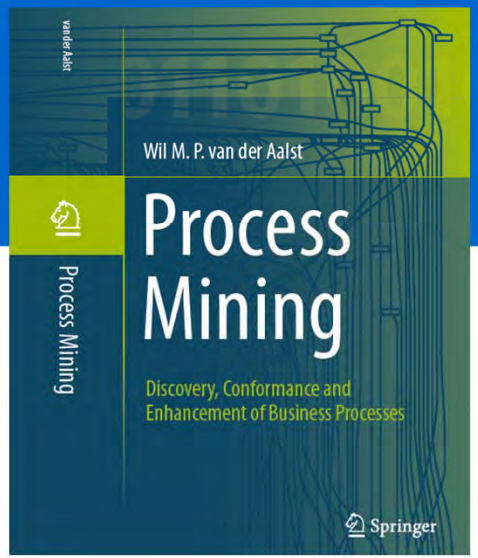
Process Mining (Manifesto)

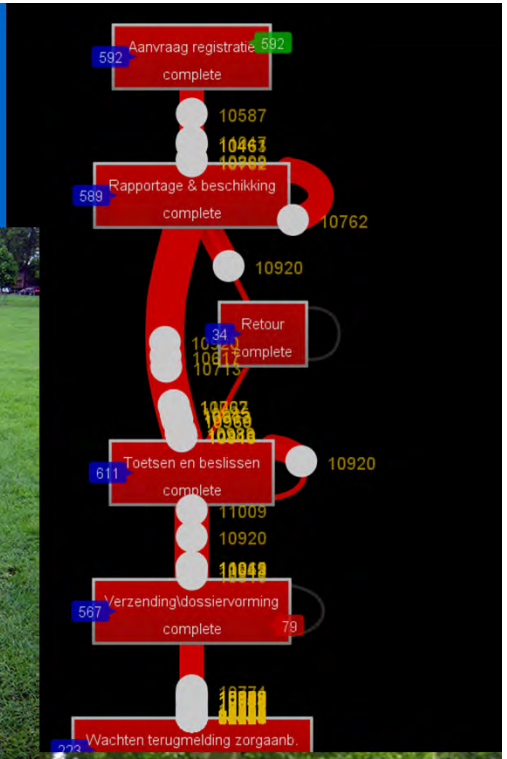
prof.dr.ir. Wil van der Aalst
www.processmining.org

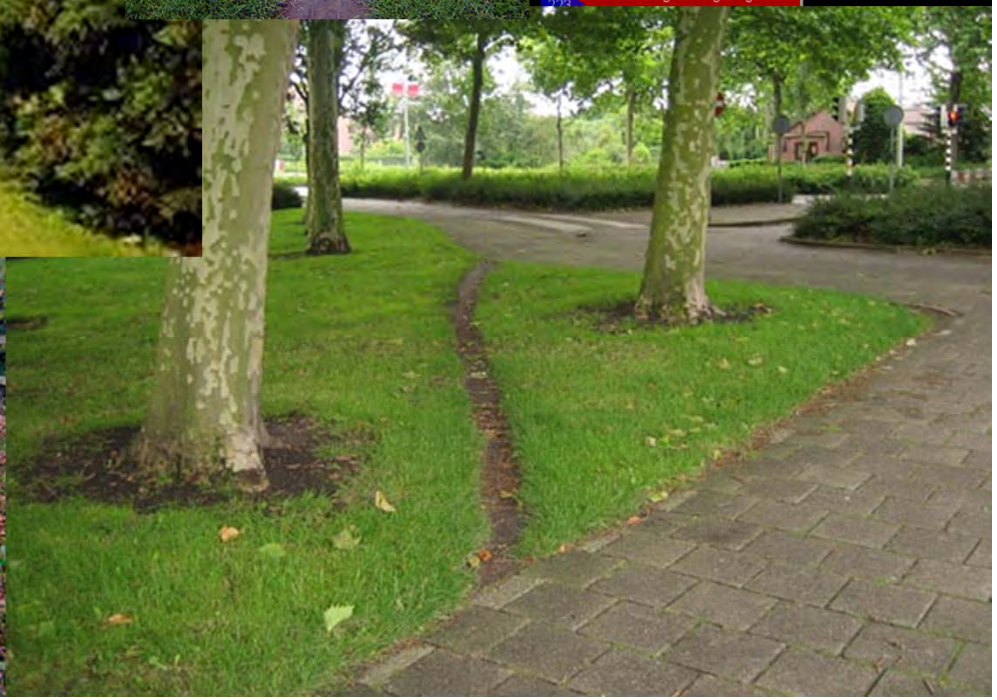


TU/e Technische Universiteit **Eindhoven** University of Technology

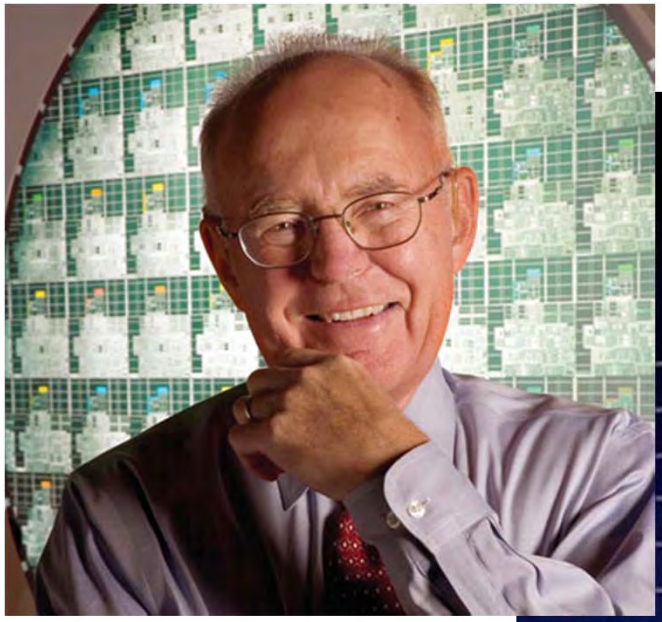
Where innovation starts

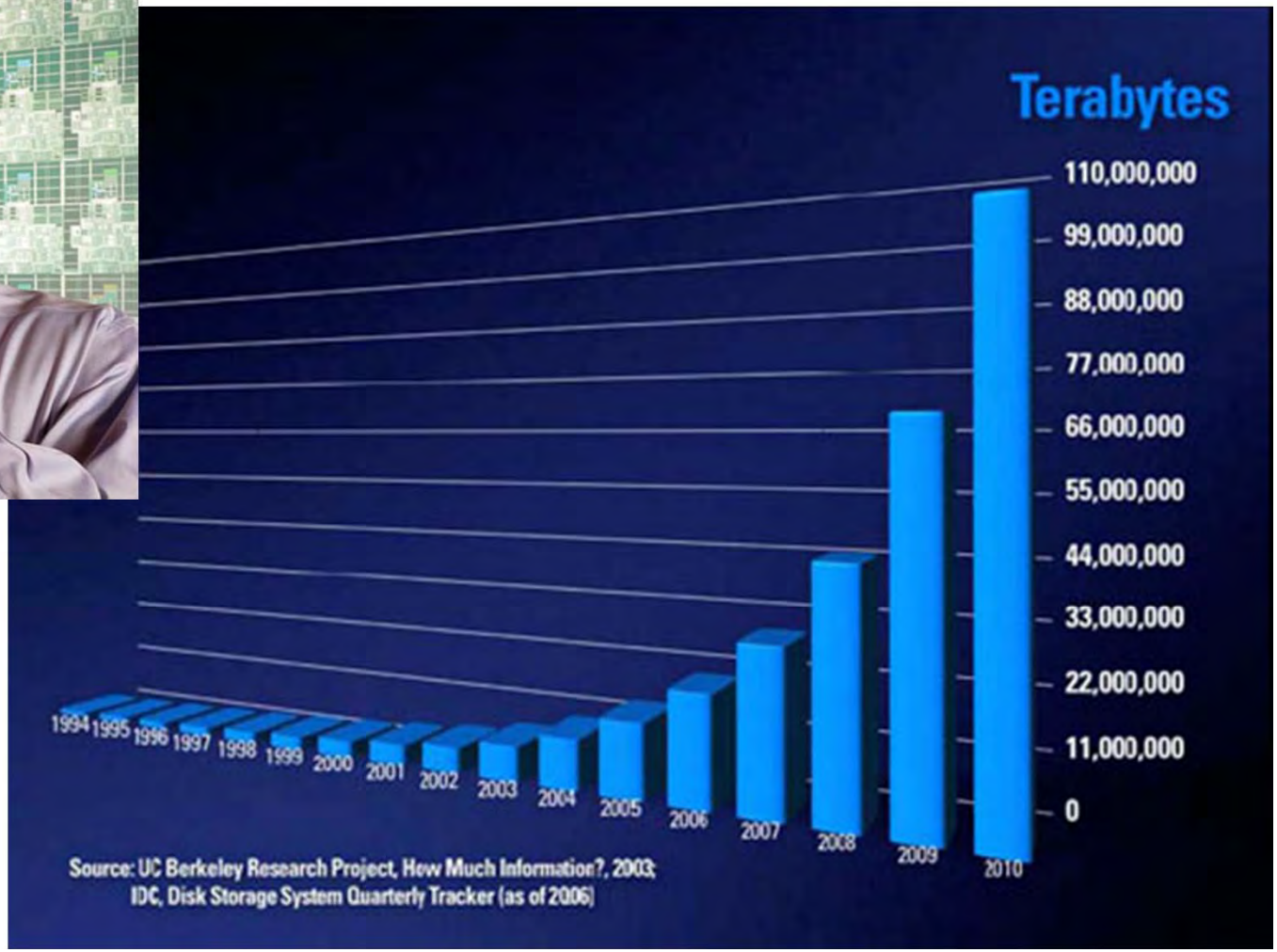
Desire lines in process models



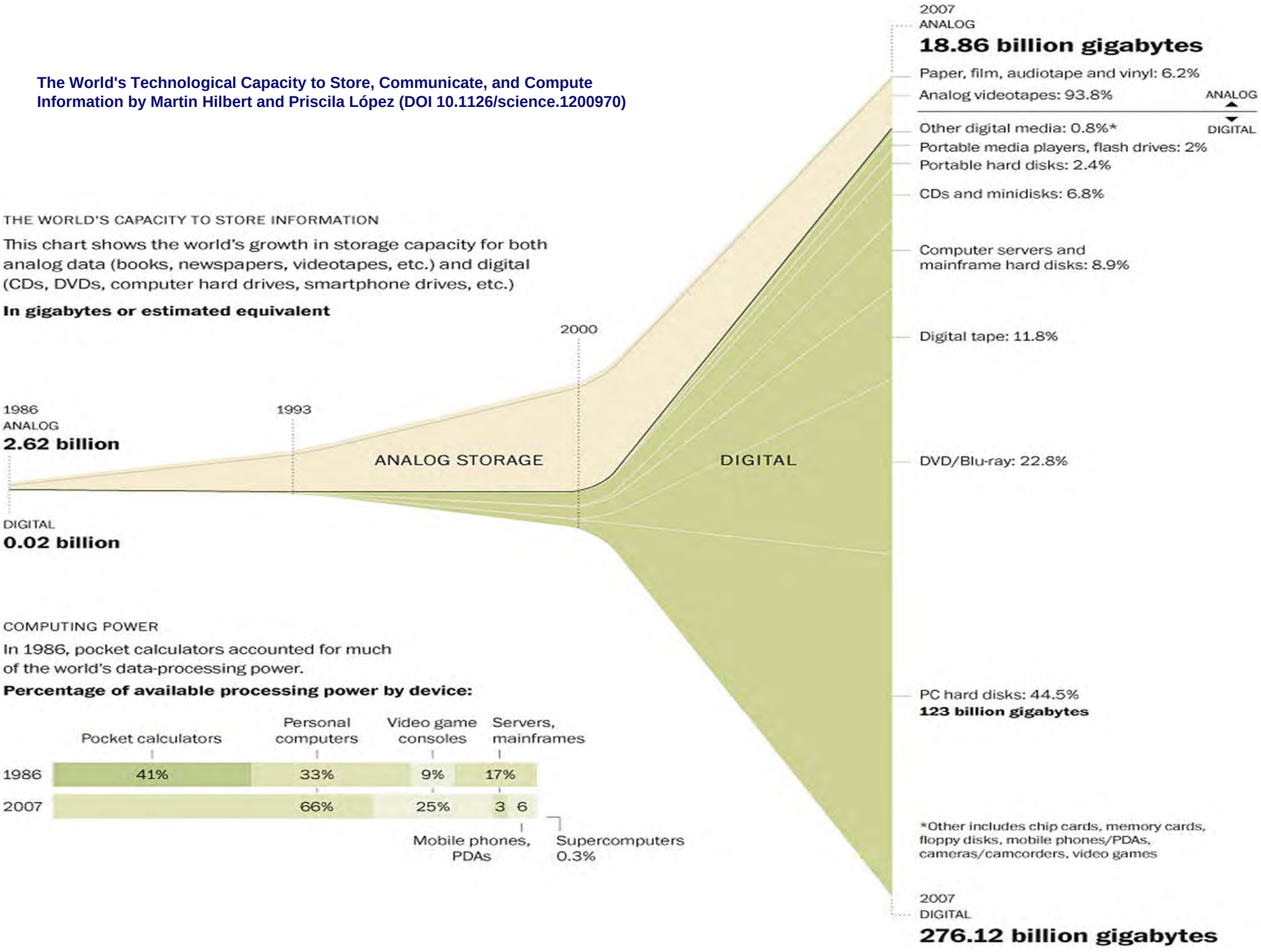


Data explosion





The World's Technological Capacity to Store, Communicate, and Compute Information by Martin Hilbert and Priscila López (DOI 10.1126/science.1200970)

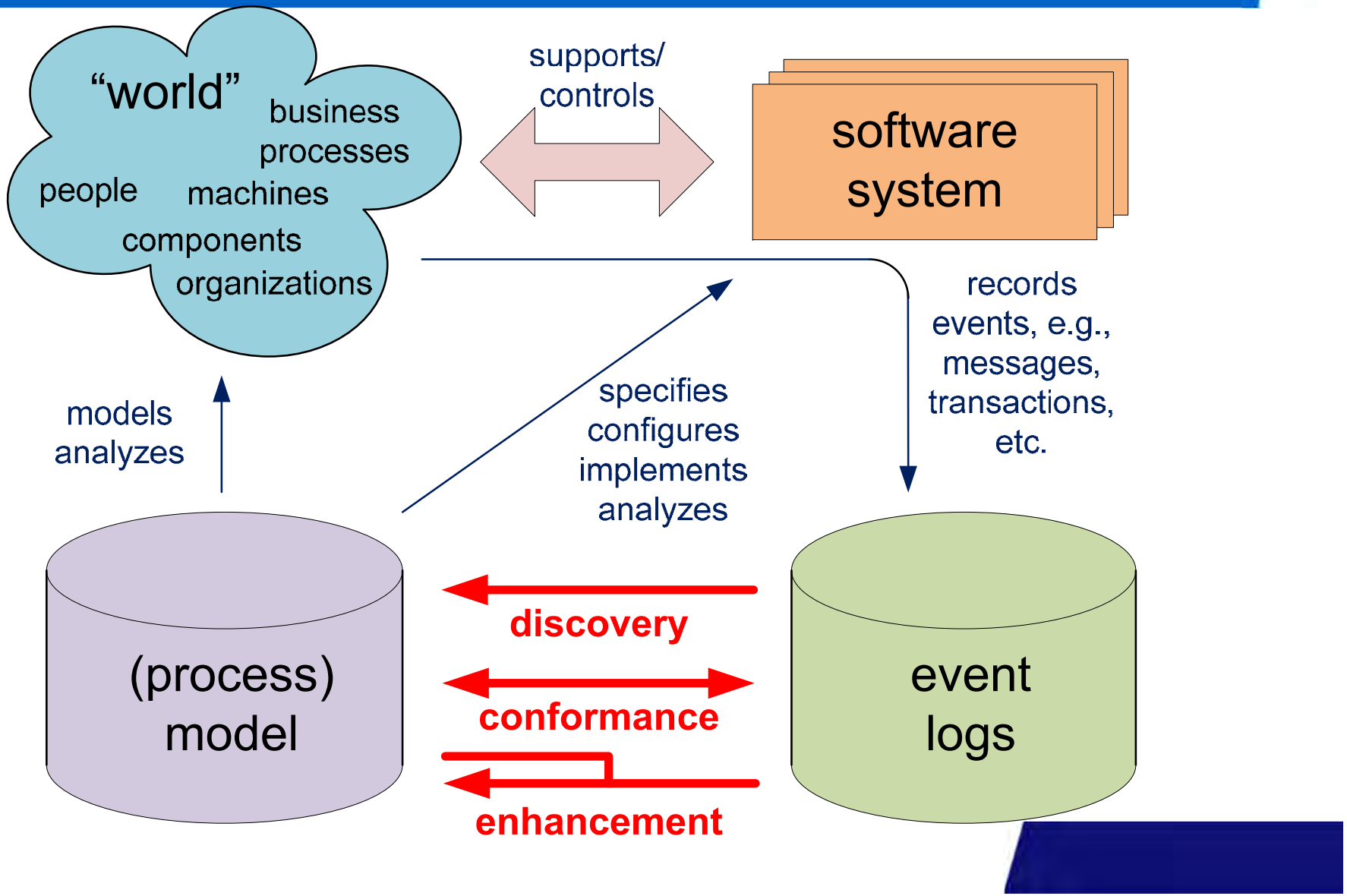



Process Mining =
Event Data + Processes
Data Mining + Process Analysis

Process Mining

- **Process discovery: "What is really happening?"**
- **Conformance checking: "Do we do what was agreed upon?"**
- **Performance analysis: "Where are the bottlenecks?"**
- **Process prediction: "Will this case be late?"**
- **Process improvement: "How to redesign this process?"**
- **Etc.**

Process Mining



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			
	35654488	05-01-2011:11.22	decide			
	35654489	08-01-2011:12.05	pay compensation			
3	35654521	30-12-2010:14.32	register request			
	35654522	30-12-2010:15.06	examine casually			
	35654524	30-12-2010:16.34	check ticket			
	35654525	06-01-2011:09.18	decide			
	35654526	06-01-2011:12.18	reinitiate request			
	35654527	06-01-2011:13.06	examine thoroughly			
	35654530	08-01-2011:11.43	check ticket			
	35654531	09-01-2011:09.55	decide			
	35654533	15-01-2011:10.45	pay compensation			
4	35654641	06-01-2011:15.02	register request			
	35654643	07-01-2011:12.06	check ticket			
	35654644	08-01-2011:14.43	examine thoroughly			
	35654645	09-01-2011:12.02	decide			
	35654647	12-01-2011:15.44	reject request			
5	35654711	06-01-2011:09.02	register request			
	35654712	07-01-2011:10.16	examine casually			
	35654714	08-01-2011:11.22	check ticket			
	35654715	10-01-2011:13.28	decide			
	35654716	11-01-2011:16.18	reinitiate request			
	35654718	14-01-2011:14.33	check ticket			
	35654719	16-01-2011:15.50	examine casually			
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...

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

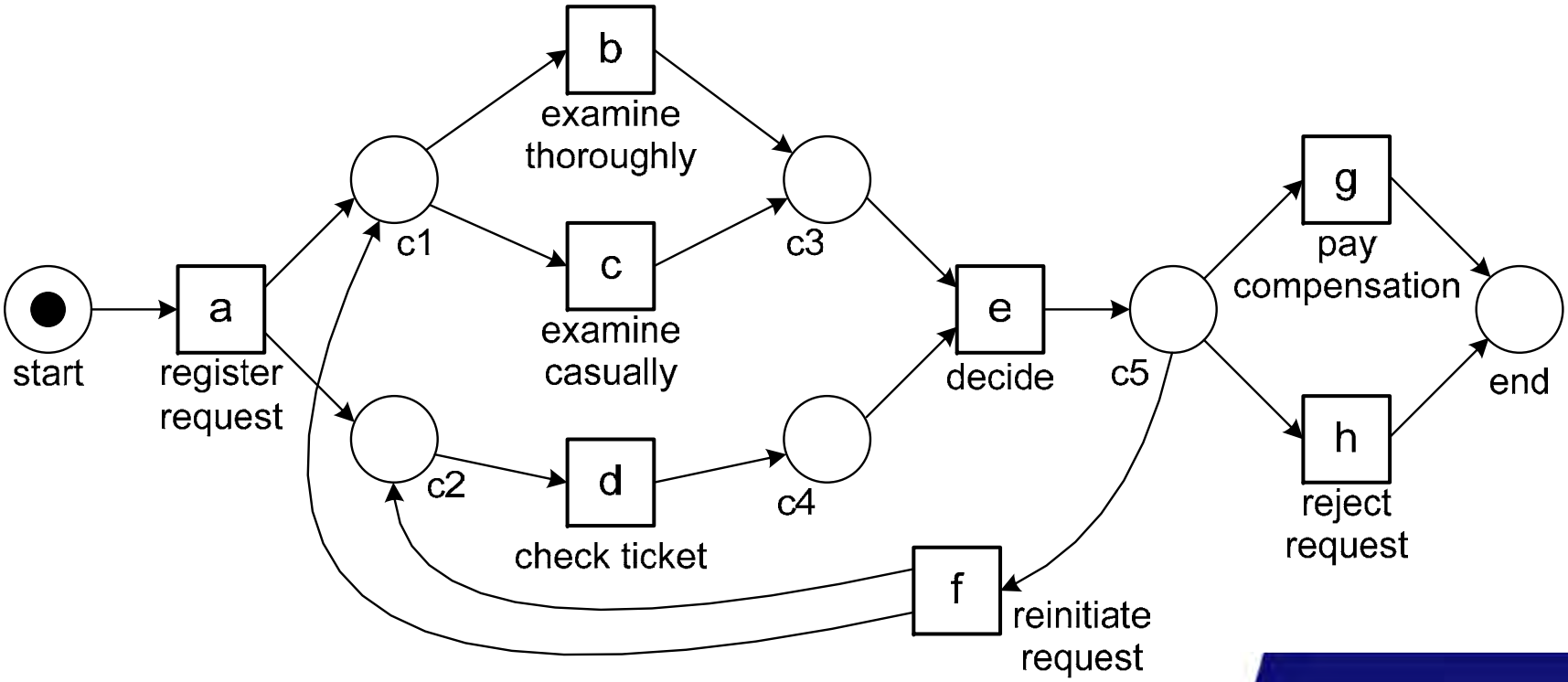
case id	event id	properties				
		timestamp	activity	resource		
1	35654423	30-12-2010:11.02	register request	Pete		
	35654424	31-12-2010:10.06	examine thoroughly	Sue		
	35654425	05-01-2011:15.12	check ticket	Mike		
	35654426	06-01-2011:11.18	decide	Sara		
	35654427	07-01-2011:14.24	reject request	Pete		
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	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	...
5	35654711	06-01-2011:09.02	register request	Ellen	50	...
	35654712	07-01-2011:10.16	examine casually	Mike	400	...
	35654714	08-01-2011:11.22	check ticket	Pete	100	...
	35654715	10-01-2011:13.28	decide	Sara	200	...
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	...
	35654718	14-01-2011:14.33	check ticket	Ellen	100	...
	35654719	16-01-2011:15.50	examine casually	Mike	400	...
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...

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

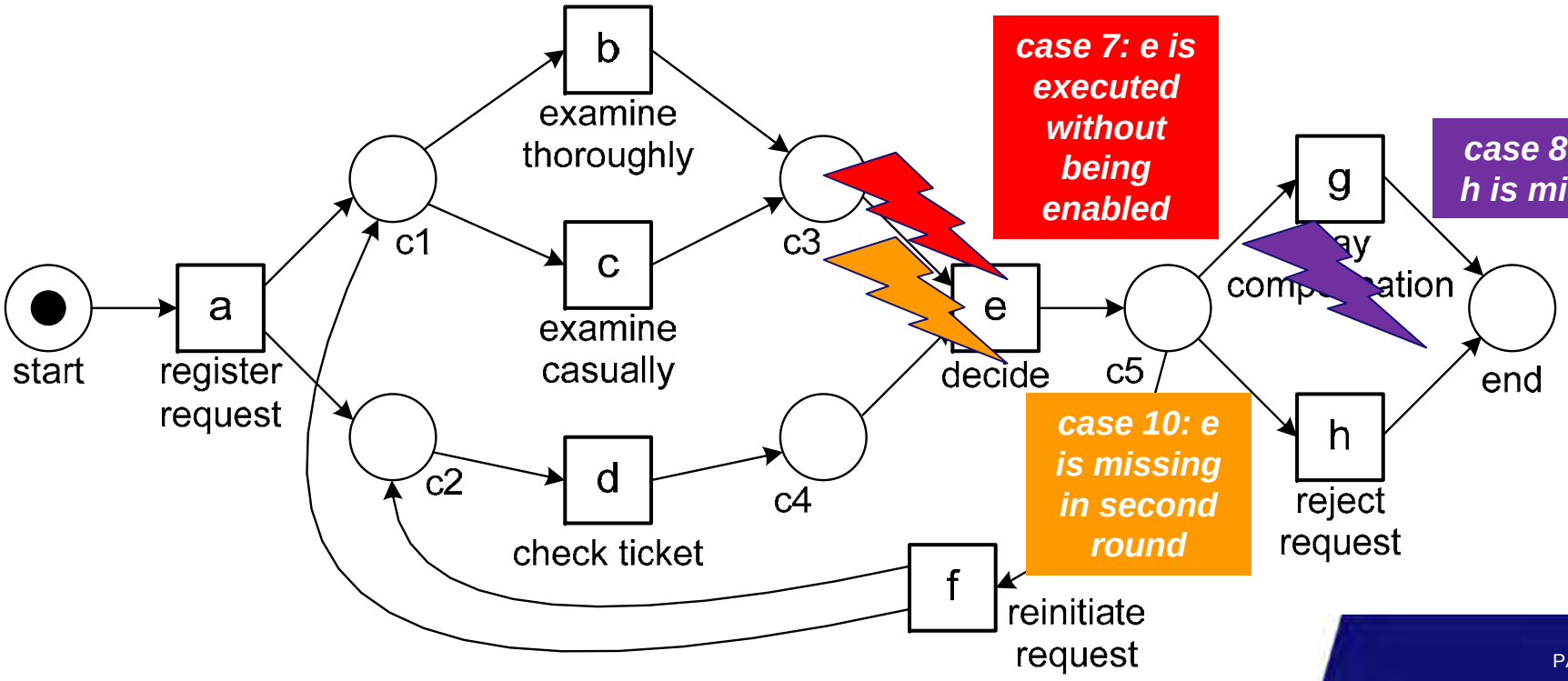
Process discovery

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	$\langle a,c,d,e,g \rangle$
...	...

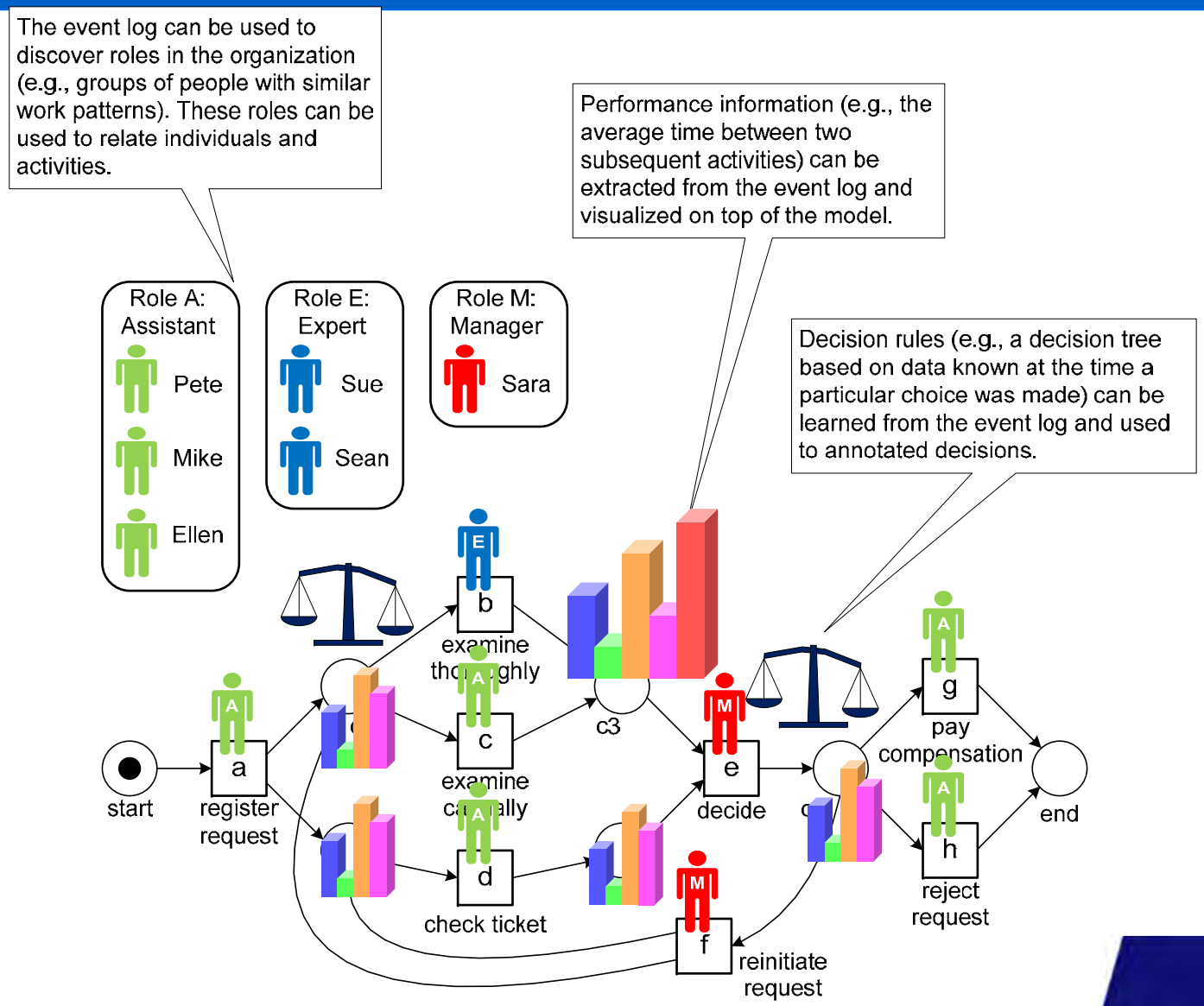


Conformance checking

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	$\langle a,c,d,e,g\rangle$
7	$\langle \mathbf{a},\mathbf{b},\mathbf{e},\mathbf{g}\rangle$
8	$\langle \mathbf{a},\mathbf{b},\mathbf{d},\mathbf{e}\rangle$
9	$\langle a,d,c,e,f,d,c,e,f,b,d,e,h\rangle$
10	$\langle \mathbf{a},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{f},\mathbf{b},\mathbf{d},\mathbf{g}\rangle$



Extension: Adding perspectives to model based on event log



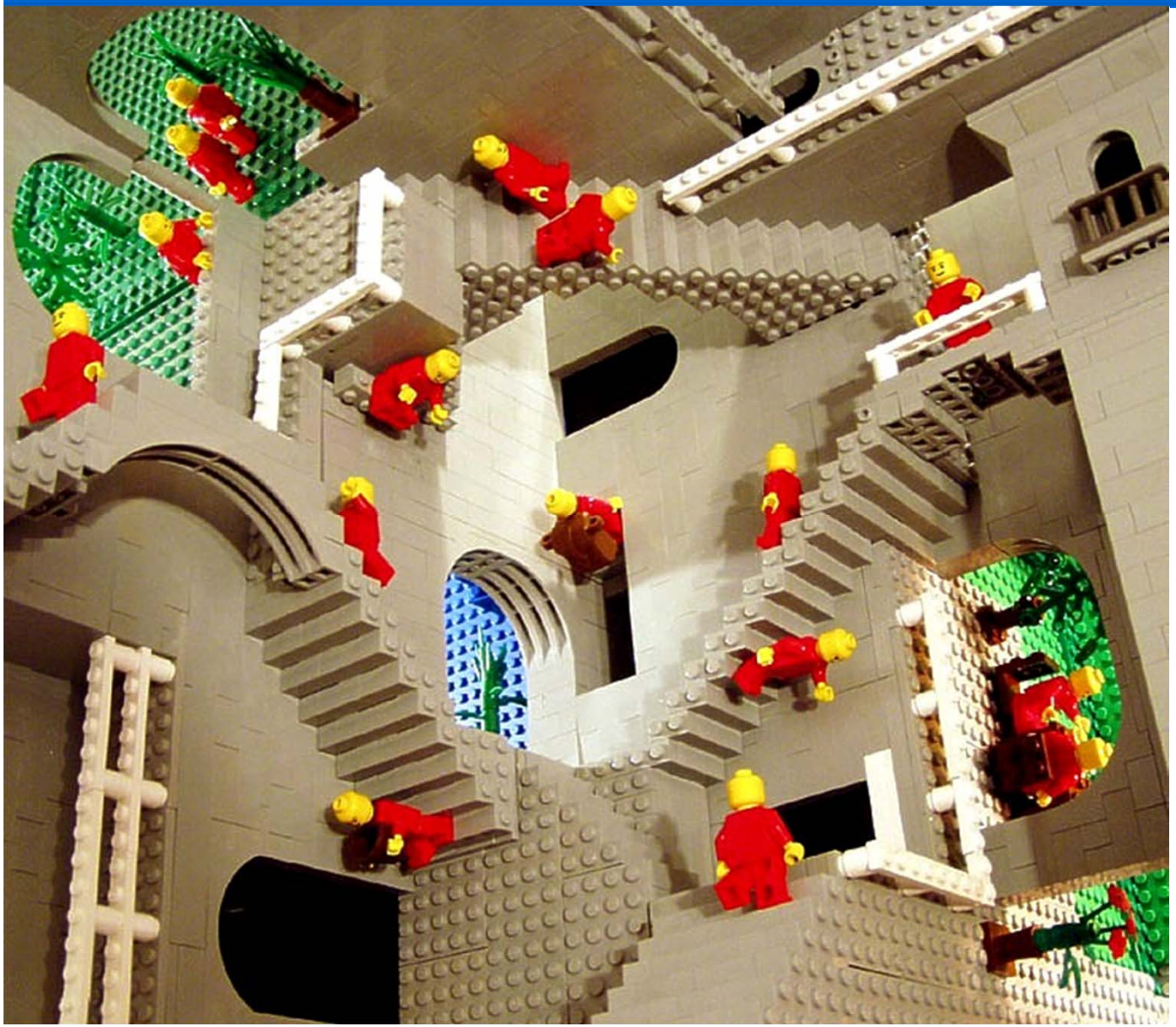
We applied ProM in >100 organizations

- **Municipalities** (e.g., Alkmaar, Heusden, Harderwijk, etc.)
- **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 and their customers** (e.g., Philips Healthcare, ASML, Ricoh, Thales)
- **Media companies** (e.g. Winkwaves)
- **...**

All supported by …

- Open-source (L-GPL), cf. www.processmining.org
- Plug-in architecture
- Plug-ins cover the whole process mining spectrum and also support classical forms of process analysis

Let us play …

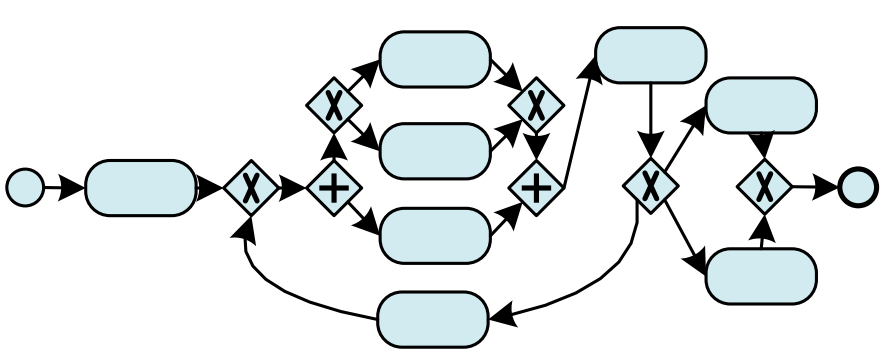


Play-Out

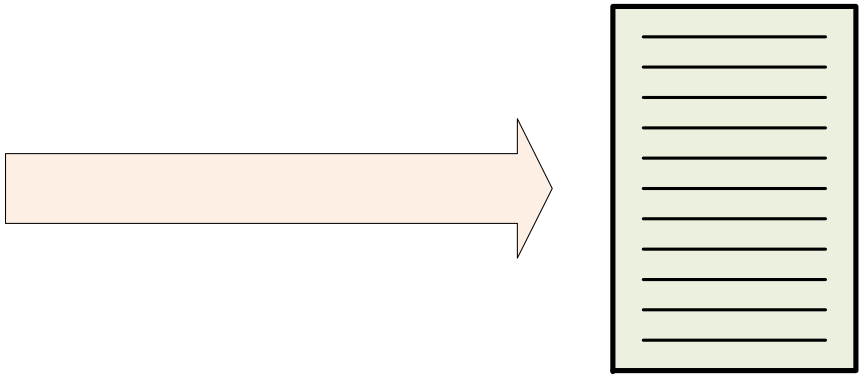
Play-In

Replay

Play-Out

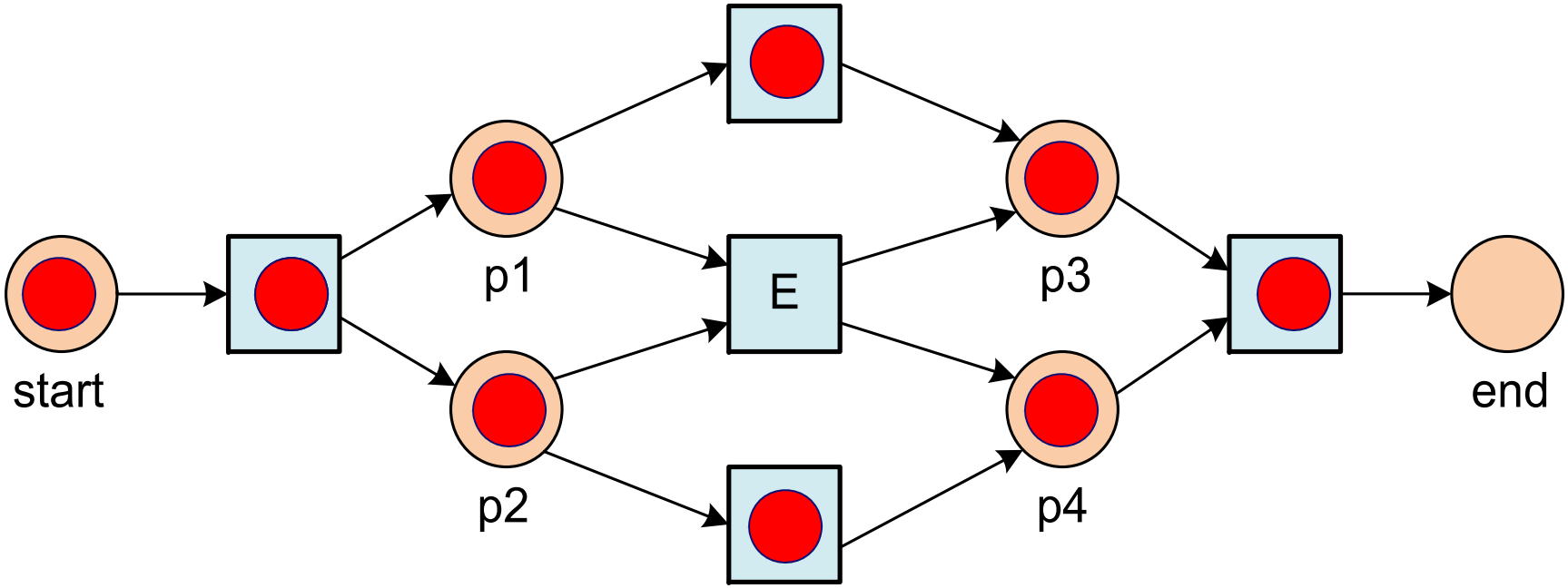


process model



event log

Play-Out (Classical use of models)

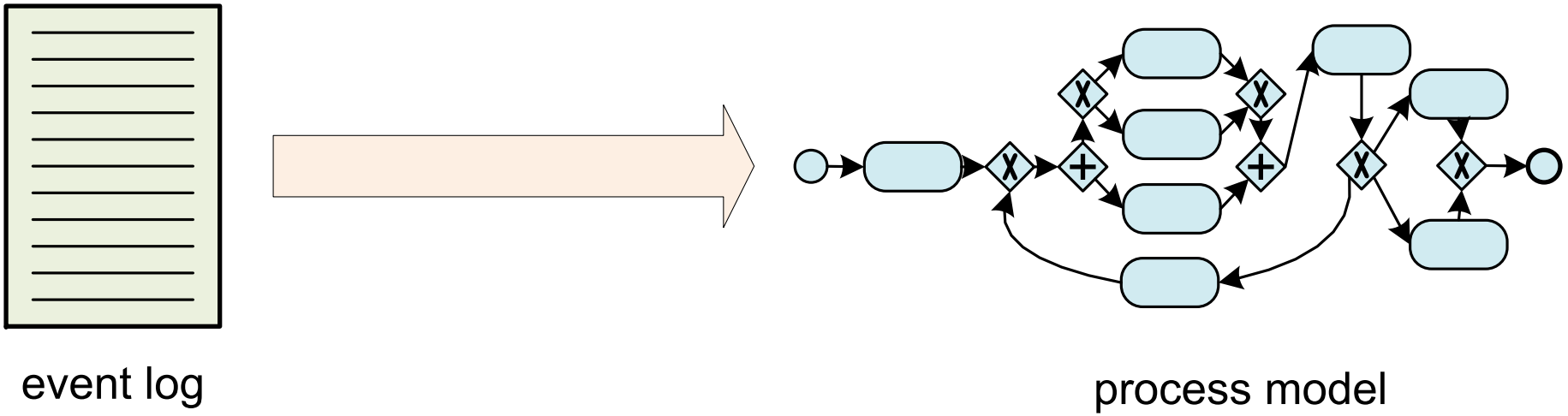


A B C D A E D A E D

A B C D A C B D

A C B D A E D A C B D

Play-In



event log

process model

Play-In

A B C D

A E D

A E D

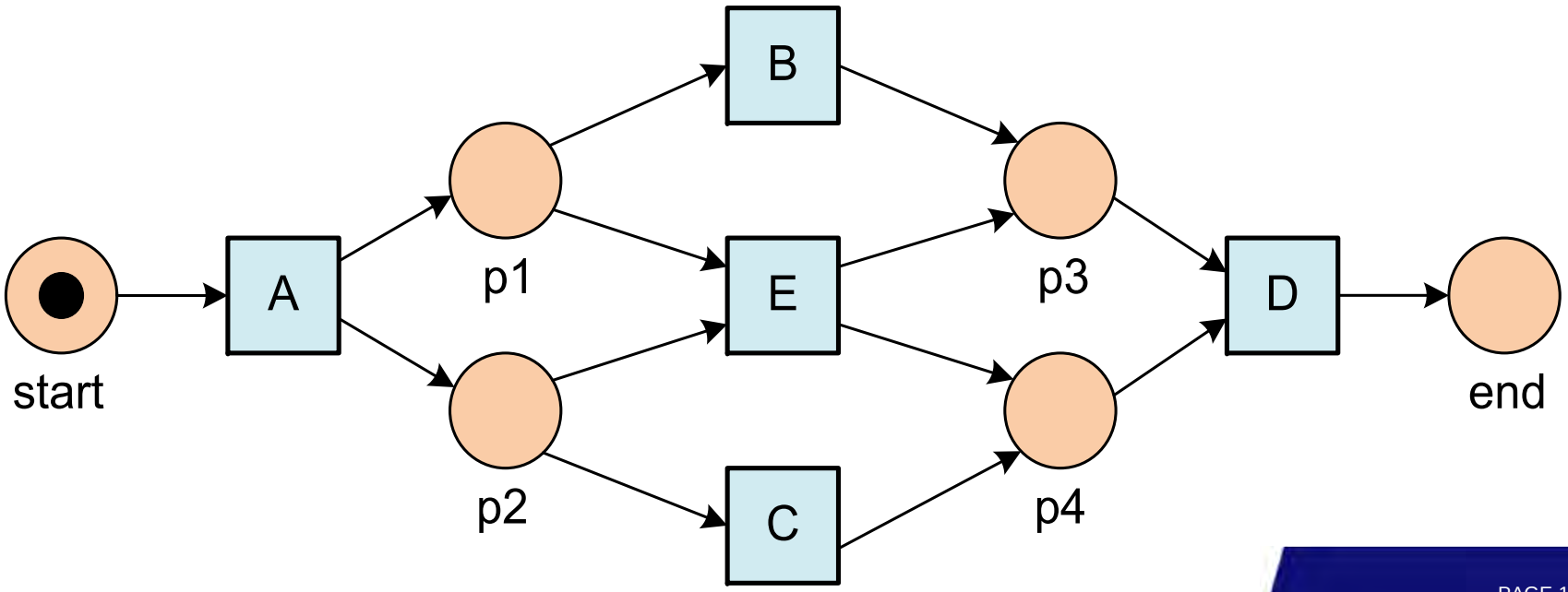
A B C D

A C B D

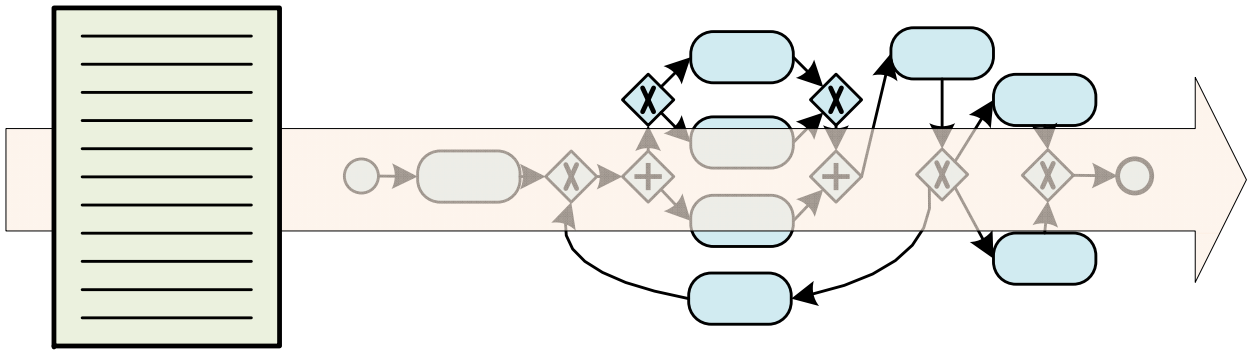
A C B D

A E D

A C B D



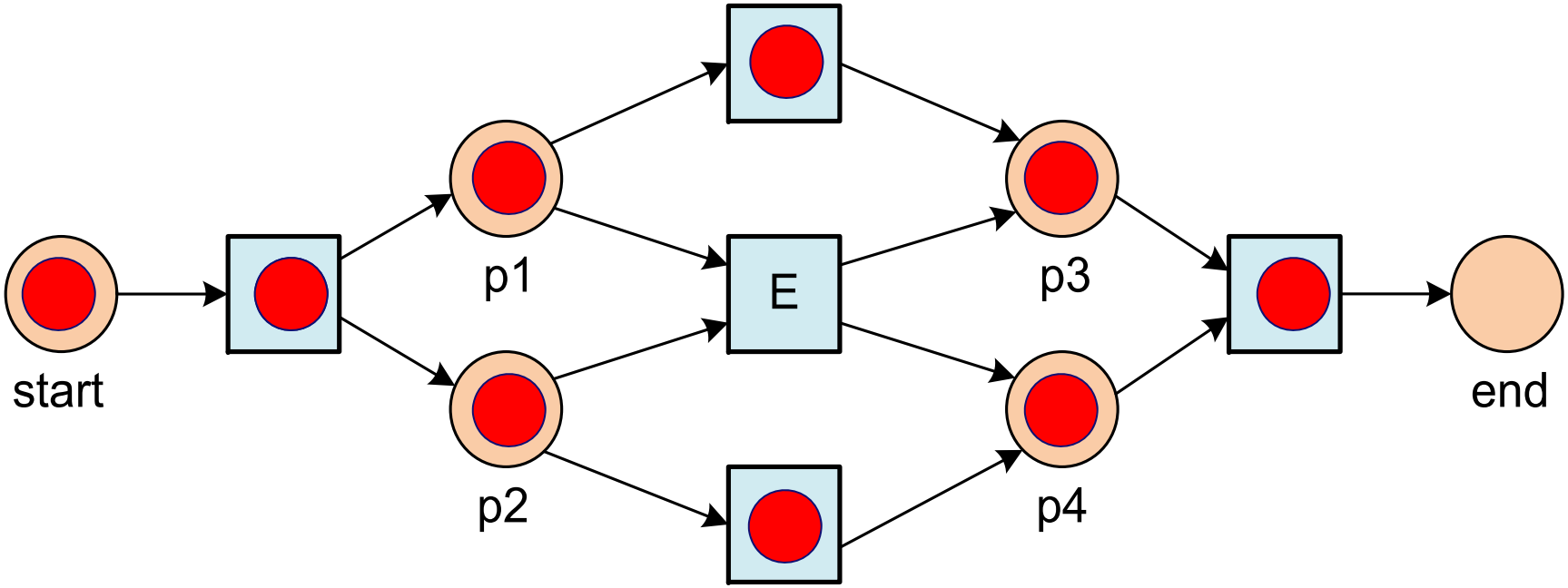
Replay



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

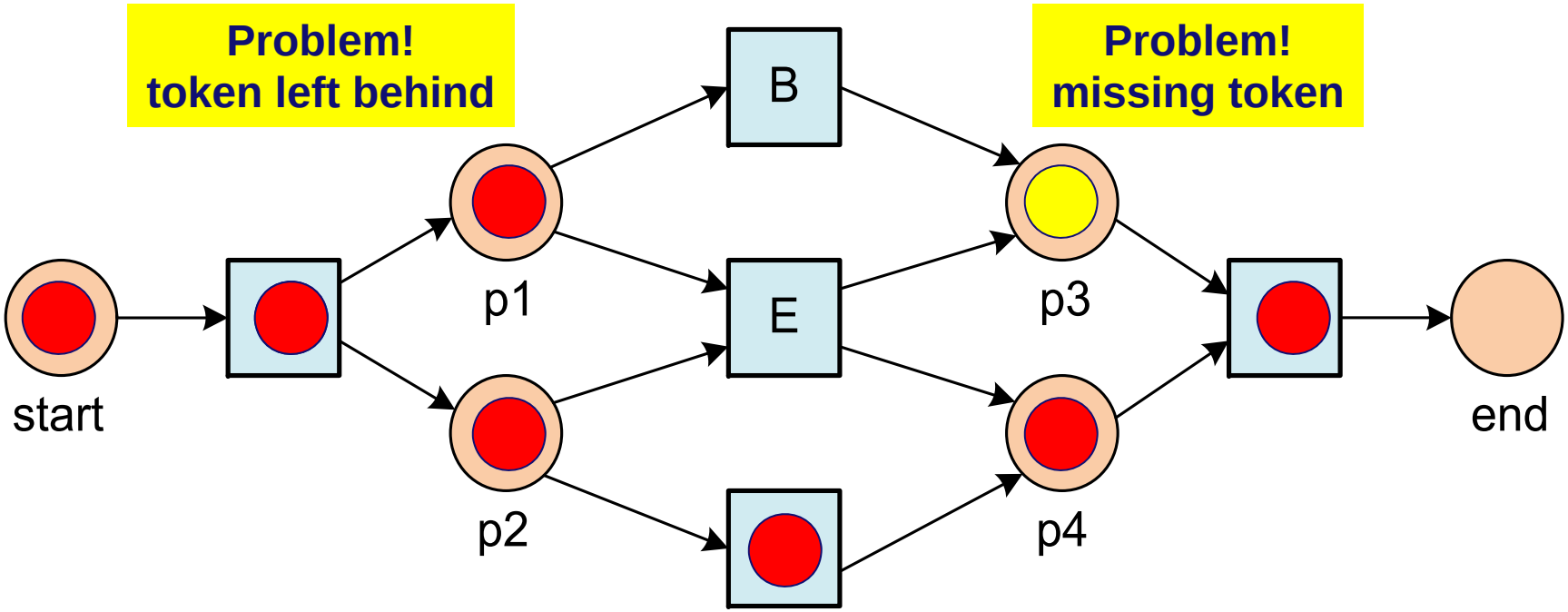
Replay

A B C D



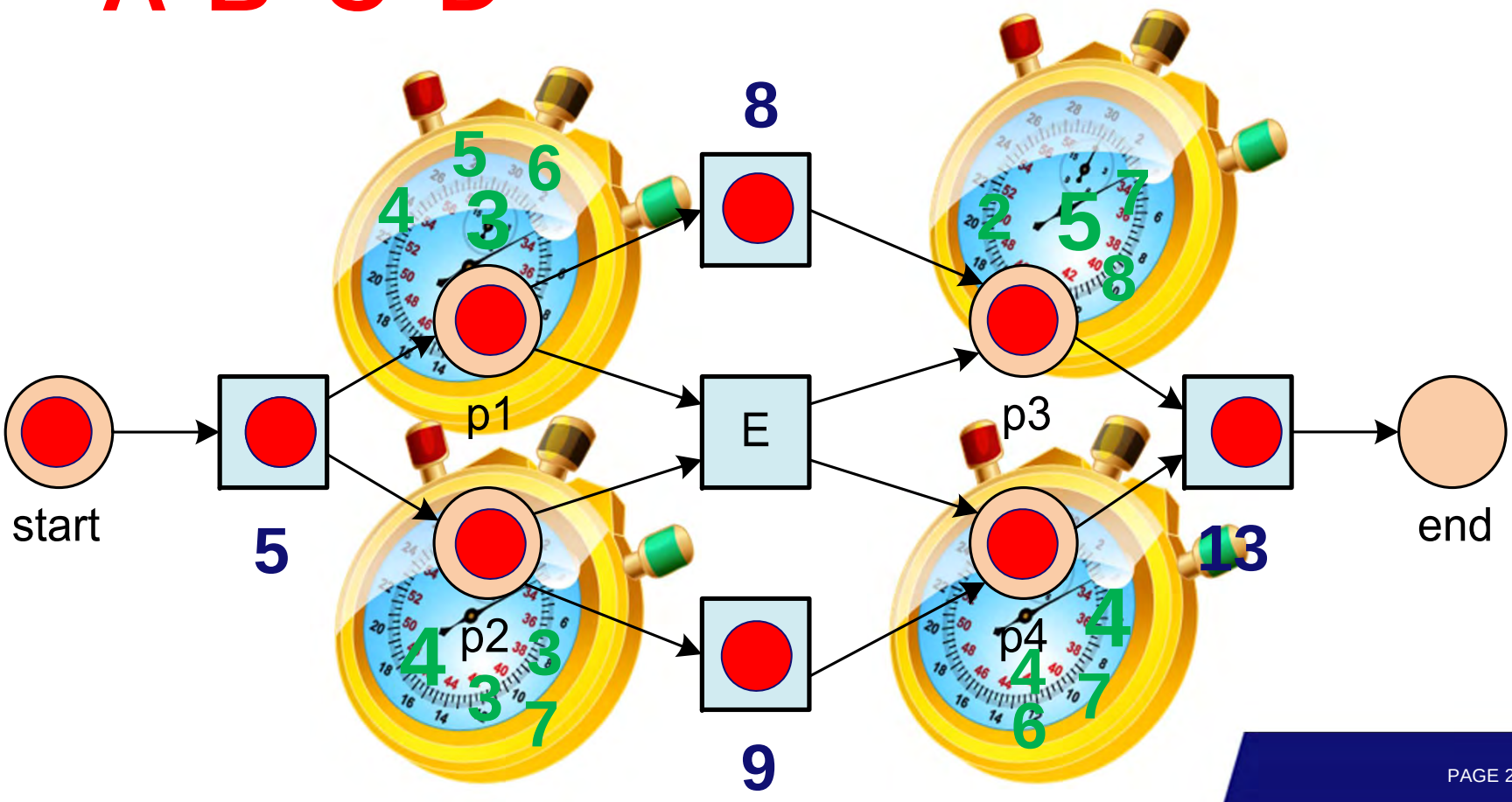
Replay can detect problems

A C D

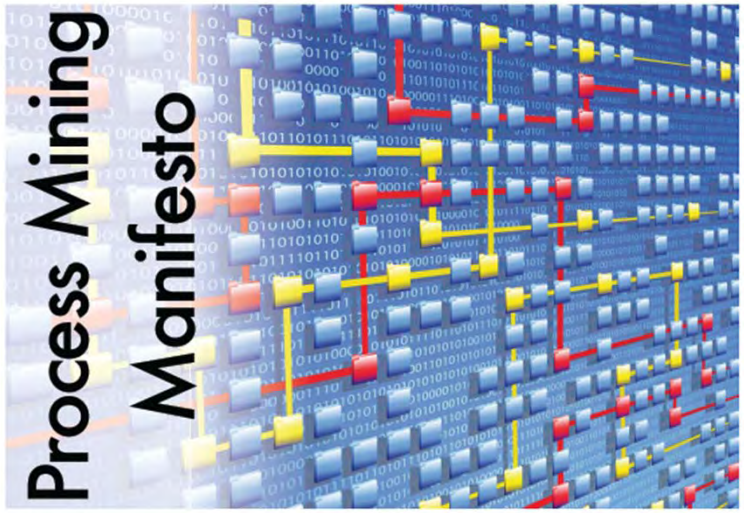


Replay can extract timing information

$A^5 B^8 C^9 D^{13}$



Process Mining Manifesto



A *manifesto* is a "public declaration of principles and intentions" by a group of people. This manifesto is written by members and supporters of the *IEEE Task Force on Process Mining*. The goal of this task force is to promote the research, development, education, implementation, evolution, and understanding of process mining.

Process mining is a relatively young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. 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 (information) systems. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/ organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations.

Contents:

Process mining techniques are able to extract knowledge from event logs commonly available in today's information systems. These techniques provide new means to *discover, monitor,* and *improve* processes in a variety of application domains. There are two main drivers for the growing interest in process mining. On the one hand, more and more events are being recorded, thus, providing detailed information about the history of processes. On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments. This manifesto is created by the *IEEE Task Force on Process Mining* and aims to promote the topic of process mining. Moreover, by defining a set of guiding principles and listing important challenges, this manifesto hopes to serve as a *guide for software developers, scientists, consultants, business managers,* and *end-users*. The goal is to increase the maturity of process mining as a new tool to improve the (re)design, control, and support of operational business processes.

1

- On 7 October 2011, the IEEE Task Force on Process Mining released the Process Mining Manifesto
- 53 organizations support the manifesto
- 77 process mining experts contributed to it

Guiding Principles

GP1 : Event data should be treated as first-class citizens.	Events should be trustworthy; that is, it should be safe to assume that the recorded events actually happened and that the attributes of events are correct. Event logs should be complete; given a particular scope, no events may be missing. Any recorded event should have well-defined semantics. Moreover,
GP2: Log extrac should be driver questions.	for example, SAP. Without
GP3: Process-m techniques shou support concurr choice, and othe control-flow cor	Cs, Petri nets, oice (XOR- rns.
GP4: Events sho related to model elements.	nts in the play" the event mple, some th additional mps).
GP5: Models should be treated as purposeful abstractions of reality.	A model derived from event data provides a view on reality. Such a view should serve as a purposeful abstraction of the behavior captured in the event log. Given an event log, multiple useful views might exist.
GP6: Process mining should be a continuous process.	Given the dynamic nature of processes, we shouldn't view process mining as a one-time activity. The goal should be not to create a fixed model, but to breathe life into process models in a way that encourages users and analysts to look at them on a daily basis.

GP4: Events should be related to model elements.

Conformance checking and enhancement rely heavily on the relationship between elements in the model and events in the log. This relationship can be used by process mining tools to "replay" the event log on the model. Replay can reveal discrepancies between event log and model (for example, some events in the log aren't possible according to the model). It can also enrich the model with additional information from the event log (for example, it can identify bottlenecks by using timestamps).

Challenges (1/2)

C1: Finding, merging, and cleaning event data	When extracting event data suitable for process mining, we must address several challenges: data can be distributed over a variety of sources, event data might be incomplete, an event log could contain outliers, logs could contain events at different
C4: Dealing with concept drift	The process might be changing while under analysis. Understanding such concept drifts is of prime importance for process management.
C5: Improving the representational bias used for process discovery	A careful and refined selection of the representational bias is necessary to ensure high-quality process-mining results.
C6: Balancing between quality criteria such as fitness, simplicity, precision, and generalization	Four competing quality dimensions exist: fitness, simplicity, precision, and generalization. The challenge is to find models that can balance all four dimensions.

C4: Dealing with concept drift

The process might be changing while under analysis. Understanding such concept drifts is of prime importance for process management.

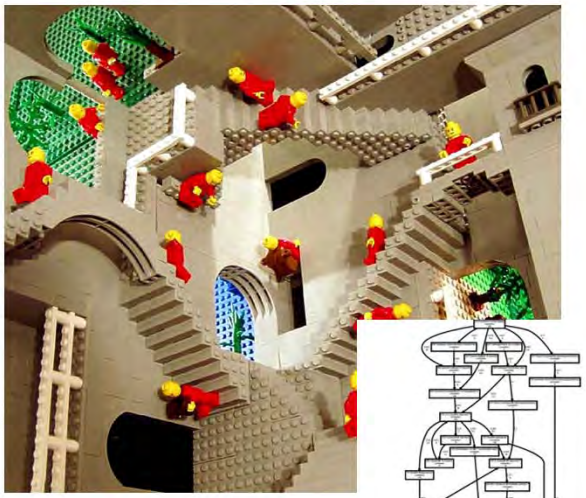
Challenges (2/2)

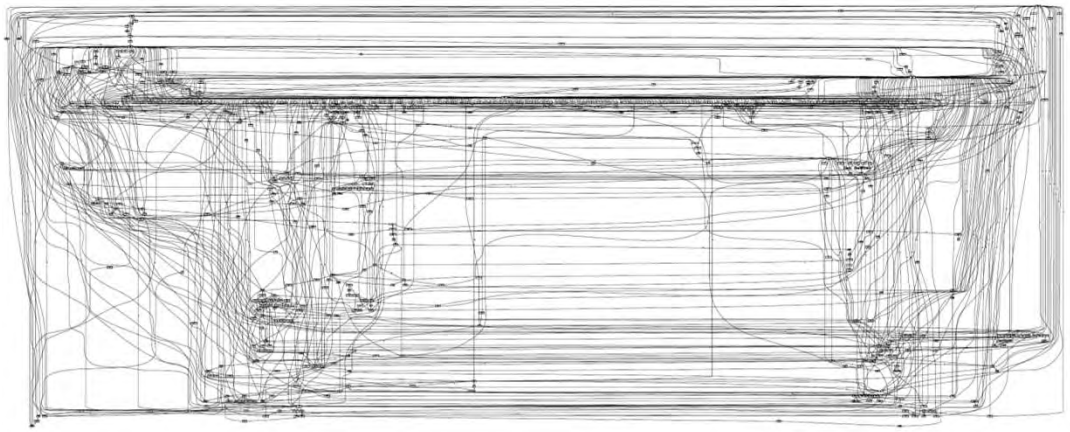
C7: Cross-organizational mining	In some use cases, event logs from multiple organizations are available for analysis. Some organizations, such as supply chain partners, work together to handle process instances; other organizations execute essentially the same process while sharing experiences, knowledge, or a common infrastructure. However, traditional process-mining techniques typically consider one event log in one organization.
C8: P opera	
C9: C proce with analy	
C10: usabi exper	
C11: Improving understandability for non-experts	conclusions. To avoid such problems, process mining tools should present results using a suitable representation and the trustworthiness of the results should always be clearly indicated.

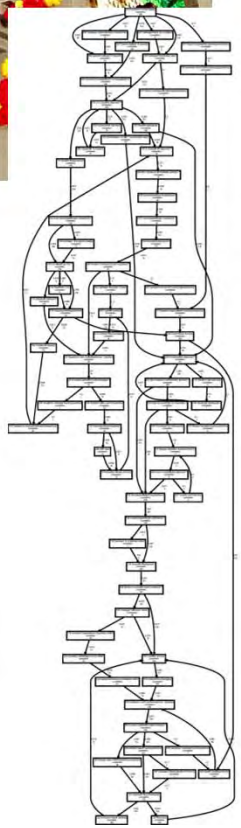
C7: Cross-organizational mining

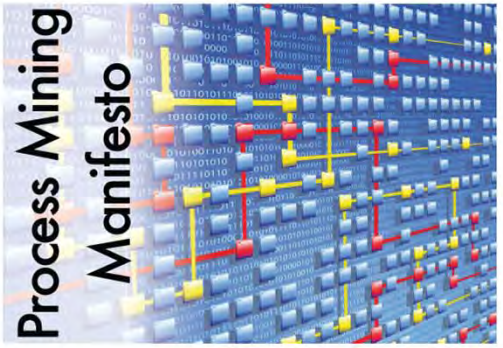
In some use cases, event logs from multiple organizations are available for analysis. Some organizations, such as supply chain partners, work together to handle process instances; other organizations execute essentially the same process while sharing experiences, knowledge, or a common infrastructure. However, traditional process-mining techniques typically consider one event log in one organization.

Conclusion









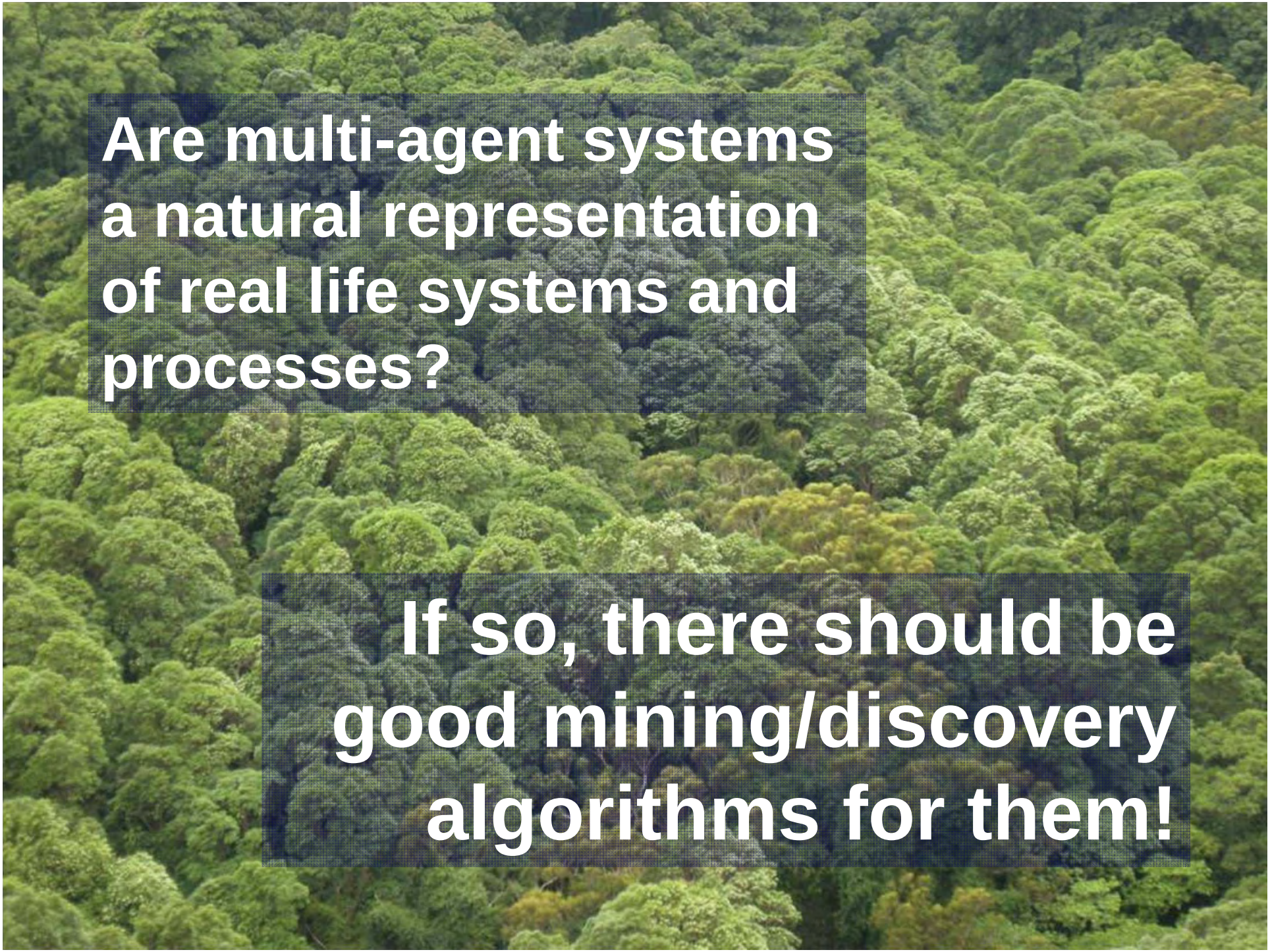
A *manifesto* is a "public declaration of principles and intentions" by a group of people. This manifesto is written by members and supporters of the *IEEE Task Force on Process Mining*. The goal of this task force is to promote the research, development, education, implementation, evolution, and understanding of process mining.

Process mining is a relatively young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. 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 (information) systems. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/ organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations.

Contents:

Process mining techniques are able to extract knowledge from event logs commonly available in today's information systems. These techniques provide new means to discover, monitor, and improve processes in a variety of application domains. There are two main drivers for the growing interest in process mining. On the one hand, more and more events are being recorded, thus, providing detailed information about the history of processes. On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments. This manifesto is created by the IEEE Task Force on Process Mining and aims to promote the topic of process mining. Moreover, by defining a set of guiding principles and listing important challenges, this manifesto hopes to serve as a *guide for software developers, scientists, consultants, business managers, and end-users*. The goal is to increase the *maturity* of process mining as a new tool to improve the (re)design, control, and support of operational business processes.

1


Are multi-agent systems a natural representation of real life systems and processes?
If so, there should be good mining/discovery algorithms for them!

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts. Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems.

Wil van der Aalst delivers the first book on process mining. It aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. In Part I, the author provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Part II focuses on process discovery as the most important process mining task. Part III moves beyond discovering the control flow of processes and highlights conformance checking, and organizational and time perspectives. Part IV guides the reader in successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM. Finally, Part V takes a step back, reflecting on the material presented and the key open challenges.

Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

Features and Benefits:

- First book on process mining, bridging the gap between business process modeling and business intelligence.
- Written by one of the most influential and most-cited computer scientists and the best-known BPM researcher.
- Self-contained and comprehensive overview for a broad audience in academia and industry.
- The reader can put process mining into practice immediately due to the applicability of the techniques and the availability of the open-source process mining software ProM.

Computer Science

ISBN 978-3-642-19344-6

springer.com

van der Aalst

Process Mining

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

www.processmining.org

www.win.tue.nl/ieeetfpm/

Springer