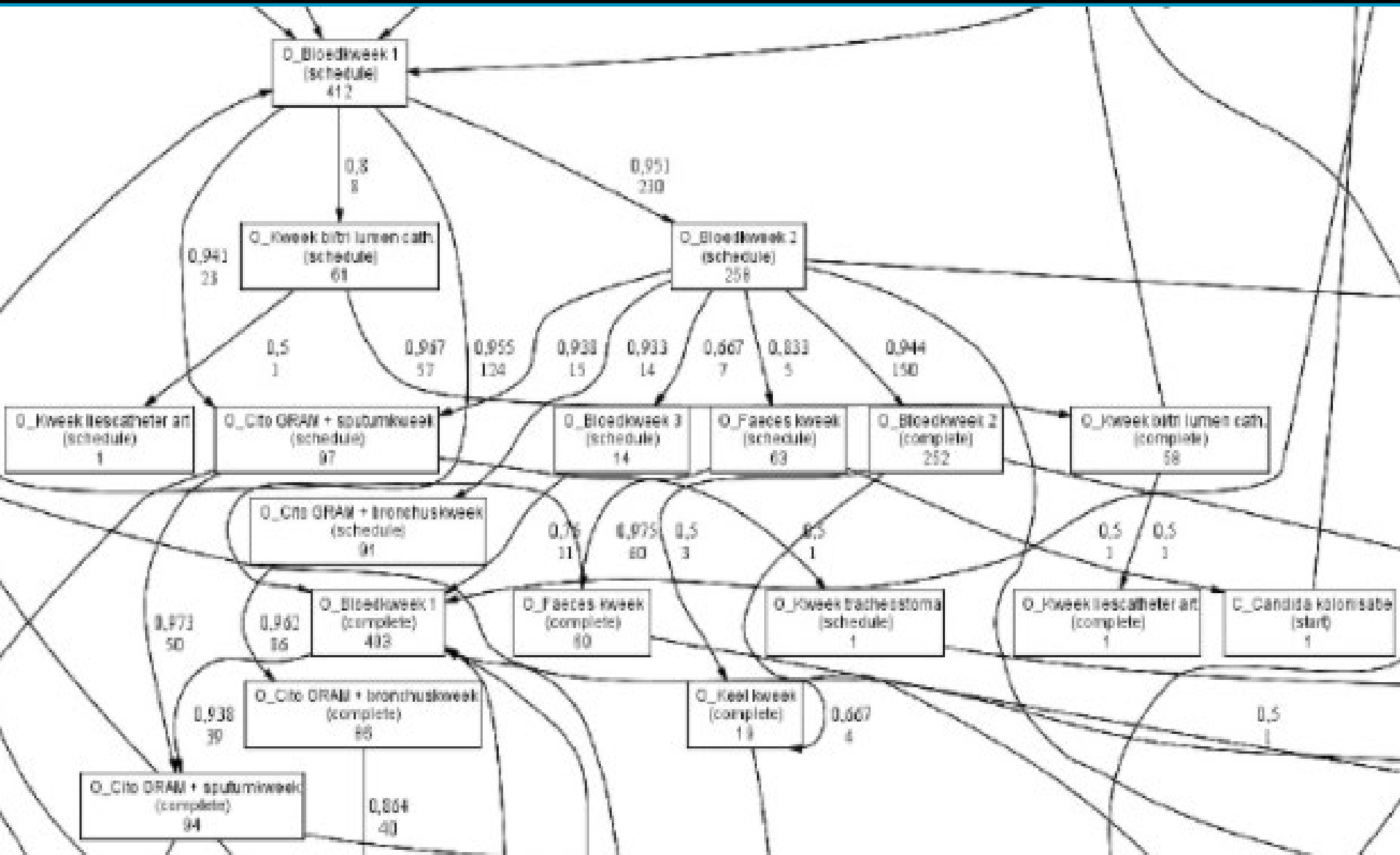
*2765 patients  
event log of 114,592 events  
619 different activities  
266 different individuals*



# Unfold a Spaghetti-Model

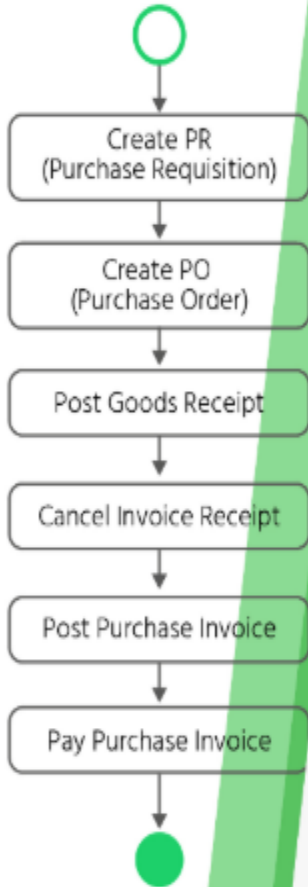


# Fragmentation : 18 activities out of 619

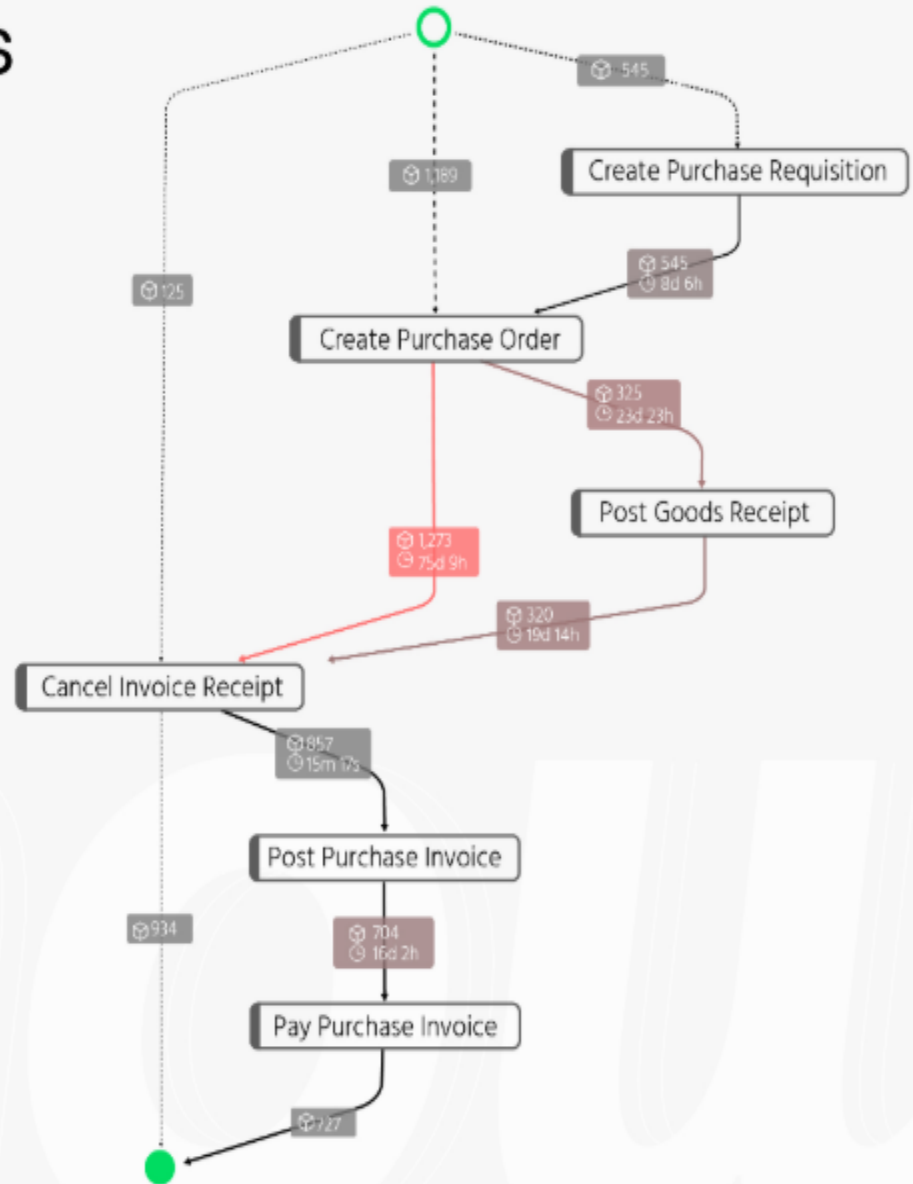




# Ideal Process



# Process Reality



# How to start?

## Collect Event Data

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

## Collect questions

- **Problem type:** cost, time, risk, compliance service etc
- **Process mining type:** discovery, conformance, enhancement?
- **Iterative process:** can be curiosity driven initially

# Conclusion

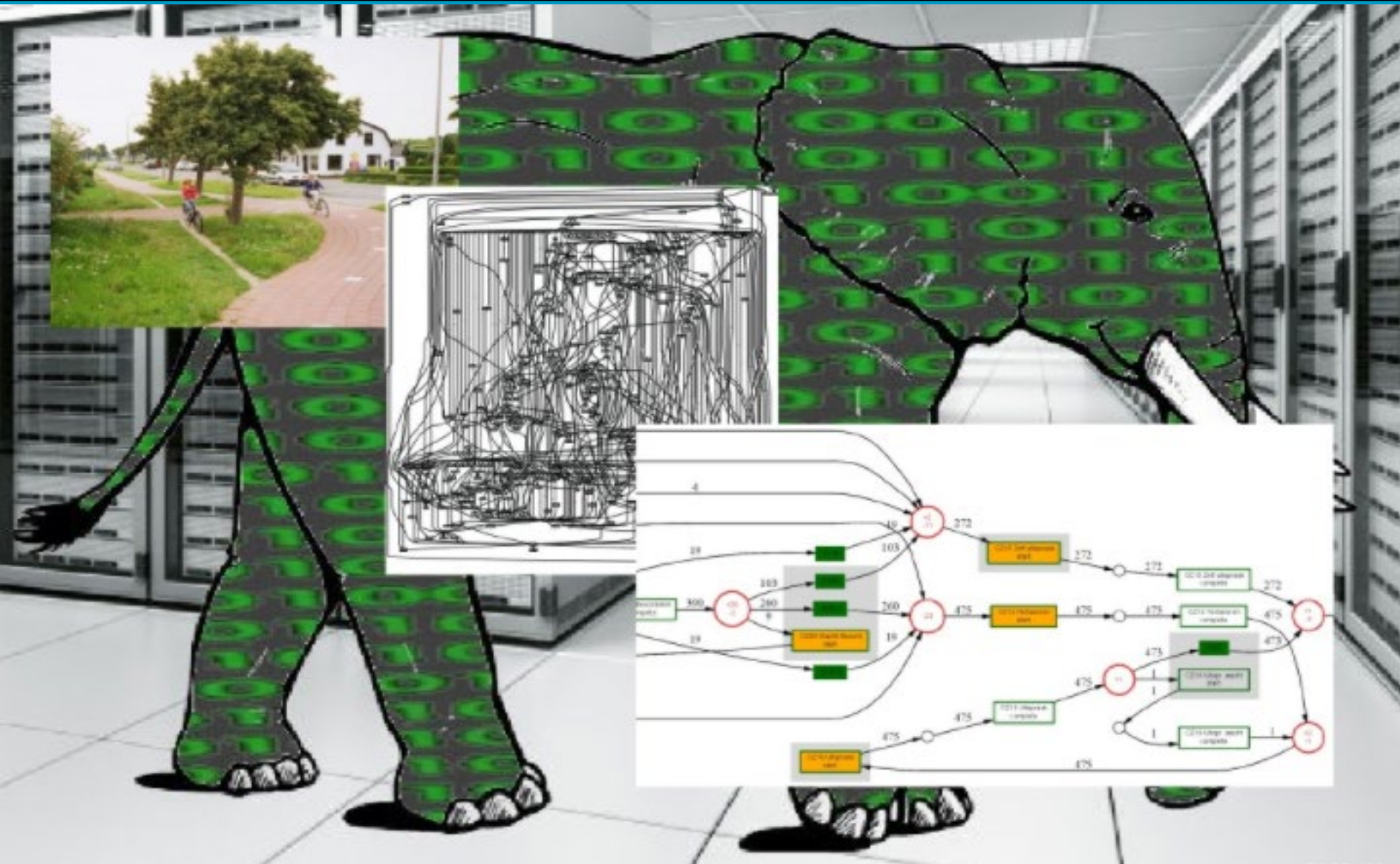
Recommend: How to reach your destination ASAP?

Detect: You are driving fast



Predict: When will you reach your destination?

# Conclusion



# What are you going to do in this course?

## Assignments

- *On paper*
- *PROM*
- *Concrete questions*
- *Discovery phase*
- *Dataset*

## Project

- *Dataset*
- *Problem is not concrete*
- *About conformance and enhancement*
- *Advisory Report*

*Curiosity driven*

UNIVERSITY OF TWENTE.

Question?

