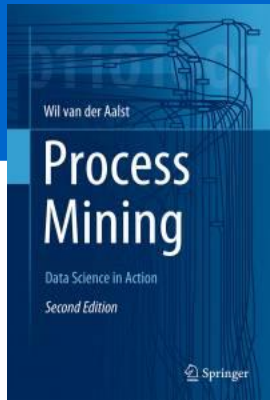


*Process Mining: Data Science in Action*

# Event Logs and Process Models

prof.dr.ir. Wil van der Aalst  
[www.processmining.org](http://www.processmining.org)



**TU/e**

Technische Universiteit  
**Eindhoven**  
University of Technology

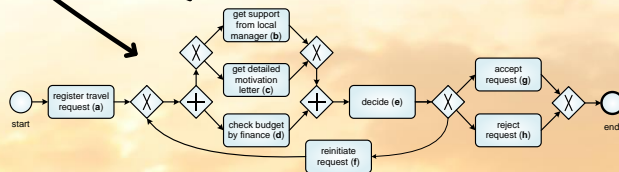
**Where innovation starts**

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary

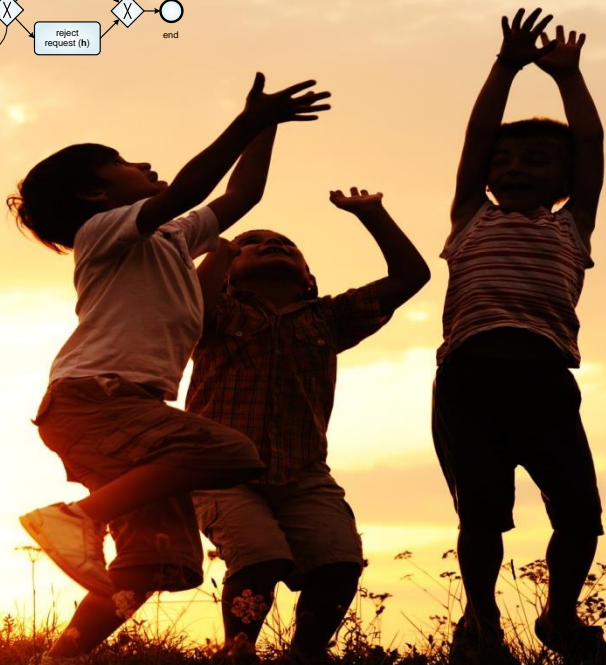
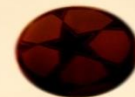
Play-In

Replay

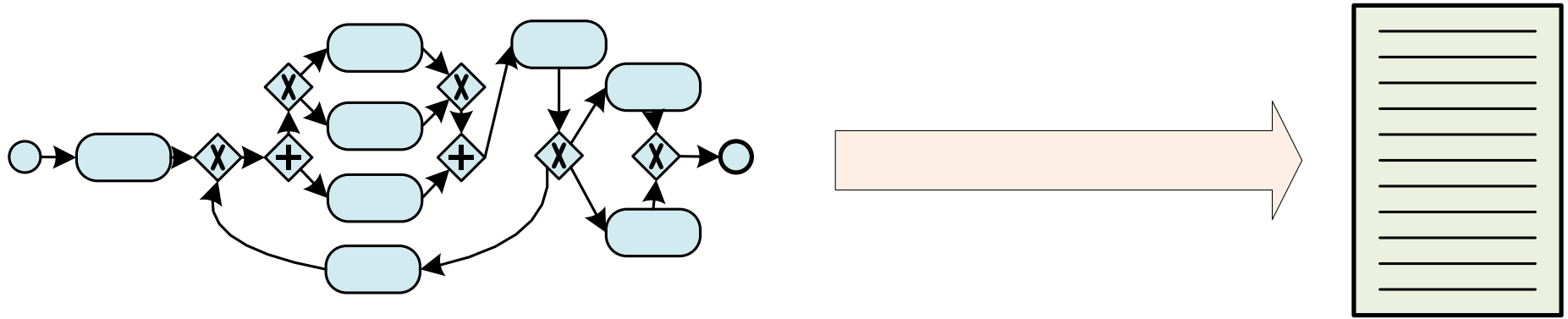
Play-Out



relating process  
models and event logs

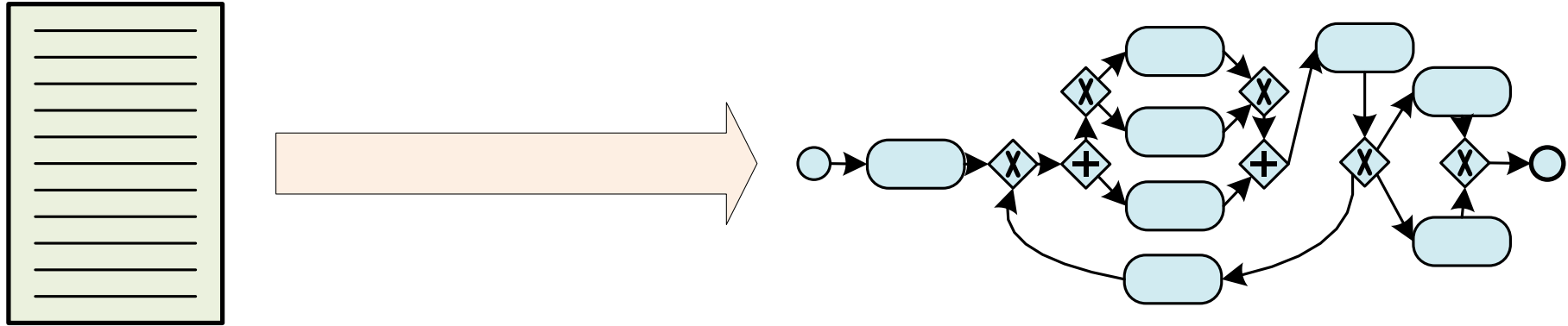


# Play-Out



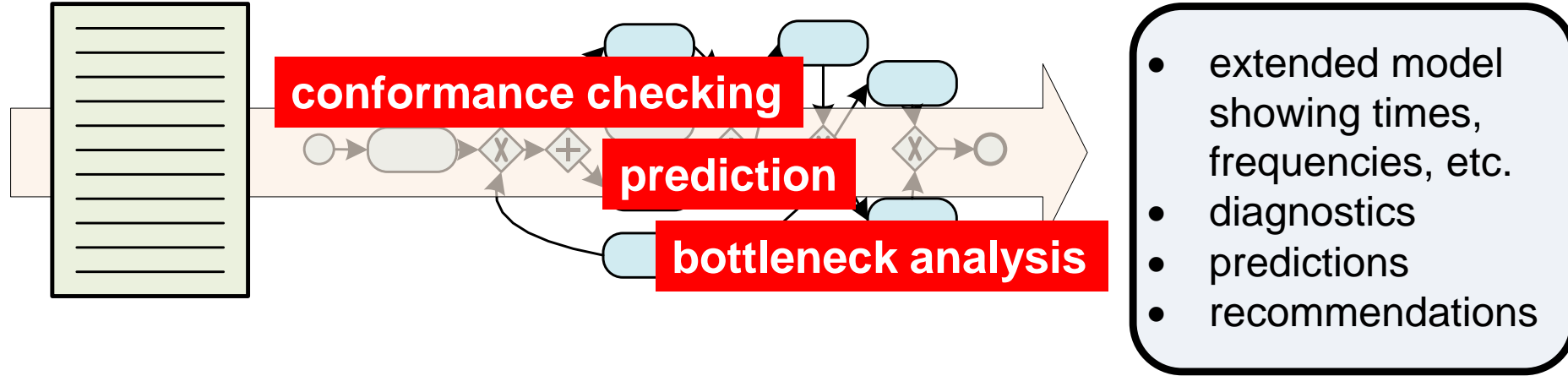
- **Simulation**
- **Workflow automation**
- **Management games**
- **Model checking**
- ...

# Play-In



**Process discovery: learning de facto process models from observed behavior.**

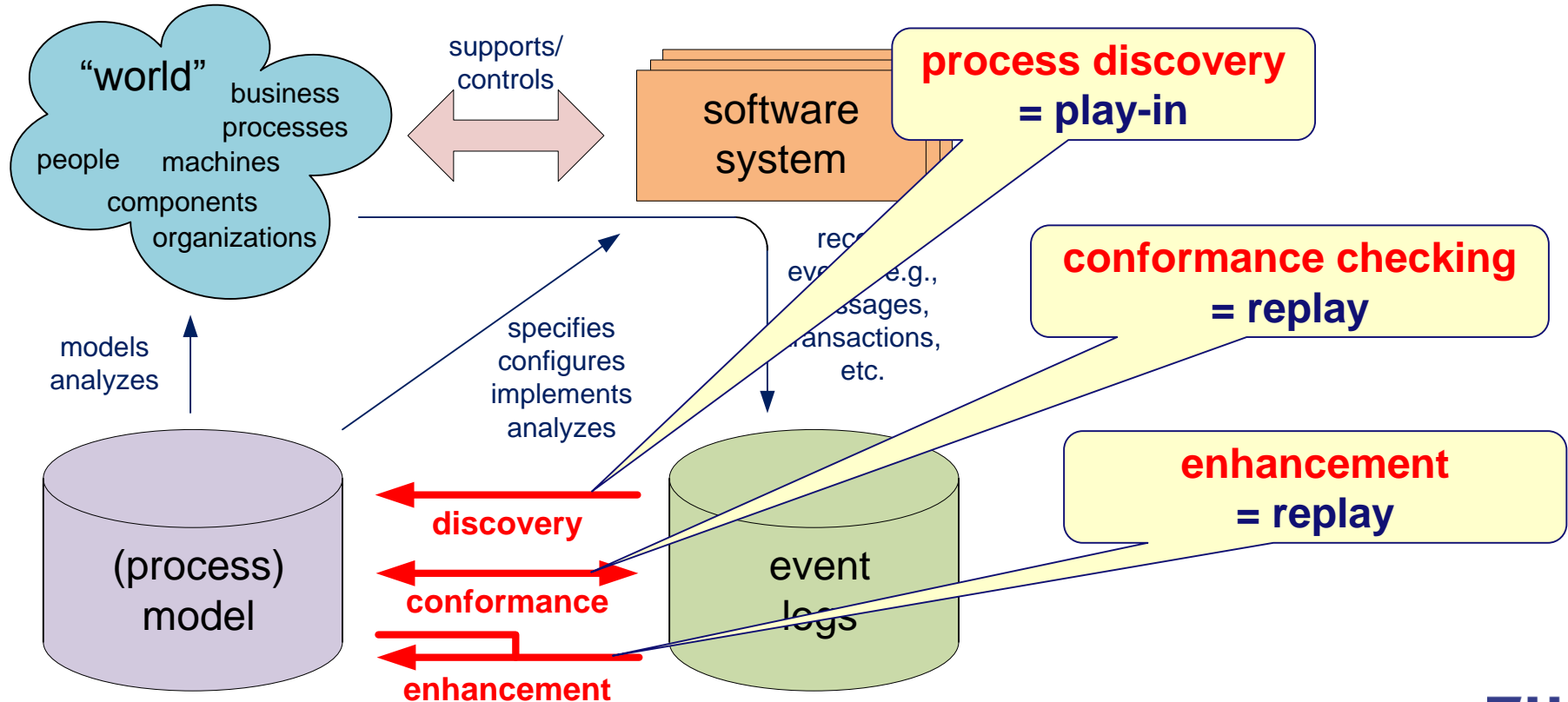
# Replay



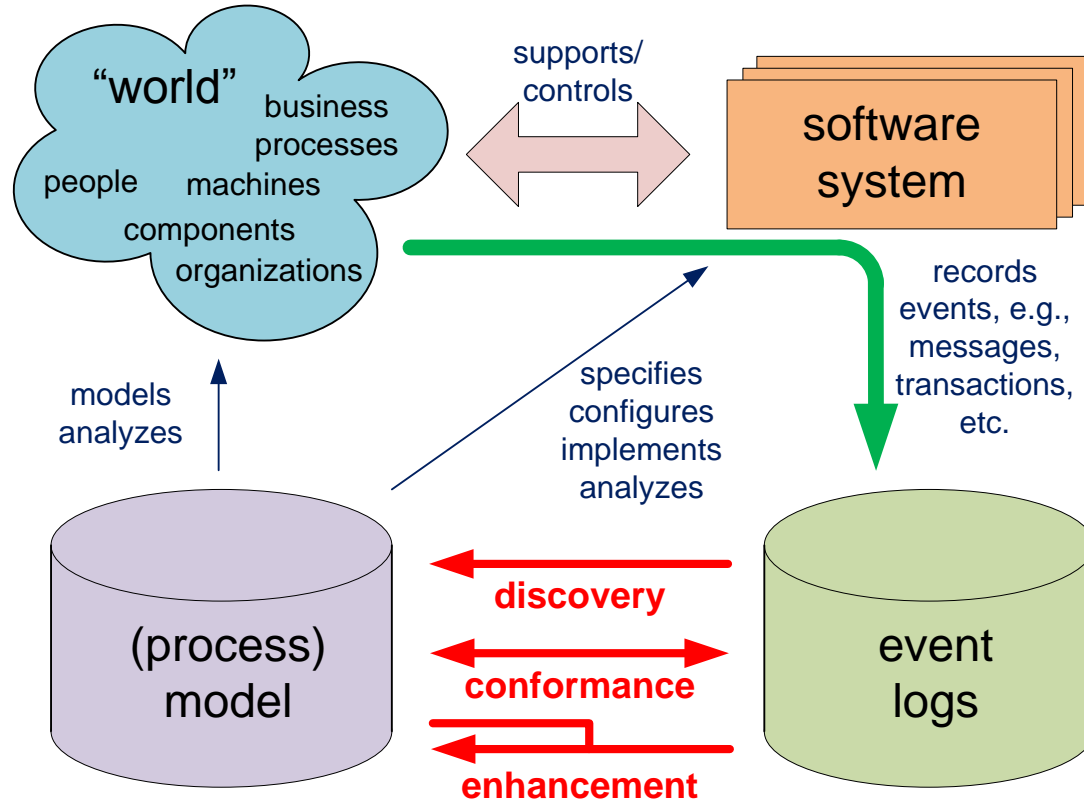
## Aligning modeled/discovered and observed behavior:

- The most important form of process mining!
- Confrontation between model and reality.

# The three main types of process mining: discovery, conformance, and enhancement



# 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 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**

**not always so clear ...**

# Question: Take your mail box

- An e-mail has:
  - a sender ("From"),
  - a set of receivers ("To"),
  - a subject,
  - a timestamp ("Date"),
  - a body,
  - etc.
- Assume an e-mail represents an event.
- What is a possible mapping onto an event log (which field is the case id, which field is the activity name, etc.)?

# Answer: Several possible mappings!

- Mapping:

- a sender ("From"),
- a set of receivers ("To"),
- a subject,
- a timestamp ("Date"),
- a body

resource

activity  
name

other data

case id

timestamp

other data

- Problems:

- Unclear what the cases are (senders, subjects, etc.).
- Unclear what the activities are.
- Context and questions needed.

# Question

student id	course	exam date	gender	nationality	mark
884362	Logic	15-4-2014	male	Dutch	6
884978	Logic	15-4-2014	female	Chinese	8
884362	Statistics	17-4-2014	male	Dutch	9
885087	Statistics	17-4-2014	female	German	9
882876	Statistics	17-4-2014	female	Dutch	5
..	..	..	..	..	..

- Assume each row (an exam attempt) represents an event.
- What is a possible mapping onto an event log (which field is the case id, etc.)?



# Answer

resource			timestamp		other data	
student id	course	exam date	gender	nationality	mark	
884362	Logic	15-4-2014	male	Dutch	6	
884978	Logic	15-4-2014	female	Chinese	8	
884362	Statistics	17-4-2014	male	Dutch	9	
885087	Statistics	17-4-2014	female	German	9	
882876			female	Dutch	5	
..	..		..	other data	other data	

- Again alternative mappings possible!  
For example, the course is the case, the student is the activity, and the nationality is the resource.
- Context and questions needed.

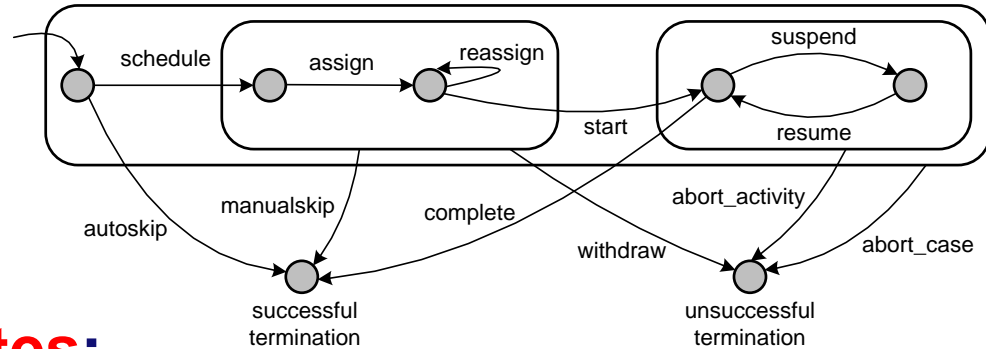
# Extensions

- **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.



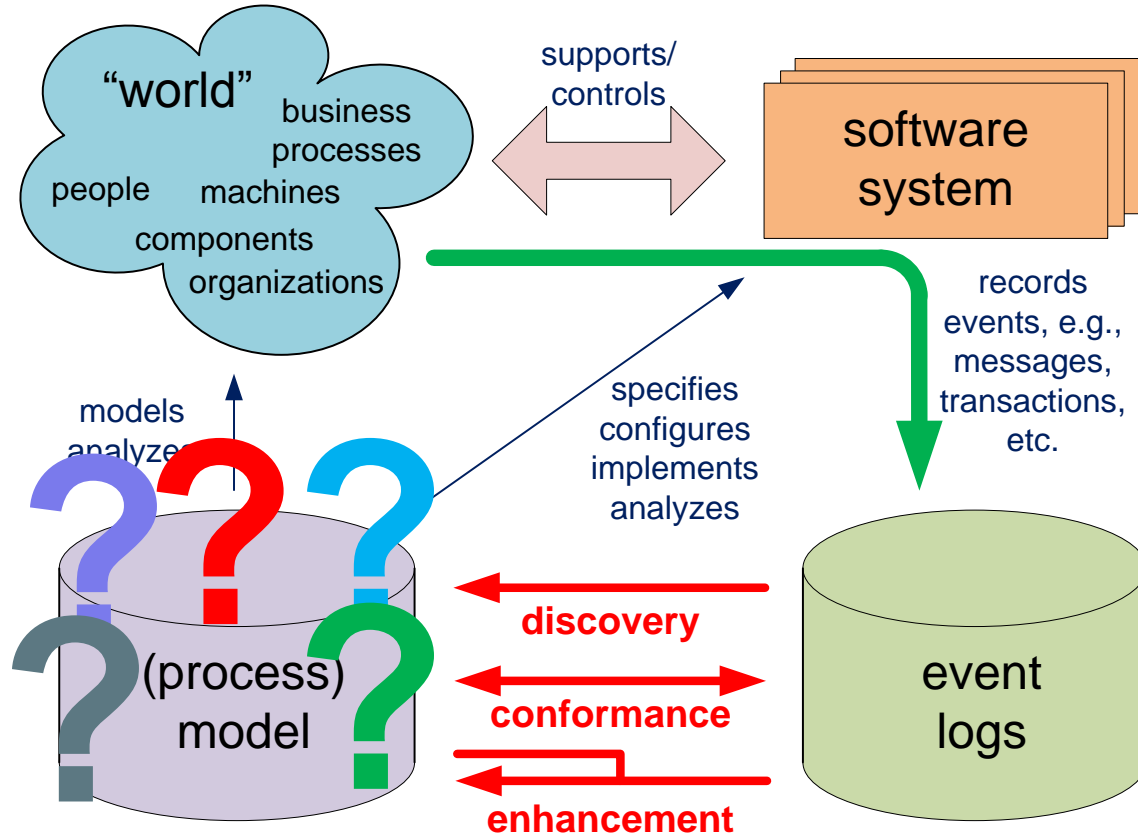
# XES (eXtensible Event Stream)

- 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](http://www.xes-standard.org)**.



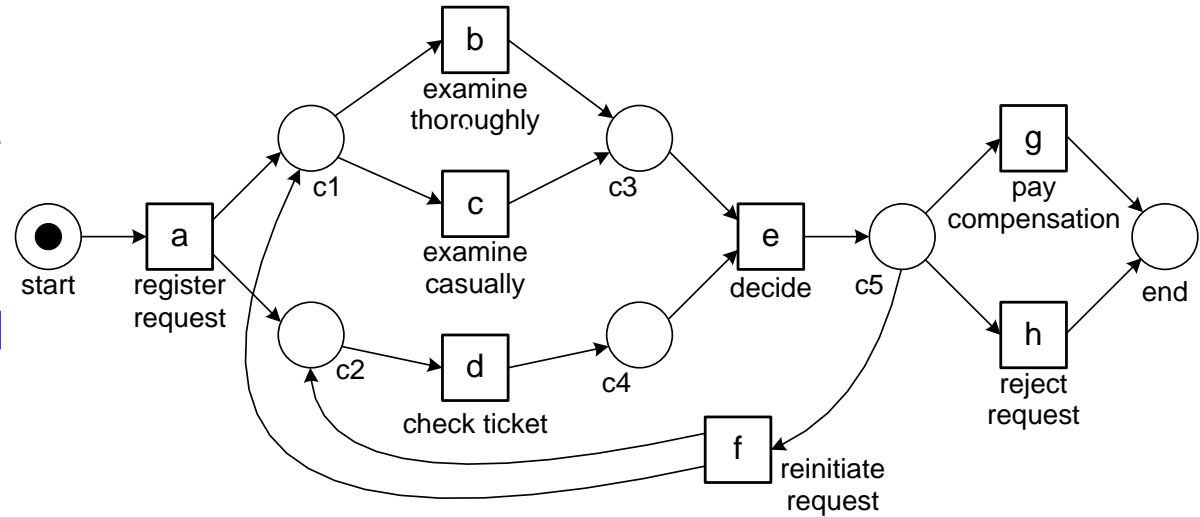
Extensible Event Stream

# Selecting the "right" representation...



# Just control-flow (represented as a Petri net)

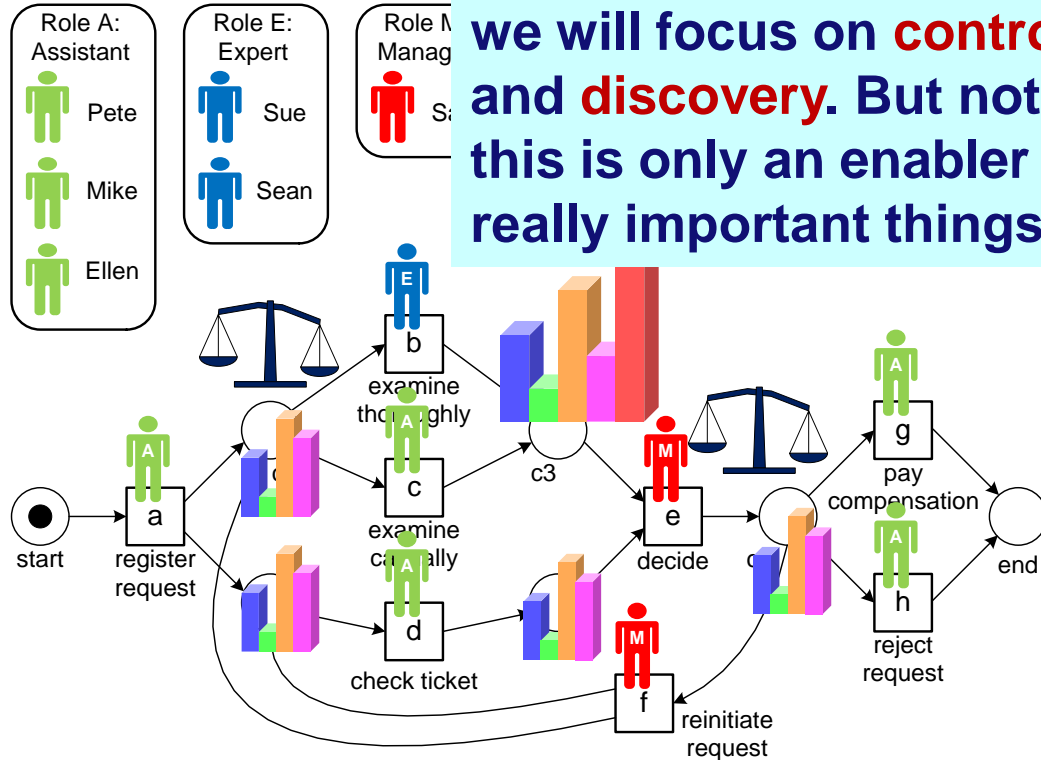
- Case starts with **a** and ends with **g** or **h**.
- Activity **d** is concurrent with **b** or **c**.
- Activity **e** has to wait until (**d** and **b**) or (**d** and **c**) have completed.
- There are three possible decisions: **f**, **g**, or **h**.
- ...



# Additional perspectives

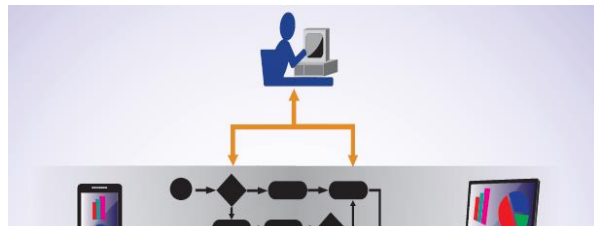
In the first part of this course, we will focus on **control-flow** and **discovery**. But note that this is only an enabler for the really important things ...

- control-flow
- data-flow
- time
- resources
- costs
- risks
- ...



# Alternative control-flow notations

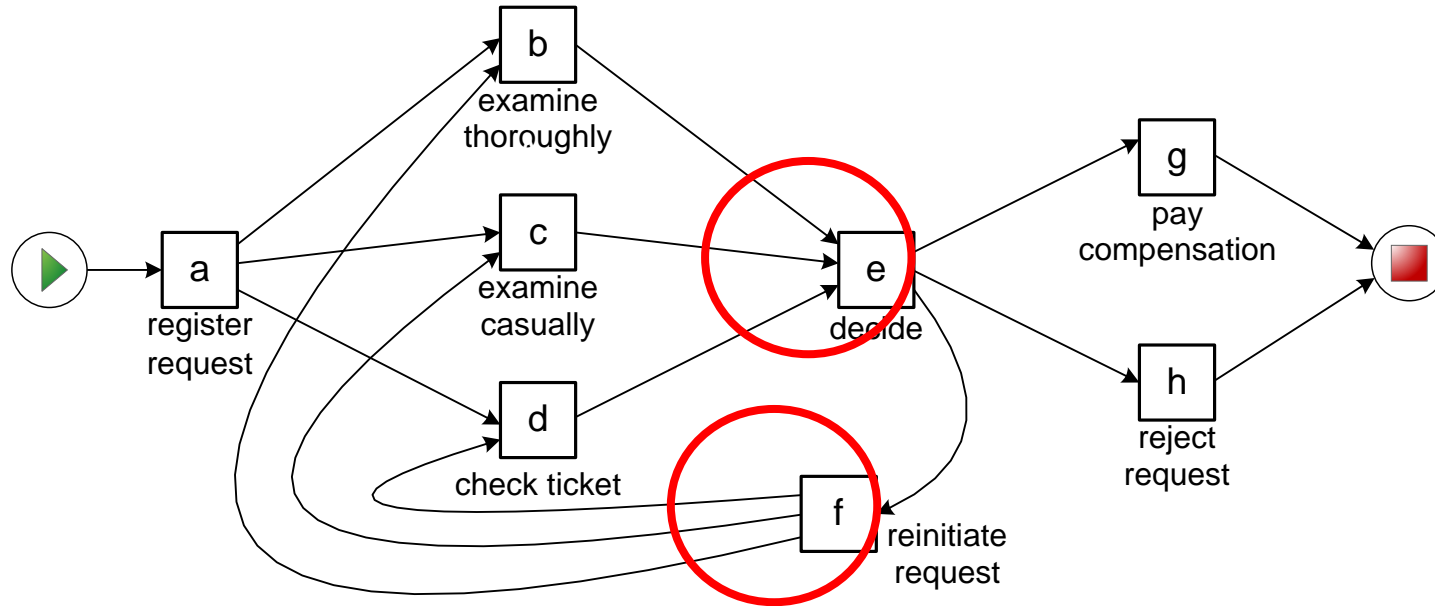
- BPMN (Business Process Model and Notation) diagrams
- UML activity diagrams
- Event-driven process chains
- Petri net variants
- Causal nets (C-nets)
- Transition systems
- (Hidden) Markov chains
- Process algebras (CSP, CCS, etc.)
- Fuzzy models
- YAWL models
- Declare models
- ...



## Notation:

- **search space: finding a model that captures reality well**
- **visualization: what do end-users need to see?**

# Class of "Fuzzy models"



**Fuzzy models are not executable like Petri nets, but they allow for a simpler representation.**



# Install ProM and Disco



# ProM: 600+ plug-ins

**dozens of model types**

**XES, MXML, CSV files**



[www.promtools.org](http://www.promtools.org)

[www.processmining.org](http://www.processmining.org)

# Disco: Simple, fast, and easy

**fuzzy models**



**XES, MXML, CSV files**

[www.fluxicon.com](http://www.fluxicon.com)

[www.fluxicon.com/academic/](http://www.fluxicon.com/academic/)

## *Part I: Introduction*

### **Chapter 1**

Data Science  
in Action

### **Chapter 2**

Process Mining:  
The Missing Link

## *Part II: Preliminaries*

### **Chapter 3**

Process Modeling  
and Analysis

### **Chapter 4**

Data Mining

## *Part III: From Event Logs to Process Models*

### **Chapter 5**

Getting the Data

### **Chapter 6**

Process Discovery:  
An Introduction

### **Chapter 7**

Advanced Process  
Discovery Techniques

## *Part IV: Beyond Process Discovery*

### **Chapter 8**

Conformance  
Checking

### **Chapter 9**

Mining Additional  
Perspectives

### **Chapter 10**

Operational Support

## *Part V: Putting Process Mining to Work*

### **Chapter 11**

Process Mining  
Software

### **Chapter 12**

Process Mining in the  
Large

### **Chapter 13**

Analyzing “Lasagna  
Processes”

### **Chapter 14**

Analyzing “Spaghetti  
Processes”

## *Part VI: Reflection*

### **Chapter 15**

Cartography and  
Navigation

### **Chapter 16**

Epilogue

