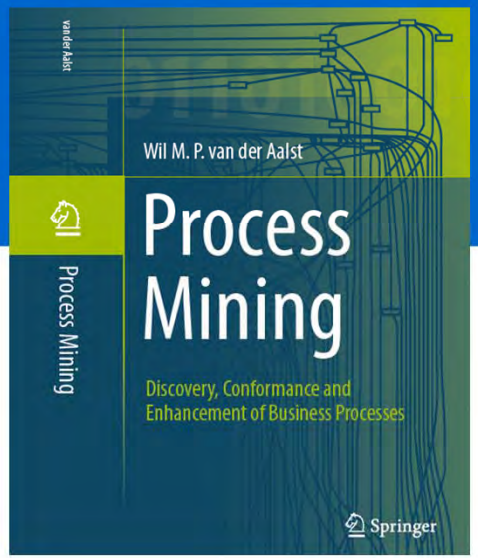
Cross-Organizational Process Configuration and Process Mining

Keynote PoEM, November 2011, Oslo, Norway

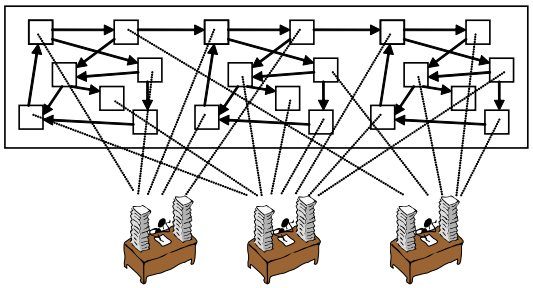
prof.dr.ir. Wil van der Aalst
www.processmining.org



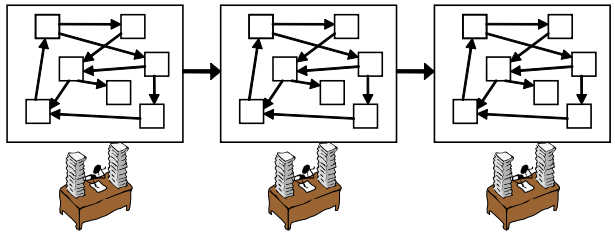
TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

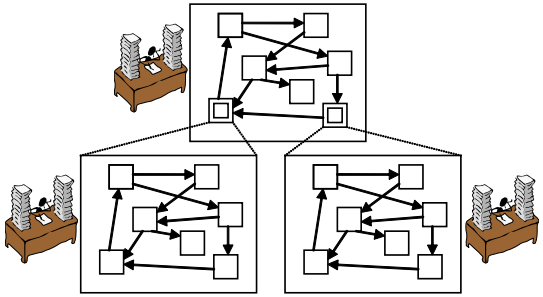
Distributing Work in a Collaborative Setting



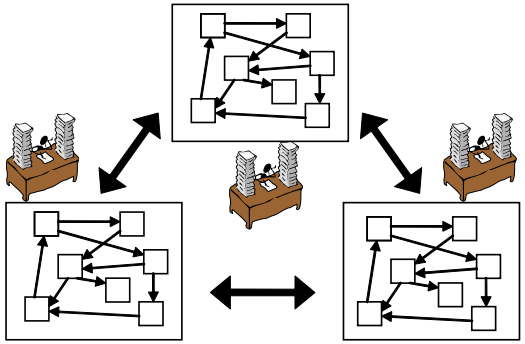
(a) Capacity sharing



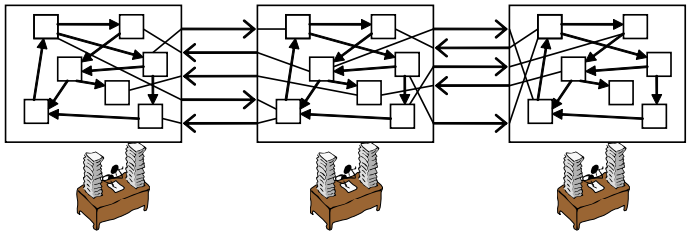
(b) Chained execution



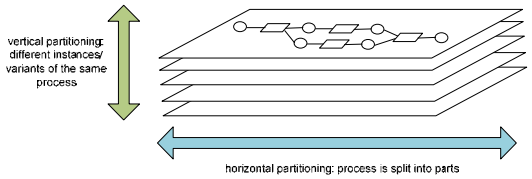
(c) Subcontracting



(d) Case transfer

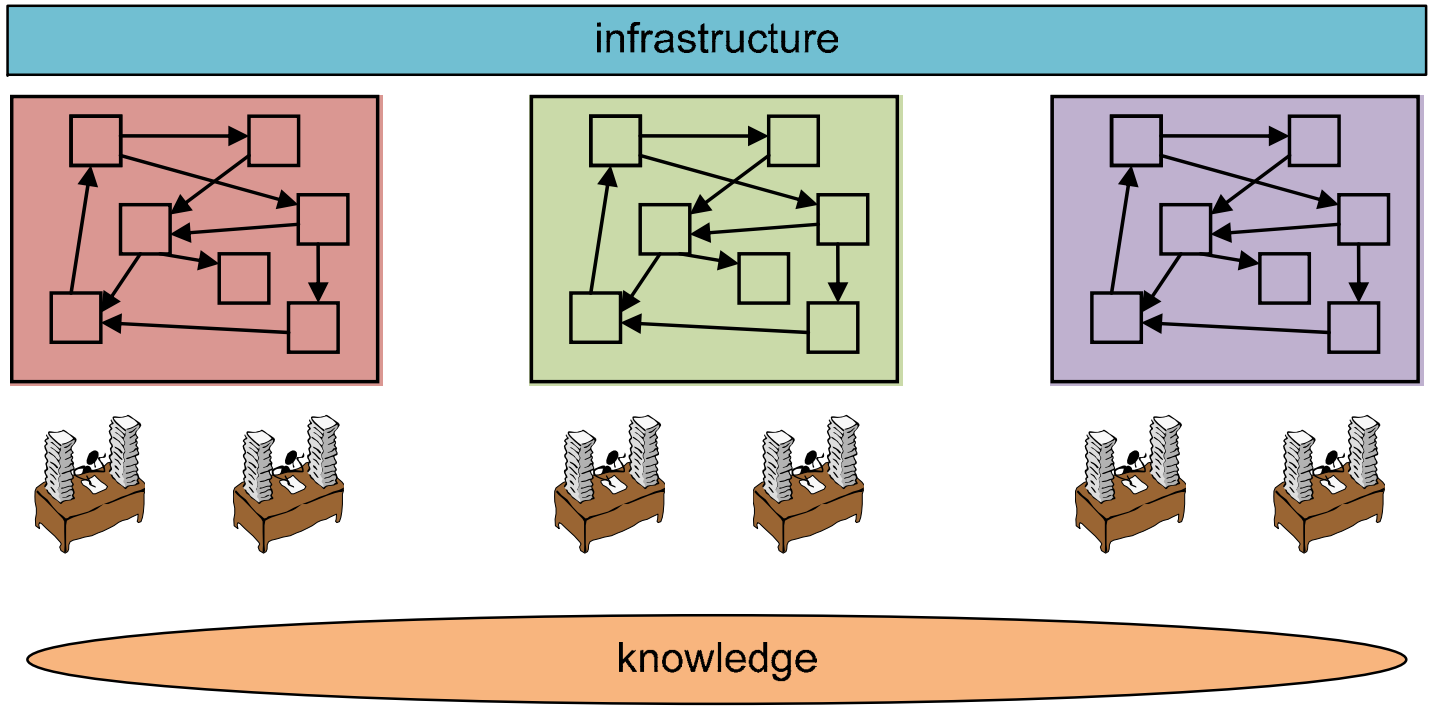


(e) Loosely coupled

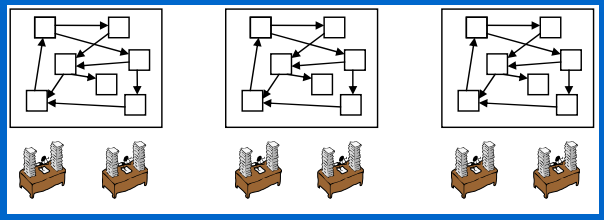


This Talk:

- **We will not focus on distributing work but on sharing knowledge and infrastructures.**



Sharing Knowledge and Infrastructures



- Consider organizations or organizational units that are executing essentially the same process(es):
  - Municipalities in the same country.
  - Large organizations having multiple brands.
  - Car rental agencies with offices in different cities/countries (Hertz, Avis, etc.).
  - Organizations using Salesforce and the like.
  - Conferences organized using Easychair.
- All share experiences, knowledge and infrastructure.
- Cloud computing will be used as an example/enabling technology. (But concepts are more generic!).

It is not just about technology ...

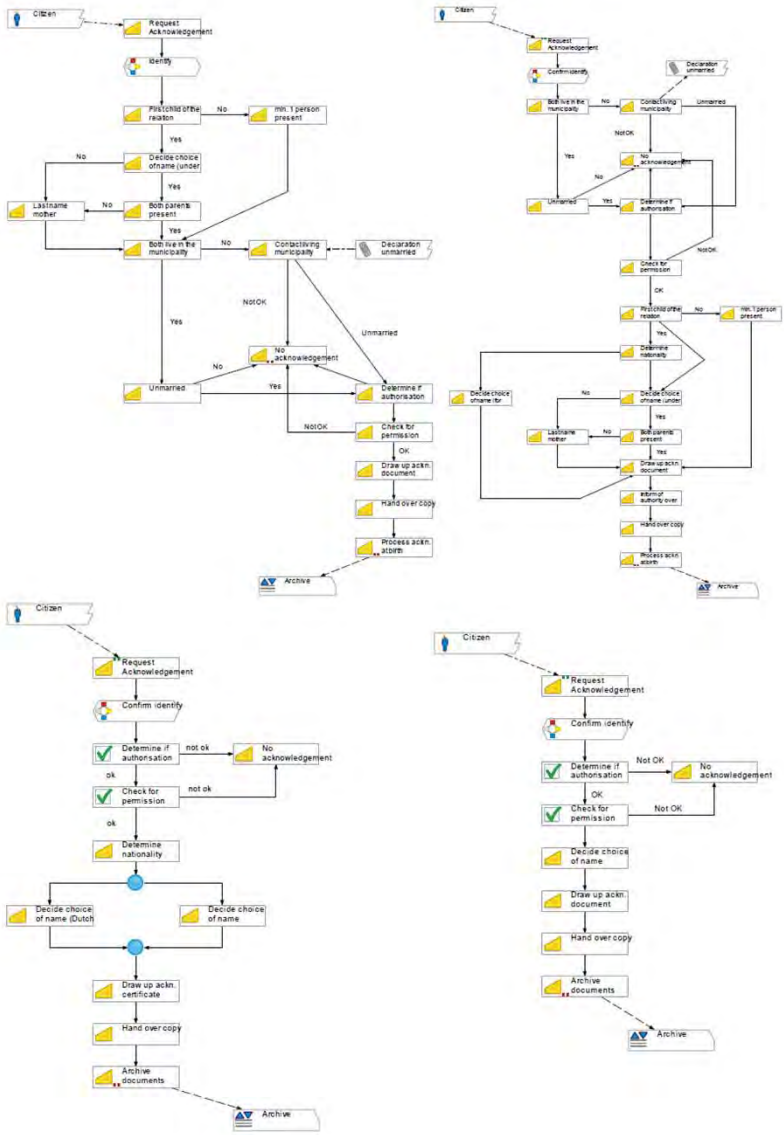

Are you afraid to look at reality?
It is also about processes!


Also applies to cloud computing!
Processes!!
Dealing with variability
Not just about technology/ infrastructure
Process variants/ configuration


New
opportunities!
Cross-organizational
process mining!!

The need for configurable process models: CoSeLoG project

+/- 430 Dutch municipalities



The need for configurable process models: Suncorp case

End to end process has between 250-1000 process steps

- 25+ steps
- 50+ steps
- 75+ steps
- 100+ steps

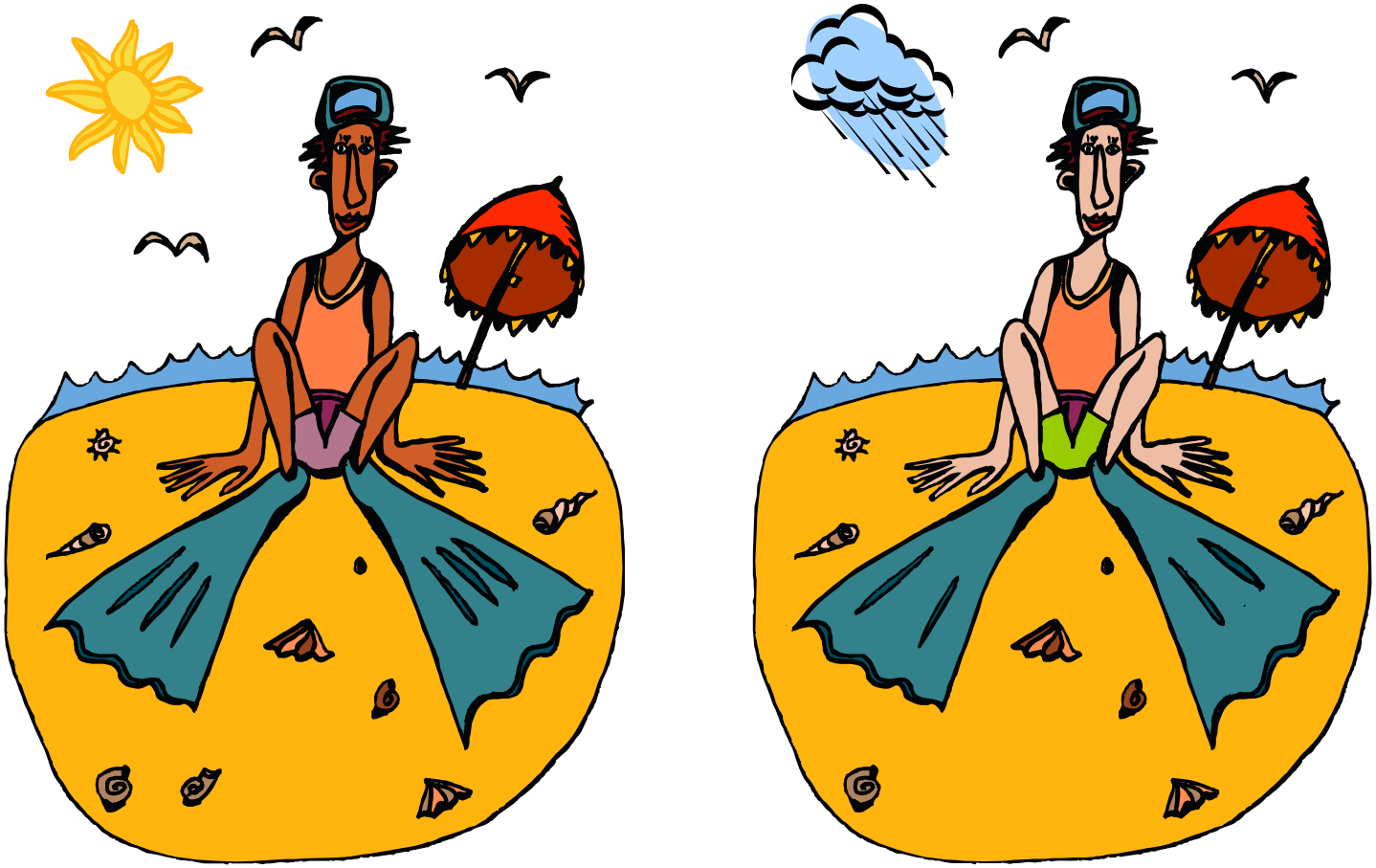
Sources: Guidewire reference models, GIO CISSS Project, CI US&S P4PI Project

	AAMI	Apia	bingle	CIL	GIO	JUST·CAR		SUNCORP	vero
Home	☑	☑			☑		☑	☑	☑
Motor	☑	☑	☑		☑	☑	☑	☑	☑
Commercial	☑			☑	☑			☑	☑
Liability	☑	☑			☑			☑	☑
CTP / WC	☑	☑			☑ ☑		☑	☑	

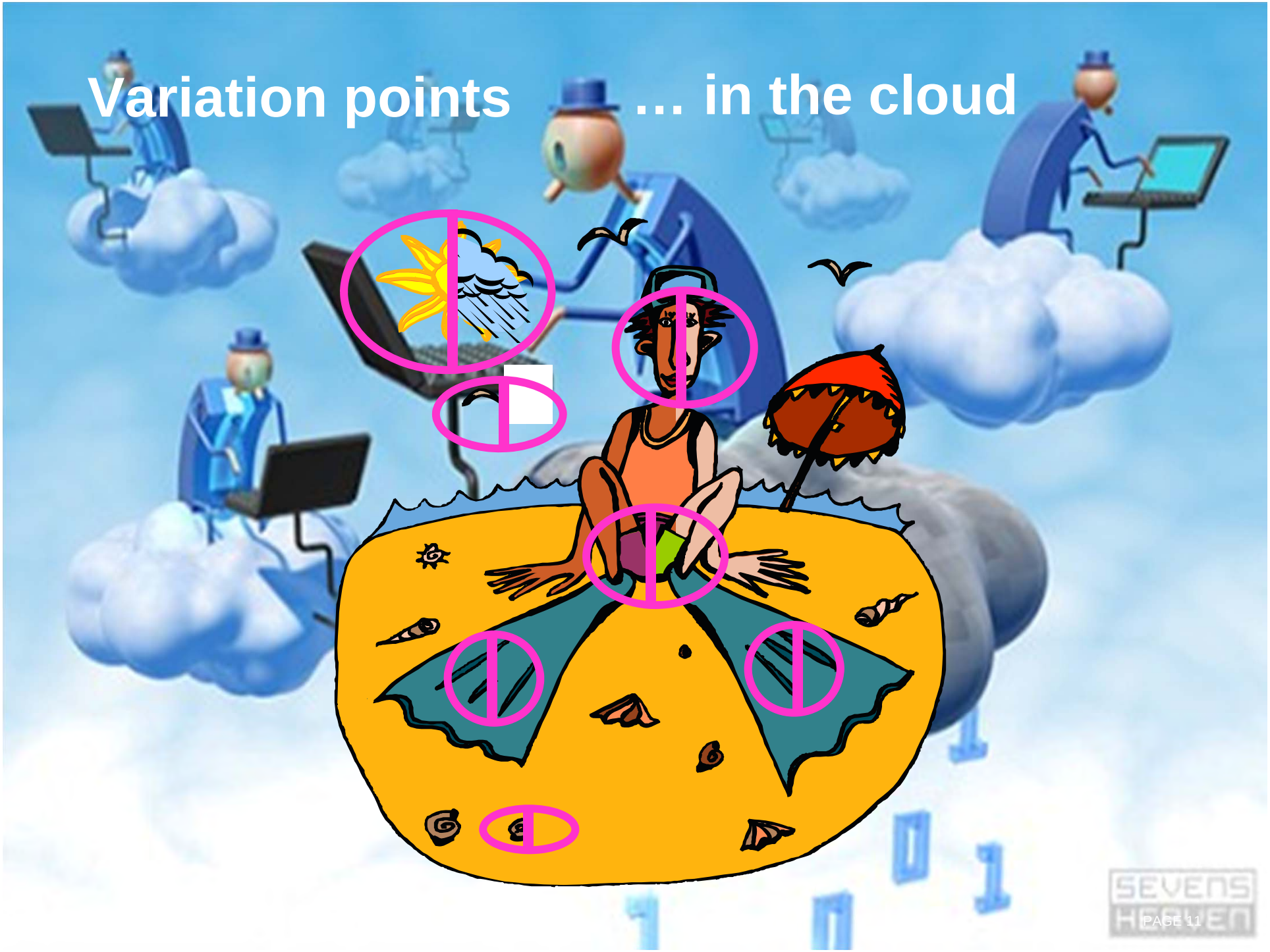
30
variations

Thanks to Marcello La Rosa

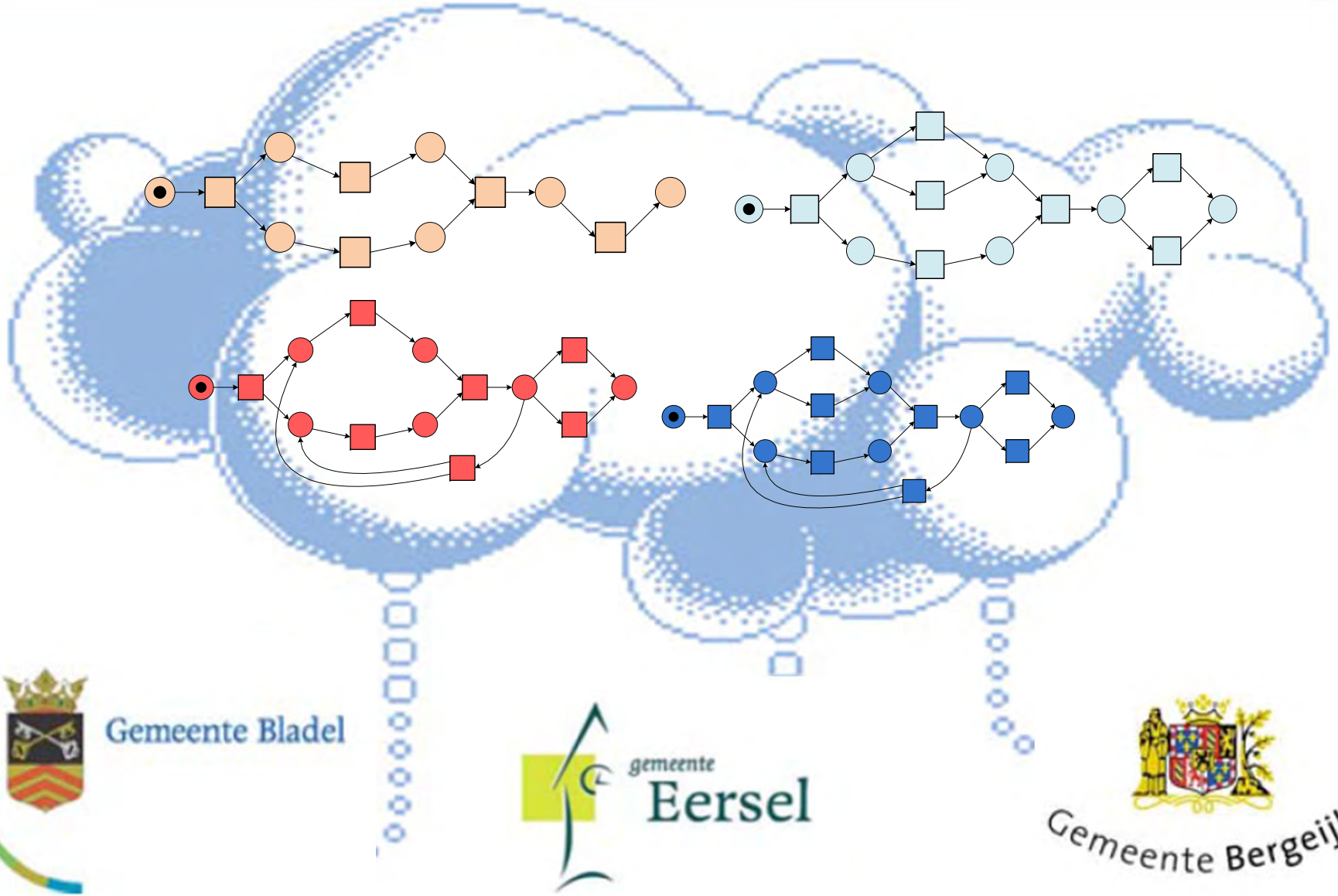
Two variants of the same process …



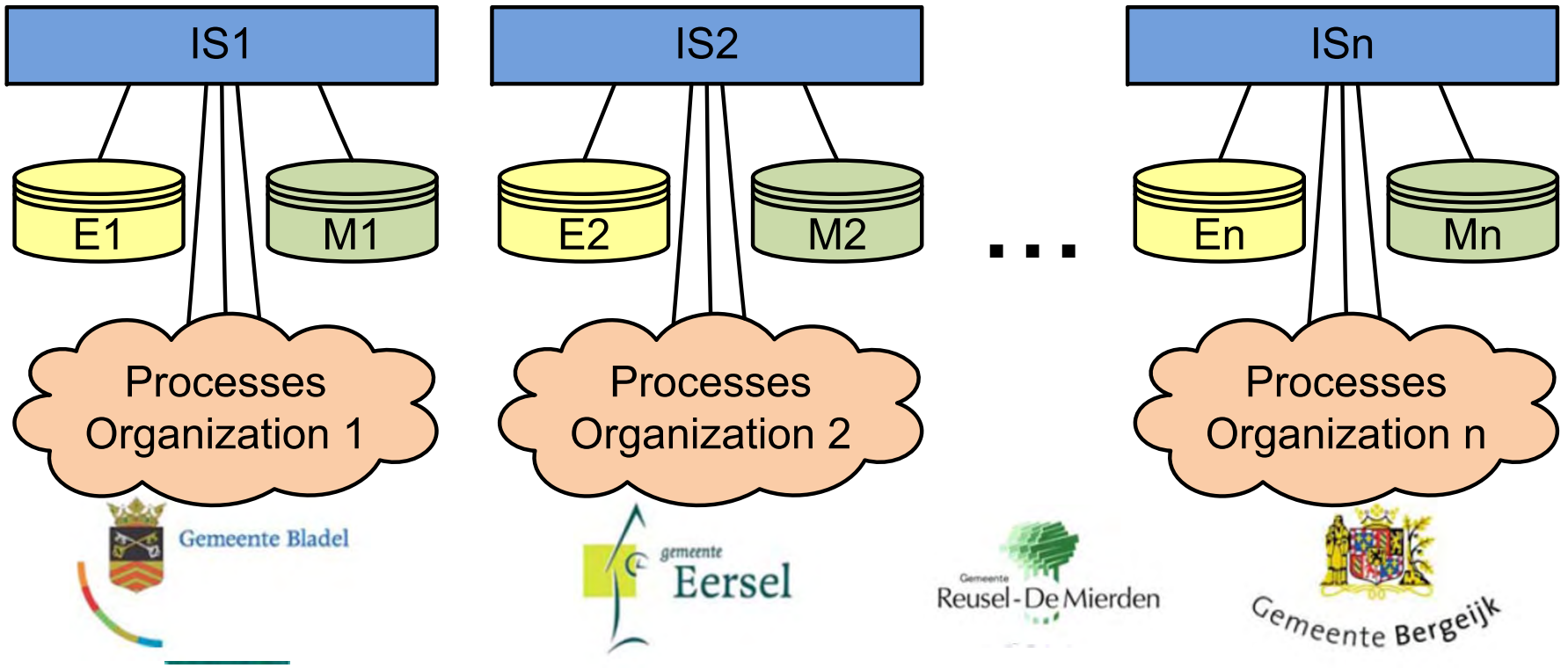
Variation points … in the cloud



Cloud computing



Traditional Situation

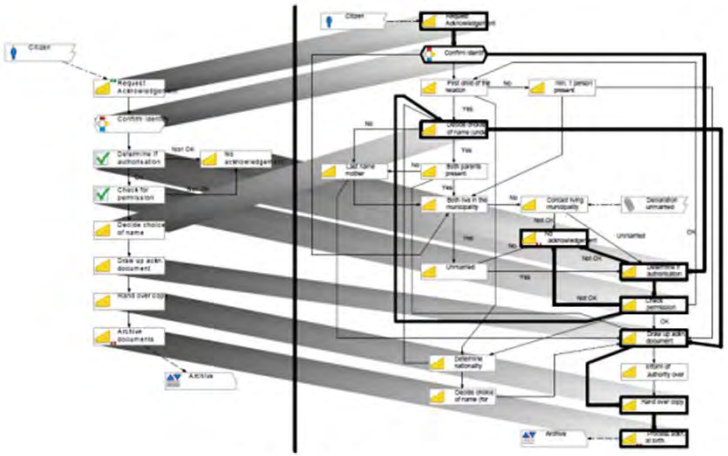


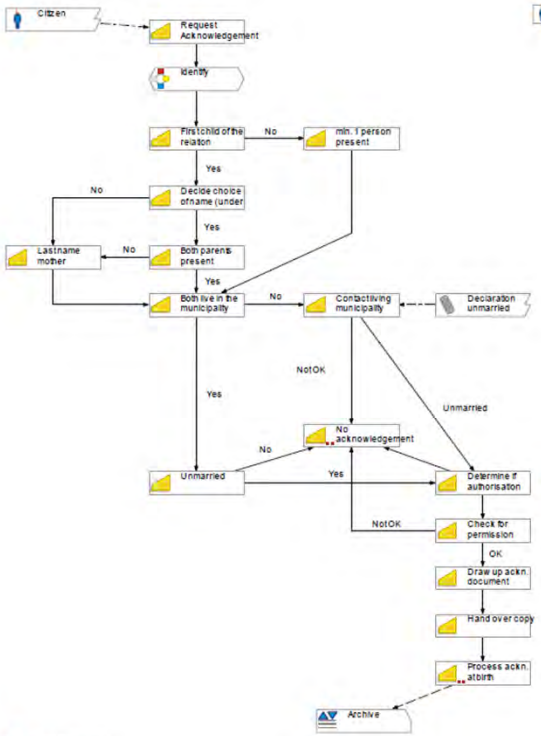
IS = Information System
E = Event log
M = Models

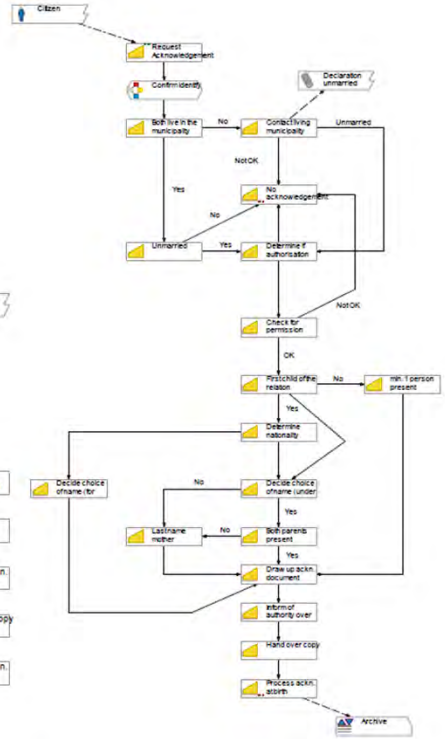
Example

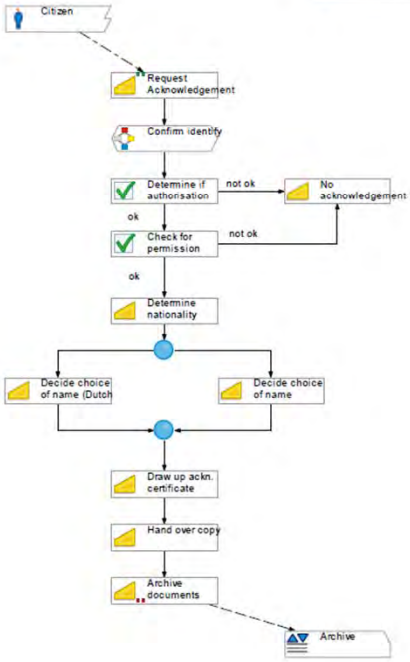
Acknowledgement of an Unborn Child

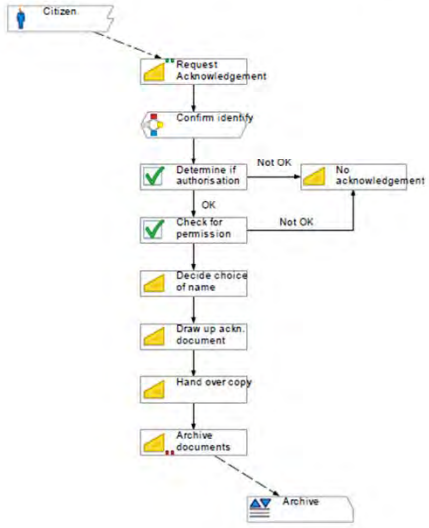
- Same but different …
- “Couleur Locale”
- Different from NVVB models.
- Configurable process models!



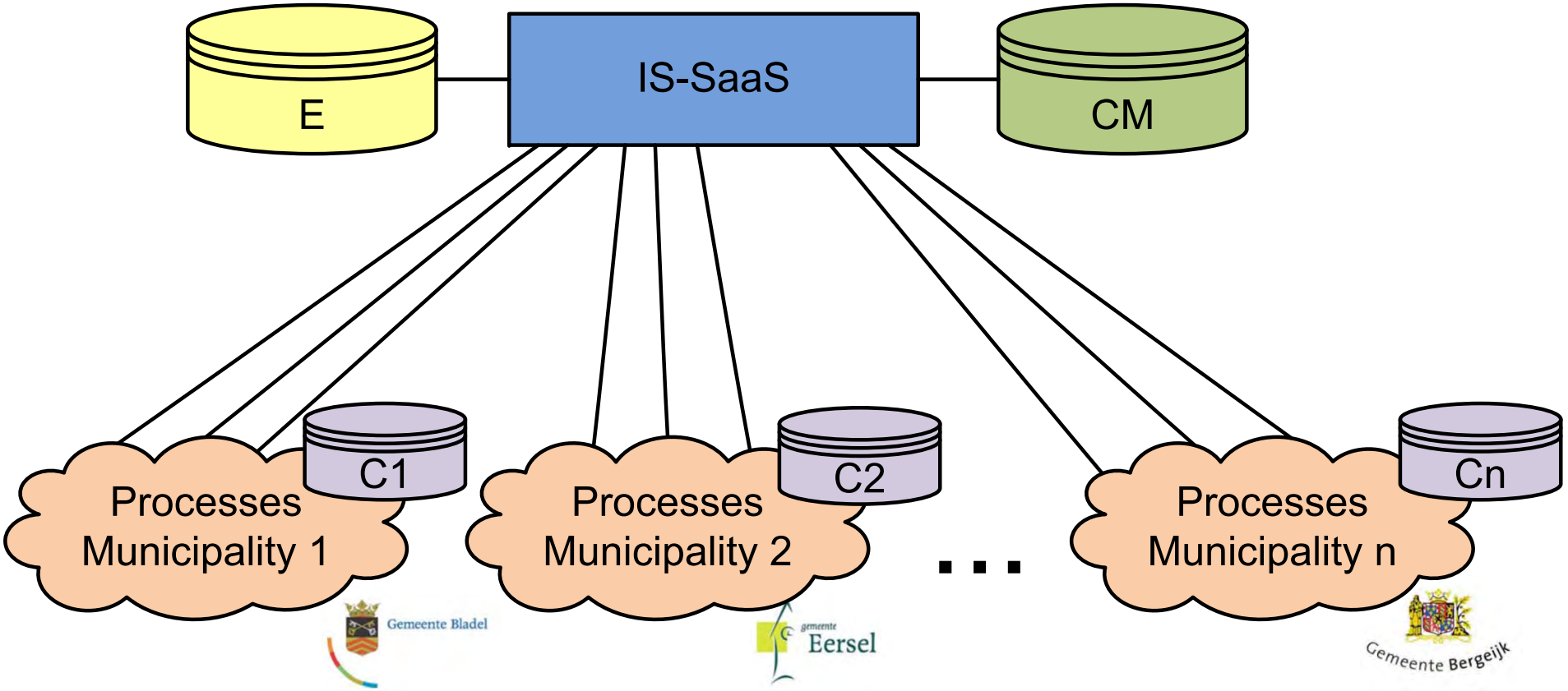






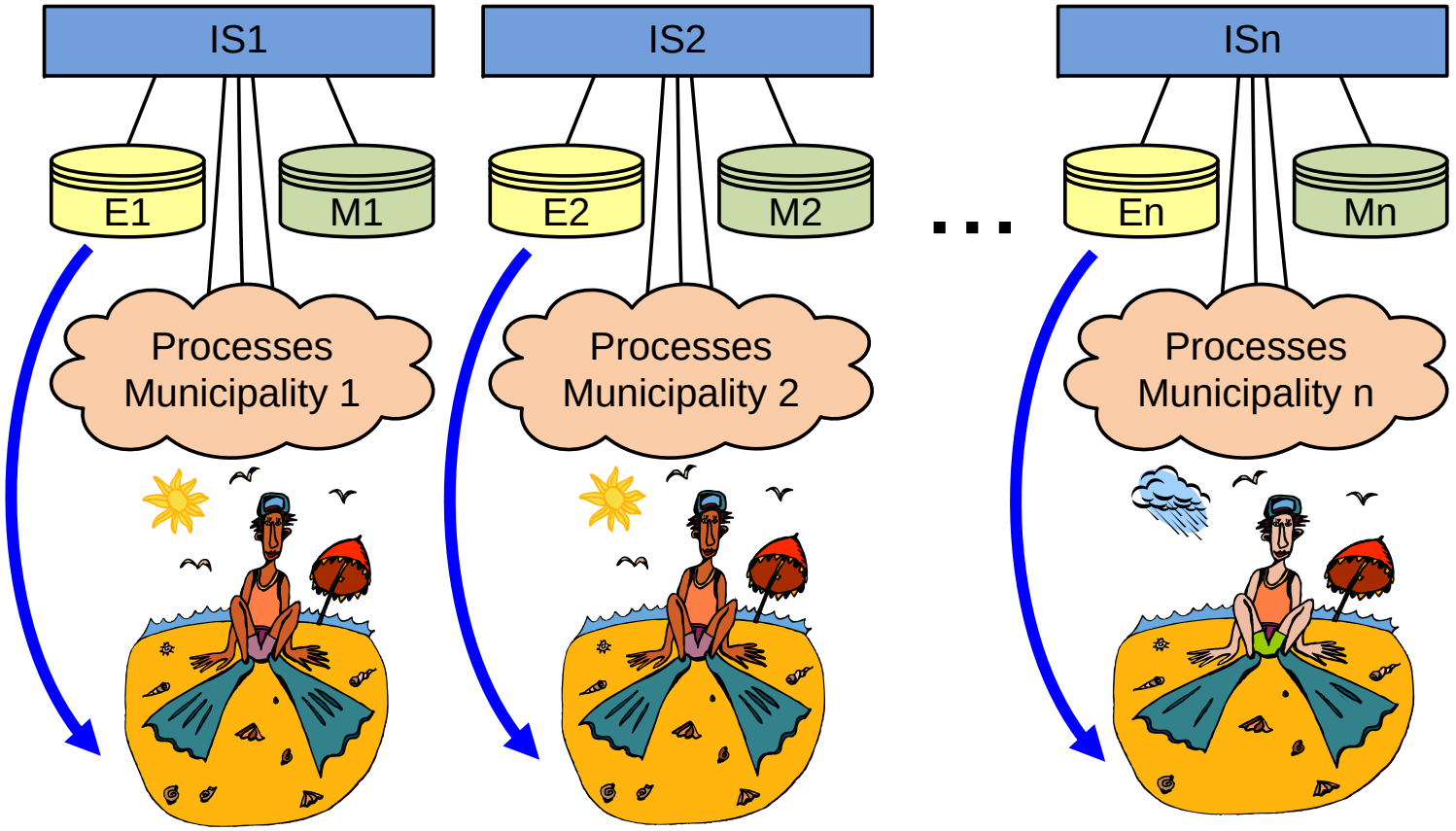


Using SaaS Technology

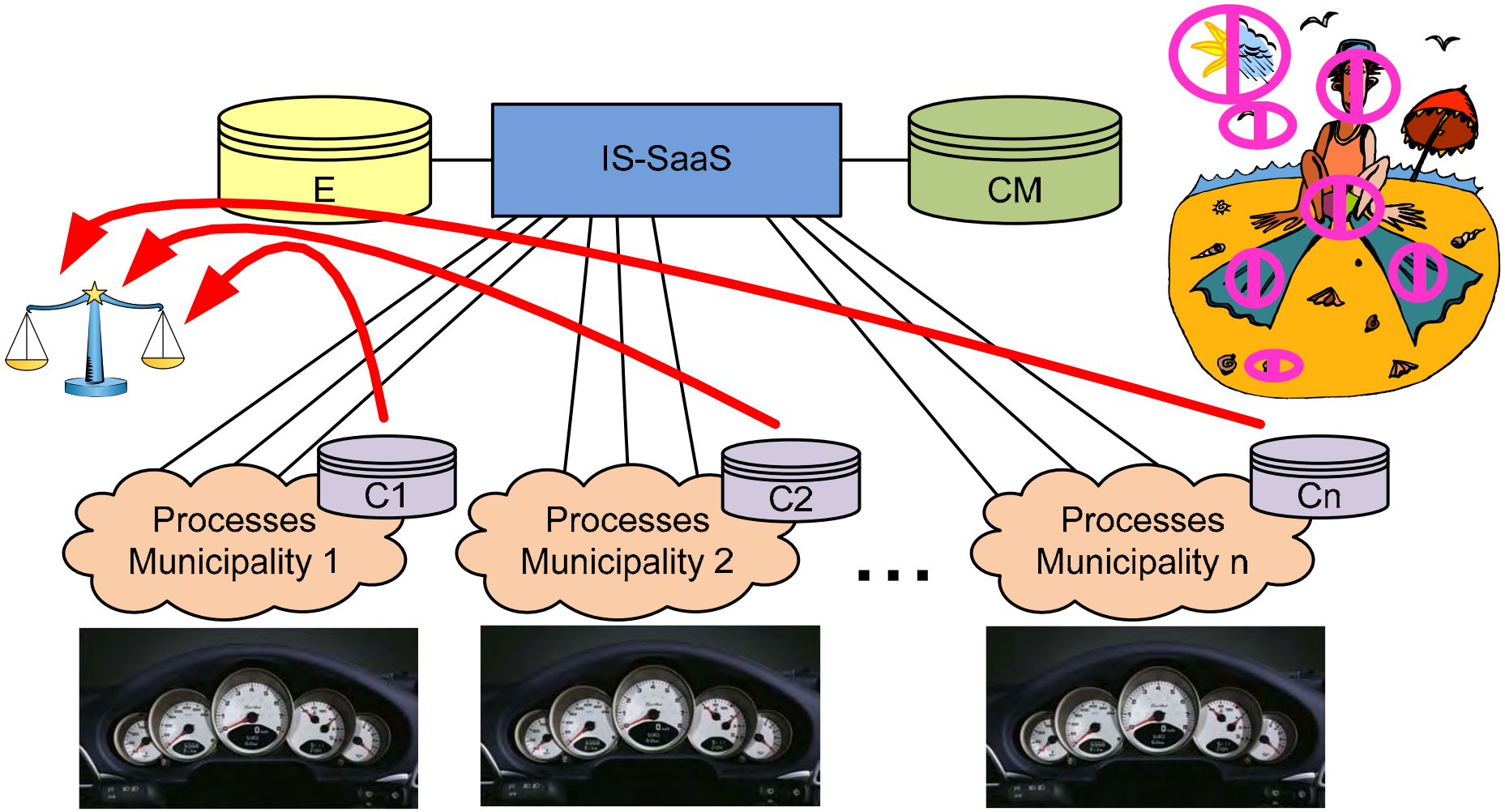


IS-SaaS = Information System (using a SaaS-based BPMS)
E = Event log
CM = Configurable Models
C = Configuration

Process Mining: Before



Process Mining: After



cross-organizational process mining


Configuration

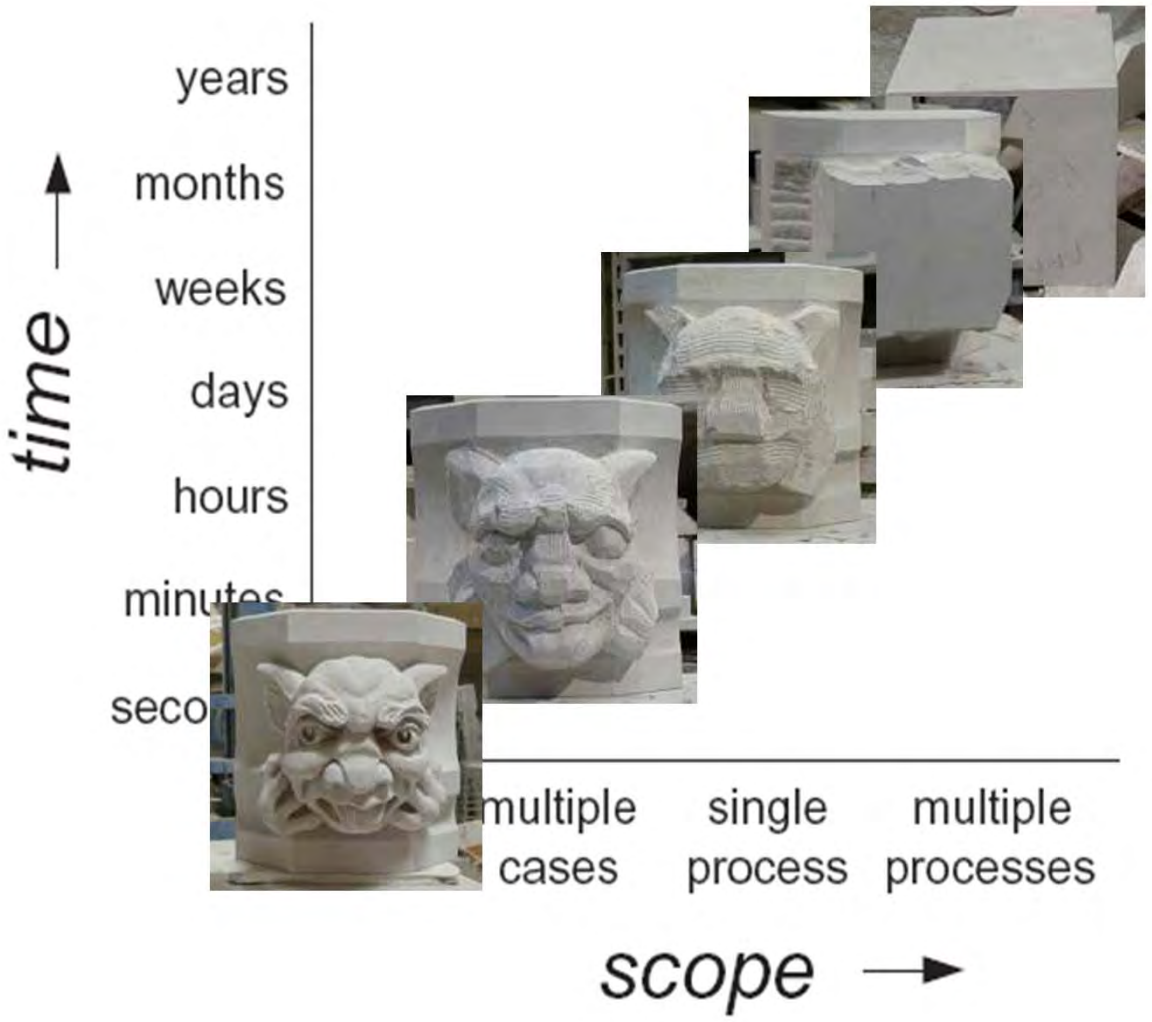
Positioning of Configuration

Some quotes from Michelangelo

- **“Every block of stone has a statue inside it and it is the task of the sculptor to discover it.”**
- **“I saw the angel in the marble and carved until I set him free.”**
- **“Carving is easy, you just go down to the skin and stop.”**

Michelangelo's David

Life is about making choices …



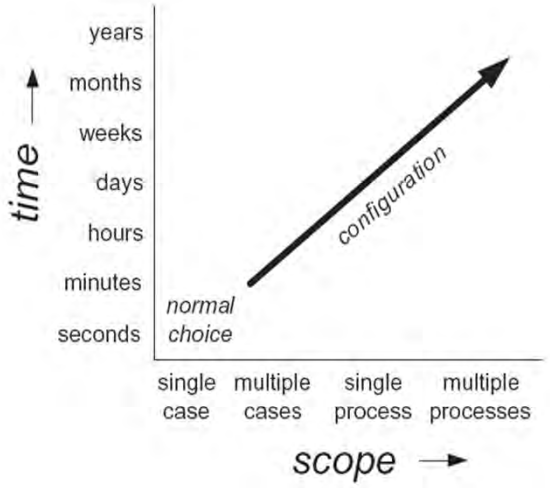
Time and artifacts

- Design time (generic model, i.e., is not released for instantiation)
- Configuration time (specific model, i.e., can be instantiated)
- Instantiation time (specific model + instance)
- Run time (specific model + instance + state/partial trace)
- Auditing time (specific model + instance + full trace)

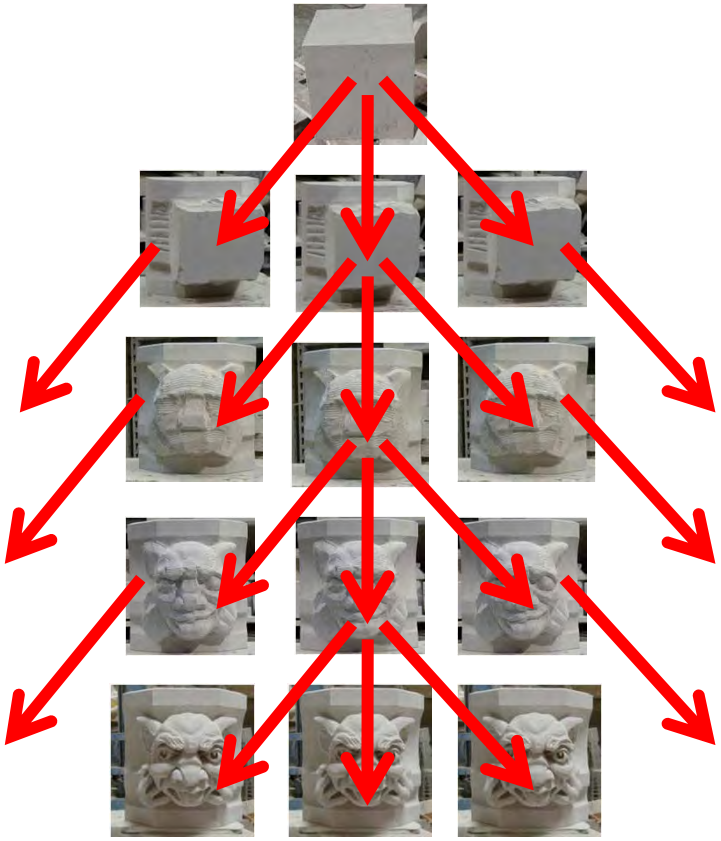




Continuum



- **In The Netherlands, …**
- **In Brisbane, …**
- **When the sun shines, …**
- **On Sunday, …**
- **When very busy, …**
- **For these customers, …**
- **…**



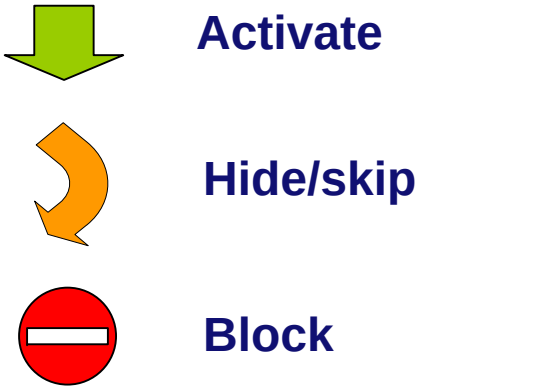
Branching structure

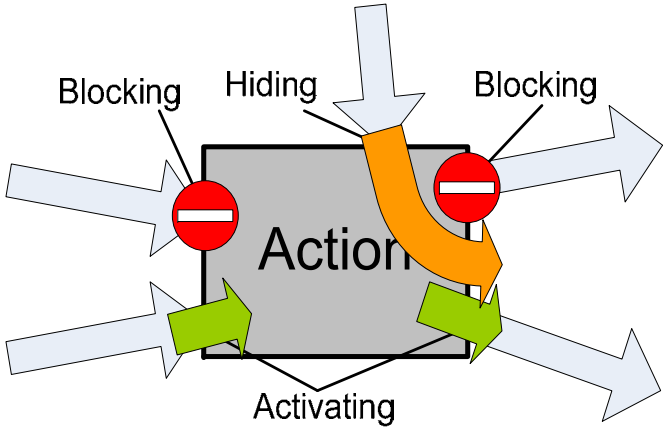
Hiding and blocking

Configuration = limiting behavior !









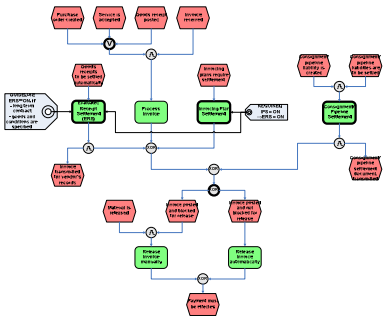
Configurable Process Models

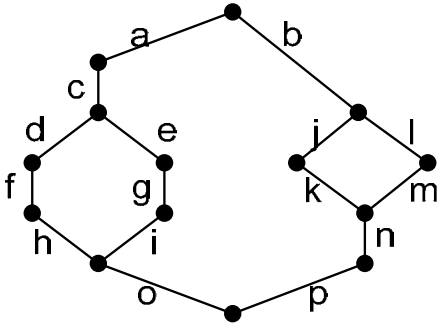
C-EPC

C-Petri Net

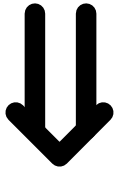
C-YAWL

C-BPEL

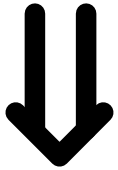




C-LTS



Configuration



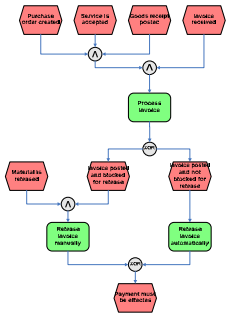
Blocking **Hiding**

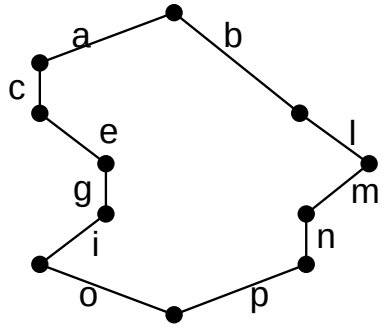
EPC

Petri Net

YAWL

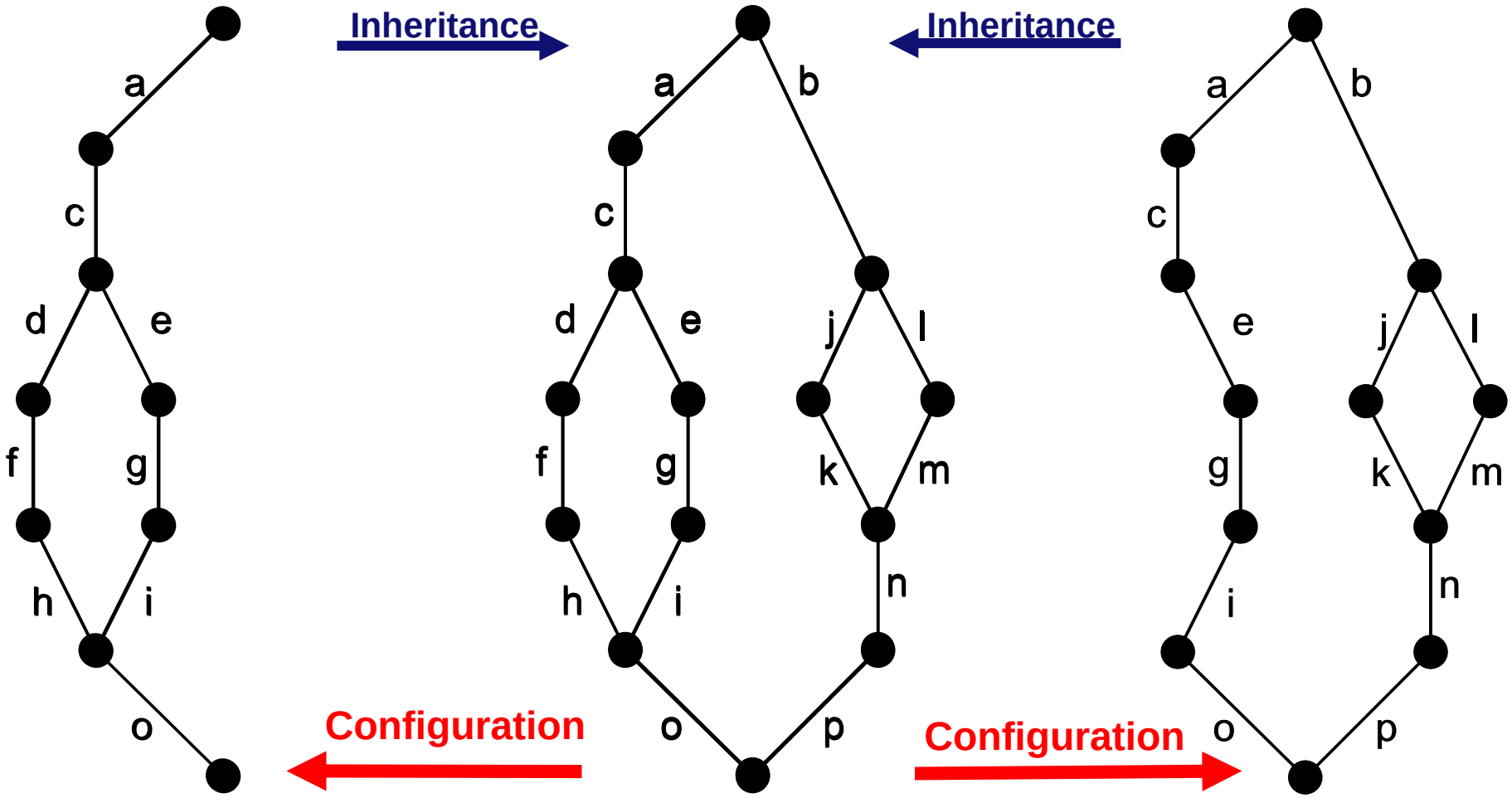
BPEL





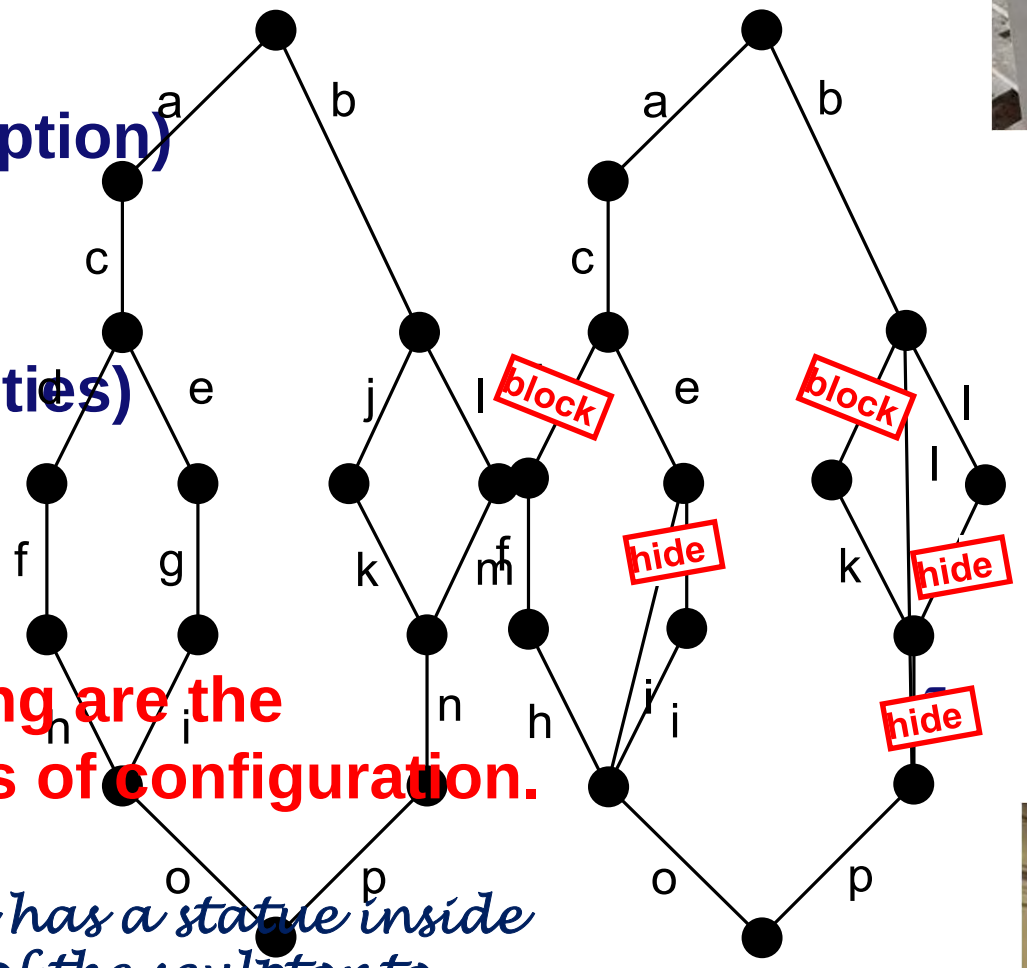
LTS

Inheritance of dynamic behavior



Configuration Techniques

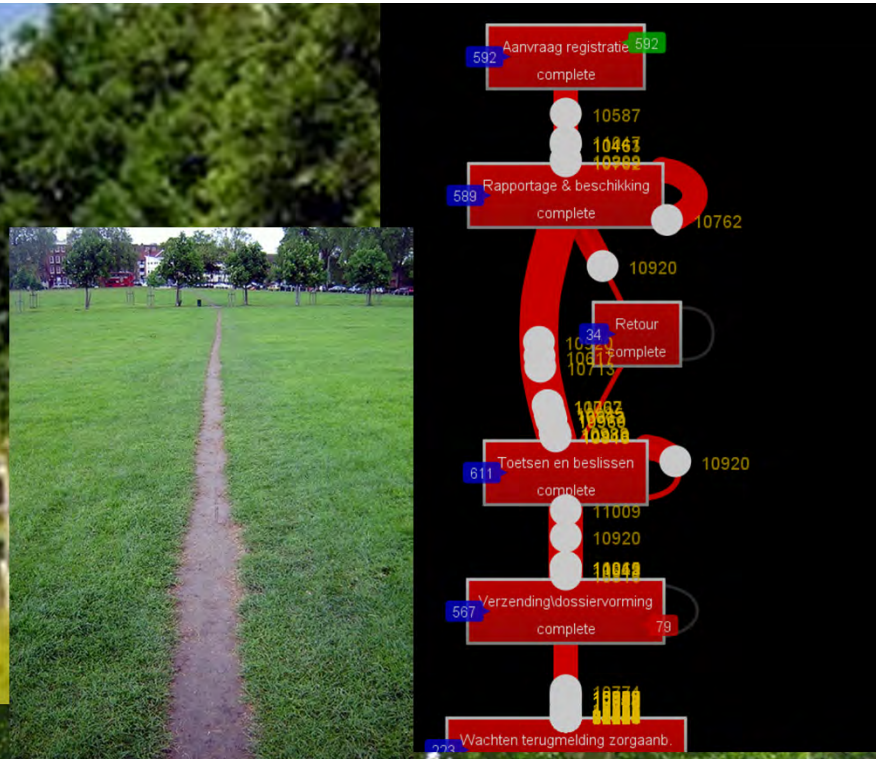
- **Blocking (removing an option)**
- **Hiding (skipping activities)**



Blocking and hiding are the essential concepts of configuration.

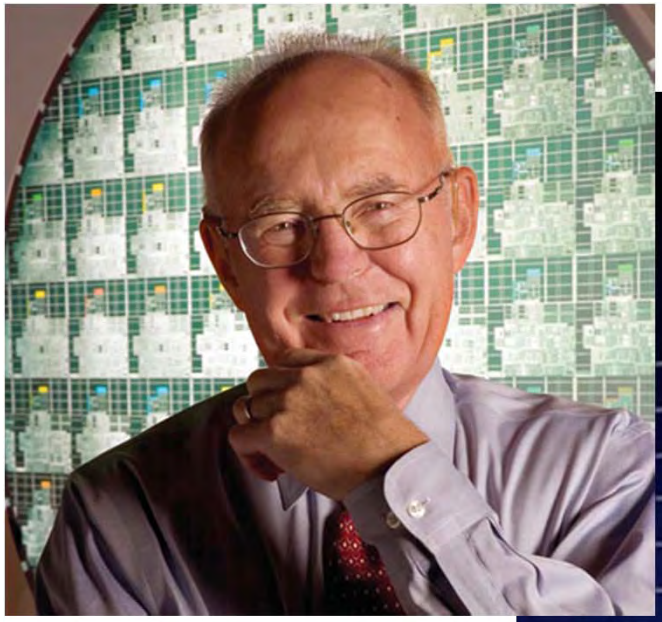
"Every block of stone has a statue inside it and it is the task of the sculptor to discover it."

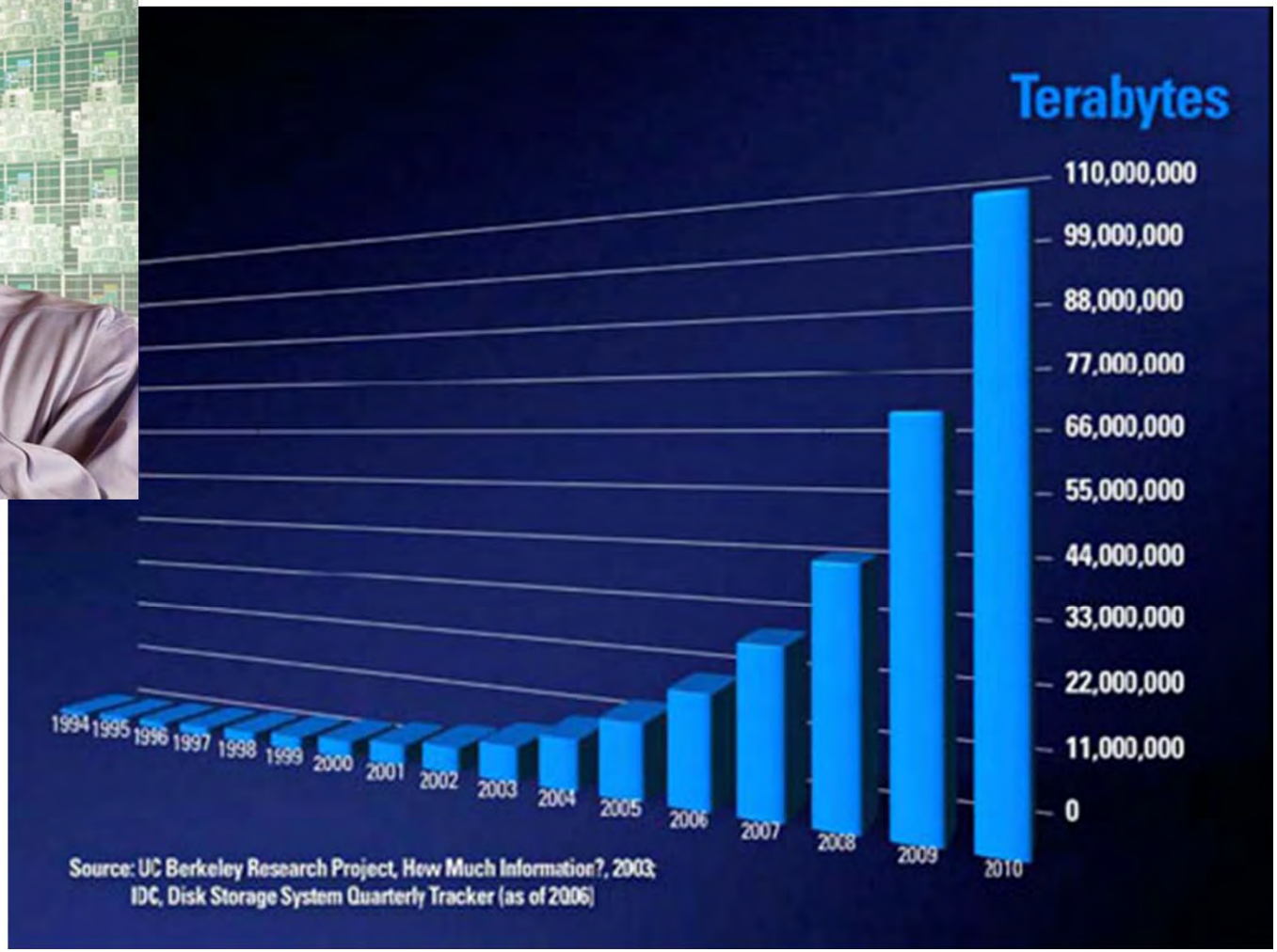
Process mining



Desire lines in process models

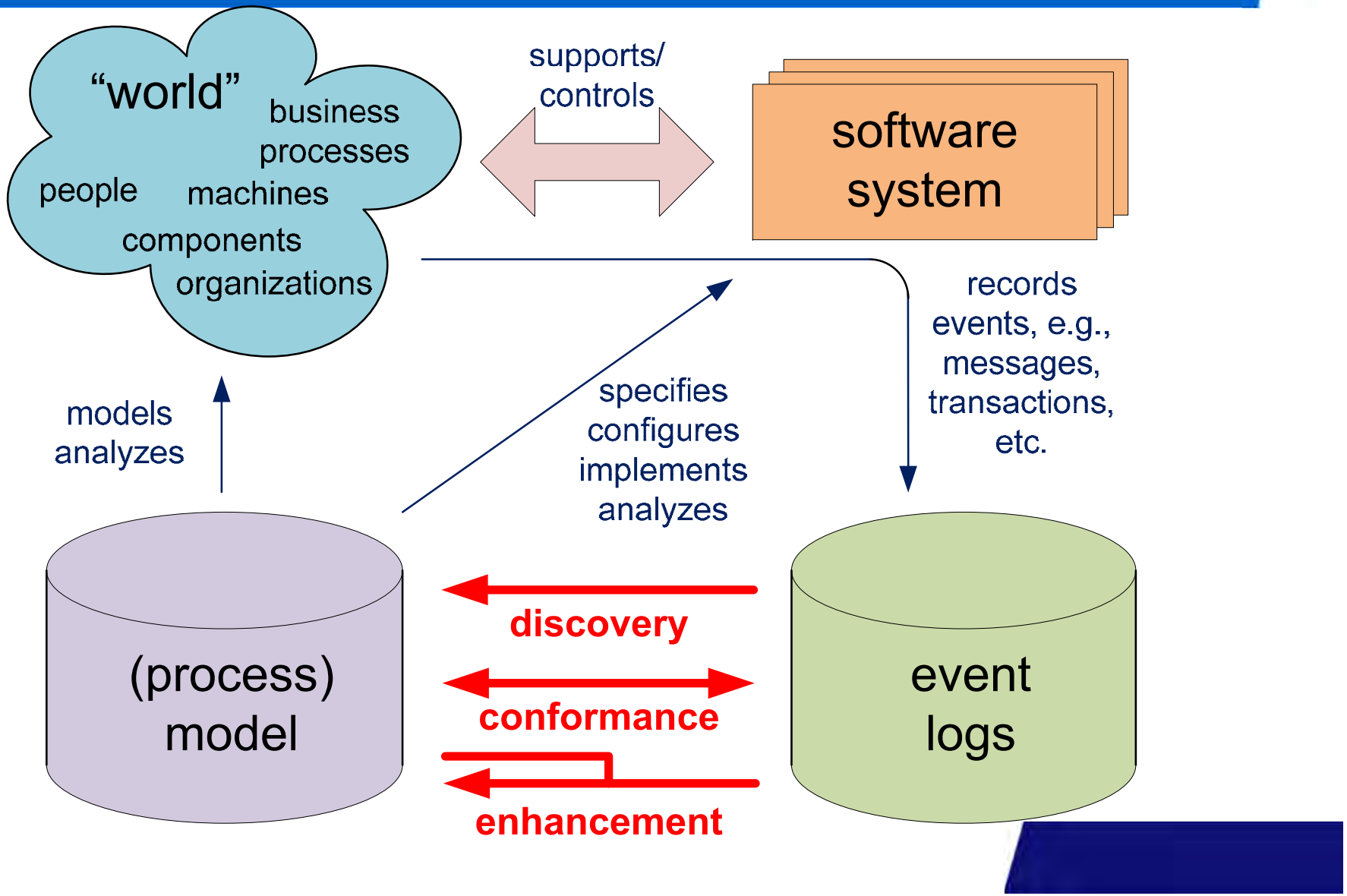
Data explosion






Process Mining =
Event Data + Processes
Data Mining + Process Analysis
Machine Learning + Formal Methods

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

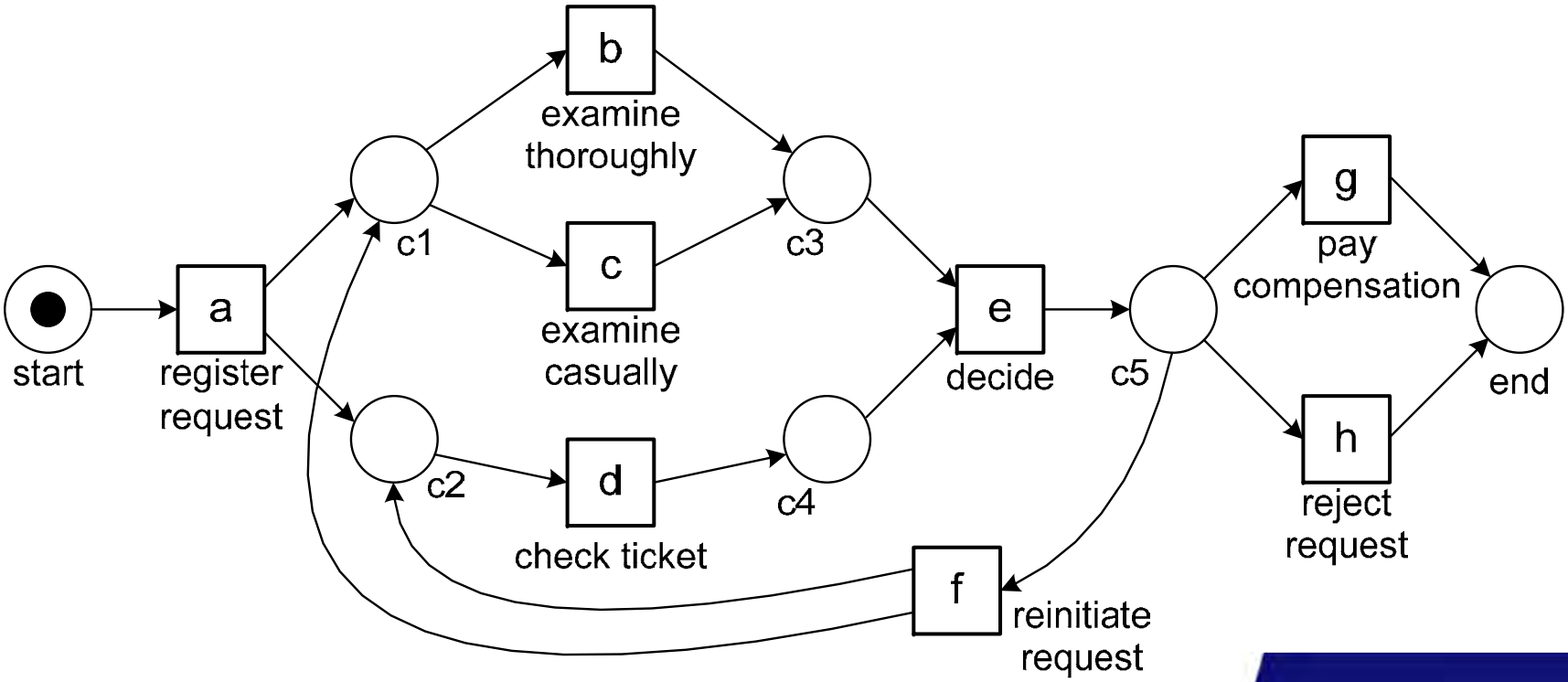
case id	event id	properties				
		timestamp	activity	resource		
1	35654423	30-12-2010:11.02	register request	Pete		
	35654424	31-12-2010:10.06	examine thoroughly	Sue		
	35654425	05-01-2011:15.12	check ticket	Mike		
	35654426	06-01-2011:11.18	decide	Sara		
	35654427	07-01-2011:14.24	reject request	Pete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

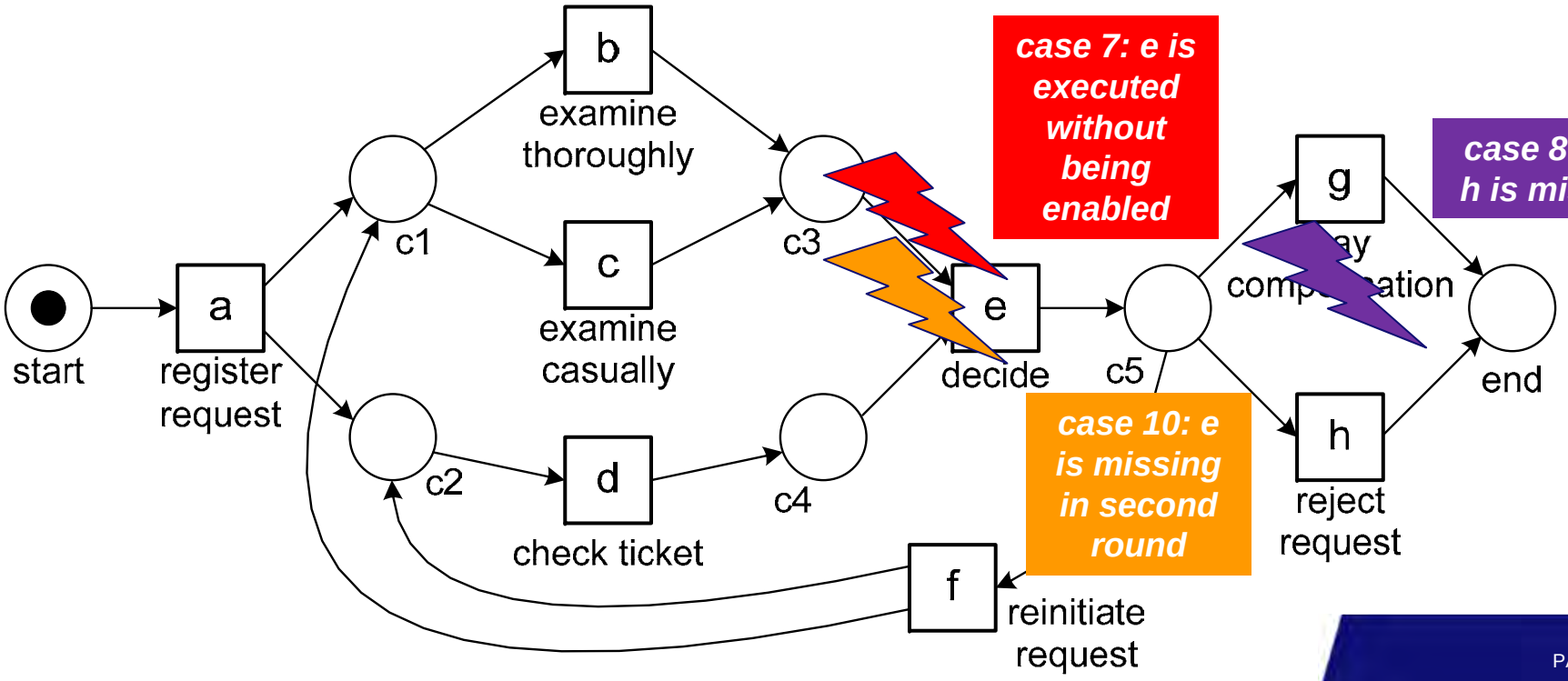
Process discovery

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

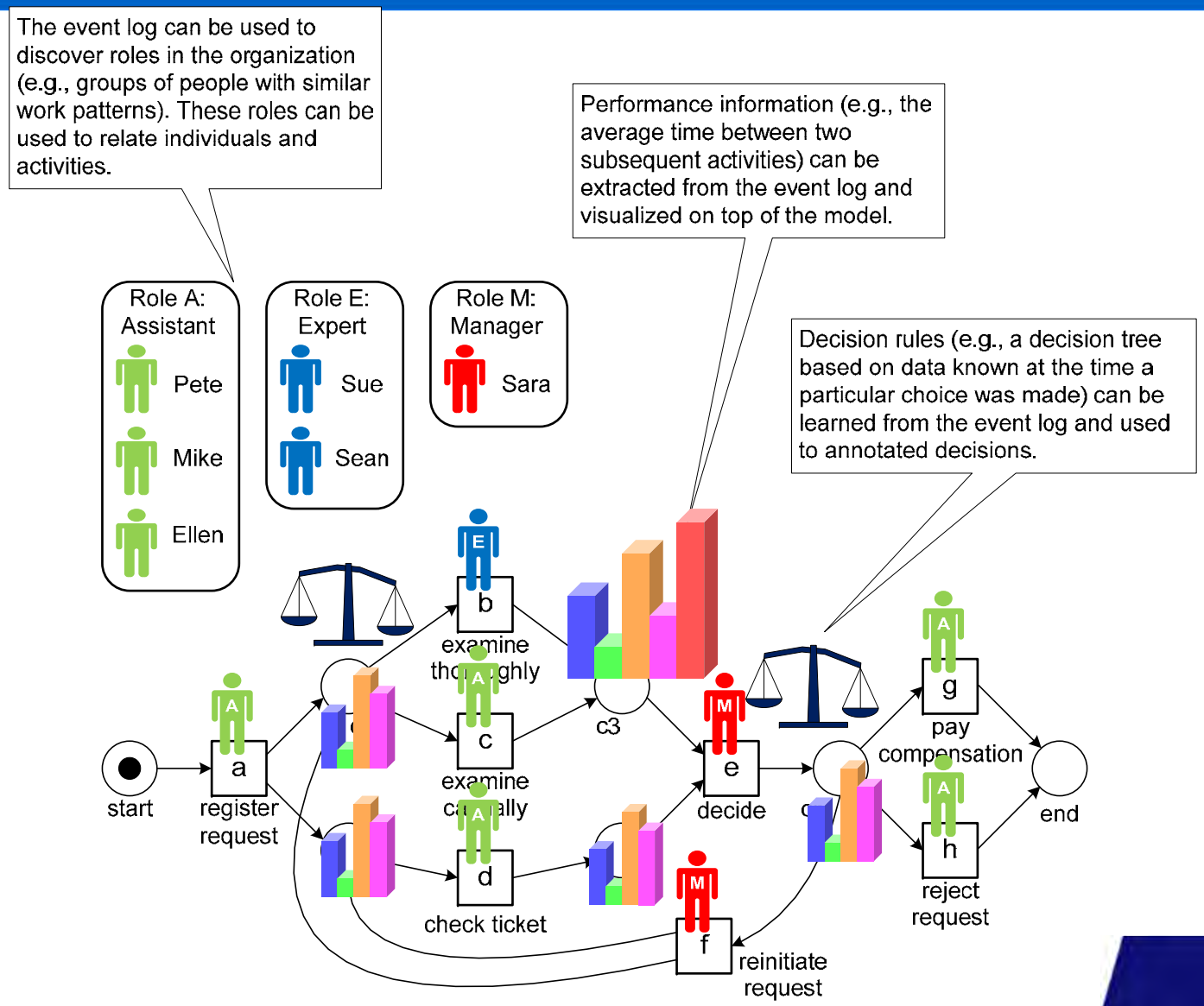


Conformance checking

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 …

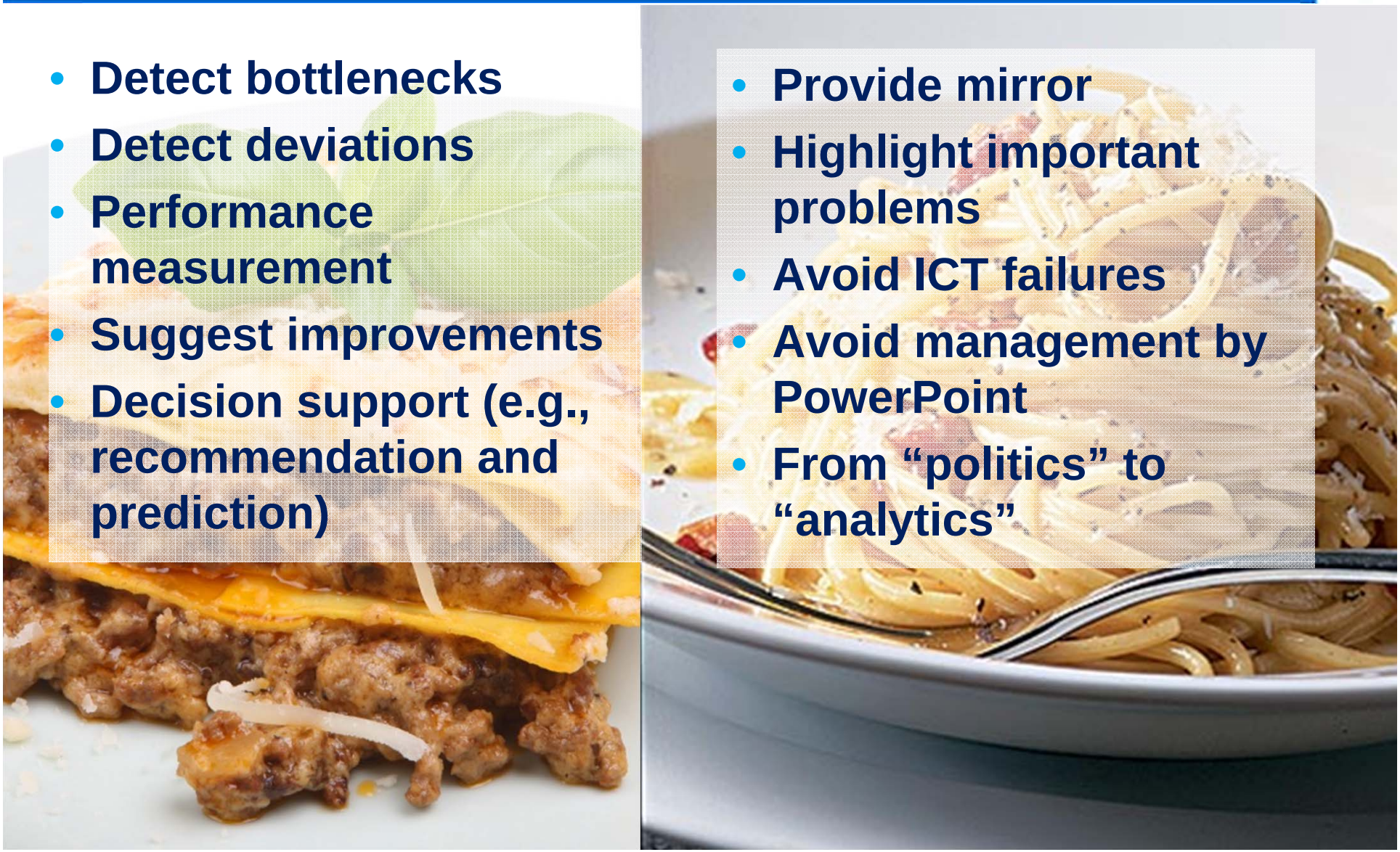
- Open-source (L-GPL), cf. www.processmining.org
- Plug-in architecture
- Plug-ins cover the whole process mining spectrum and also support classical forms of process analysis

Towards Maturity …

- **IEEE Task Force on Process Mining**
  - **Software vendors** (Pallas Athena, IDS Scheer/Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas, Fujitsu, Business Process Mining)
  - **Consultancy** (Some of the above and ProcessGold, Business Process Trends, Gartner, Deloitte, Rabobank)
  - **Universities** (TU/e, University of Padua, University of Catalunya, New Mexico State University, Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität, Queensland University of Technology, Vienna University of Economics and Business, Stevens Institute of Technology, University of Haifa, Seoul National University of Technology, Cranfield University, K.U.Leuven, Tsinghua University, Innsbruck University)
- **Various tools:** ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (Stereo-LOGIC), Flow (Fourspark), Futura Reflect (Futura Process Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/ Verint), ProcessAnalyzer (QPR), ProM (TU/e), Rbminer/Dbminer (UPC), and Reflect|one (Pallas Athena).

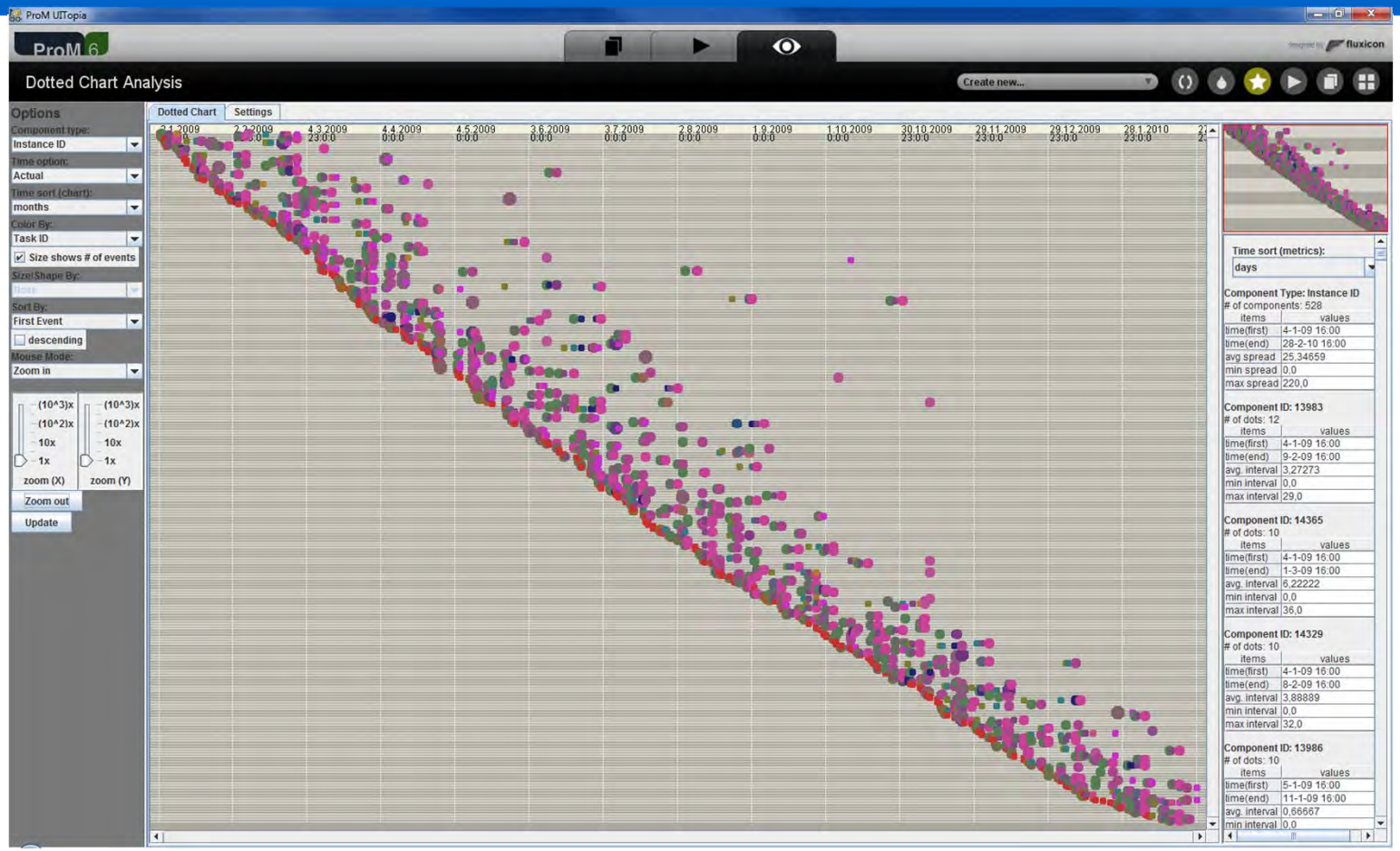
How can process mining help?

- Detect bottlenecks
- Detect deviations
- Performance measurement
- Suggest improvements
- Decision support (e.g., recommendation and prediction)



- Provide mirror
- Highlight important problems
- Avoid ICT failures
- Avoid management by PowerPoint
- From “politics” to “analytics”

Example of a Lasagna process: WMO process of a Dutch municipality

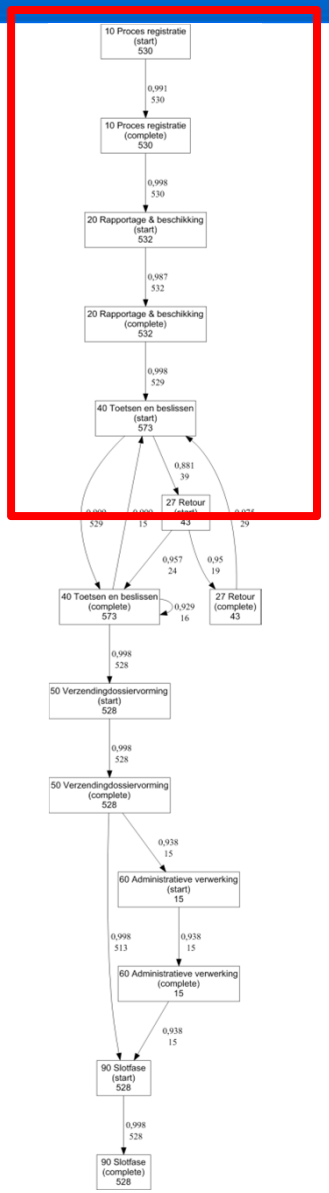


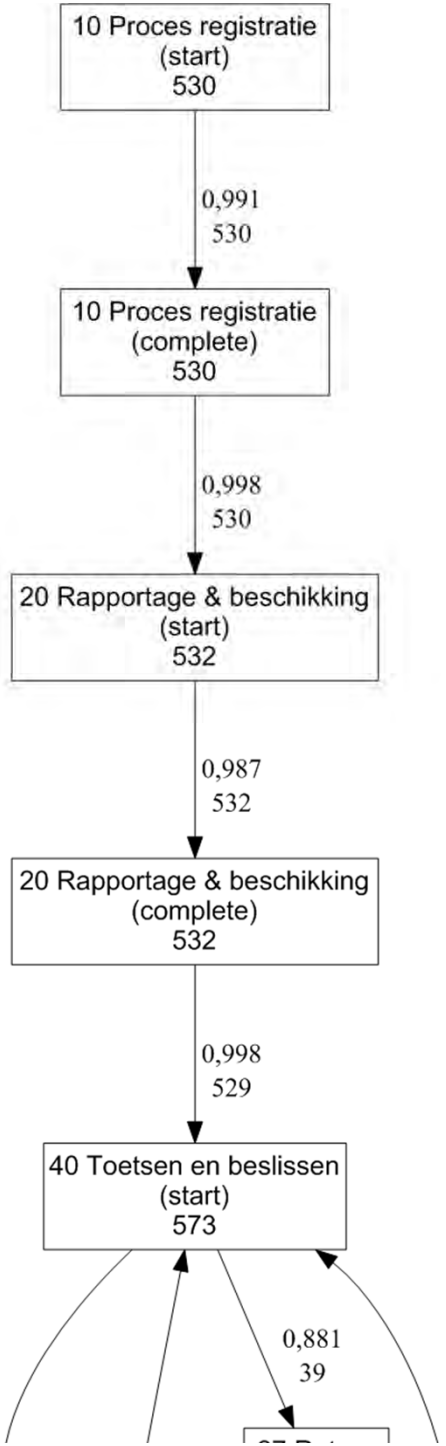
Each line corresponds to one of the 528 requests that were handled in the period from 4-1-2009 until 28-2-2010. In total there are 5498 events represented as dots. The mean time needed to handled a case is approximately 25 days.

WMO process
(Wet Maatschappelijke Ondersteuning)

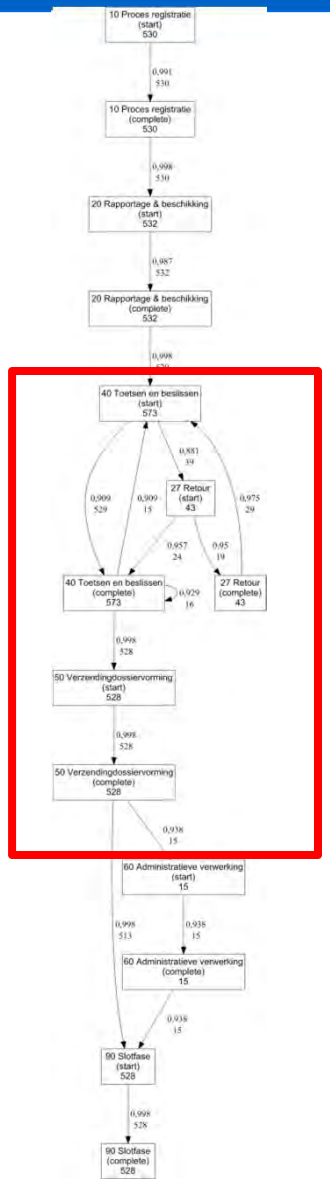
- WMO refers to the social support act that came into force in The Netherlands on January 1st, 2007.
- The aim of this act is to assist people with disabilities and impairments. Under the act, local authorities are required to give support to those who need it, e.g., household help, providing wheelchairs and scootmobiles, and adaptations to homes.
- There are different processes for the different kinds of help. We focus on the process for handling requests for household help.
- In a period of about one year, 528 requests for household WMO support were received.
- These 528 requests generated 5498 events.

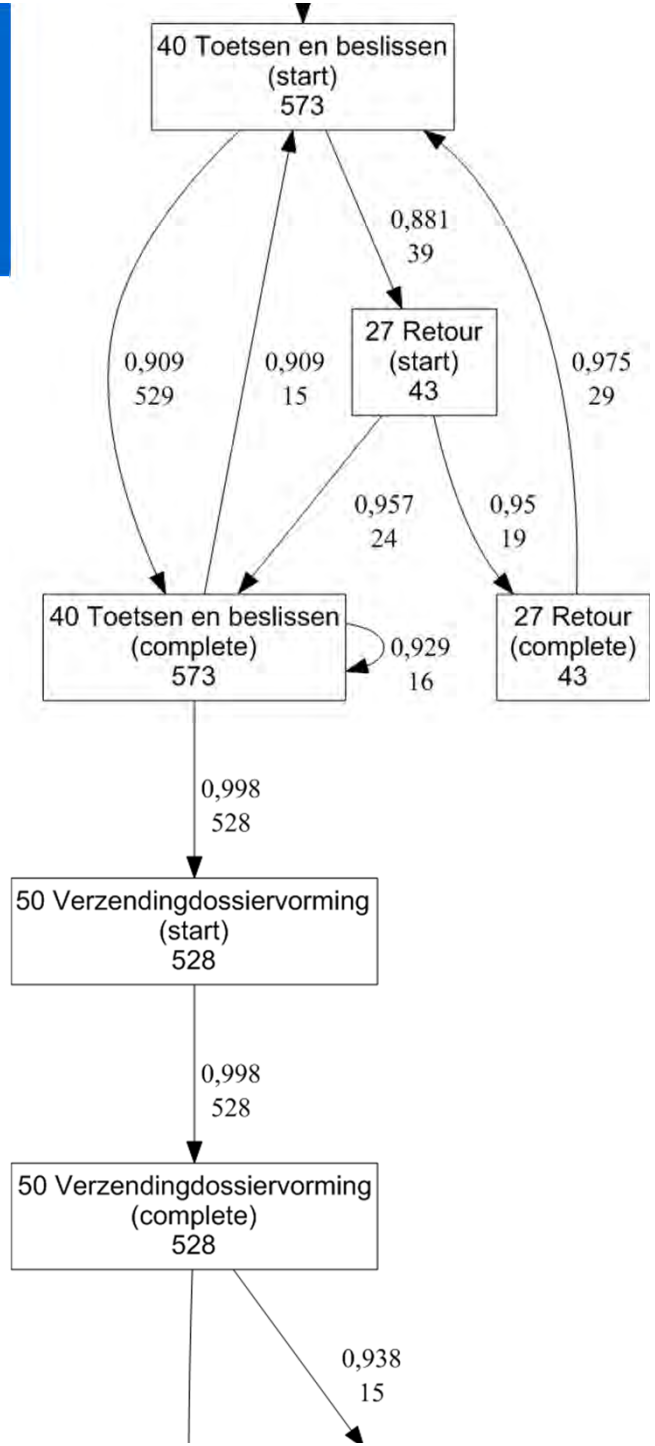
C-net discovered using heuristic miner (1/3)



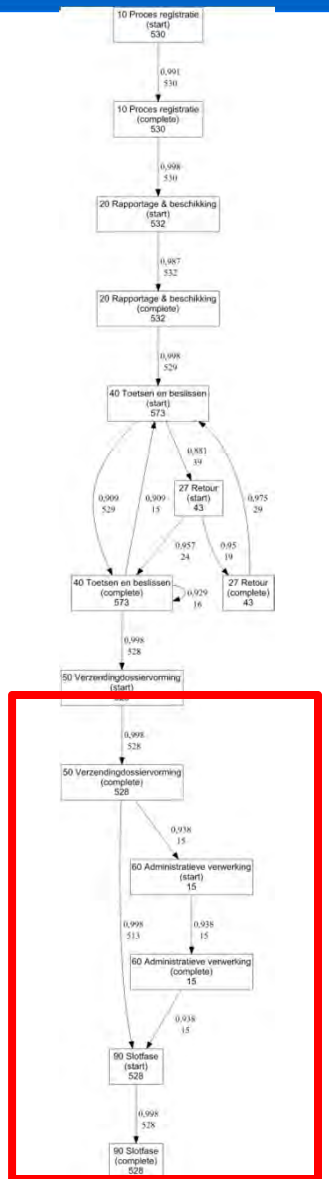


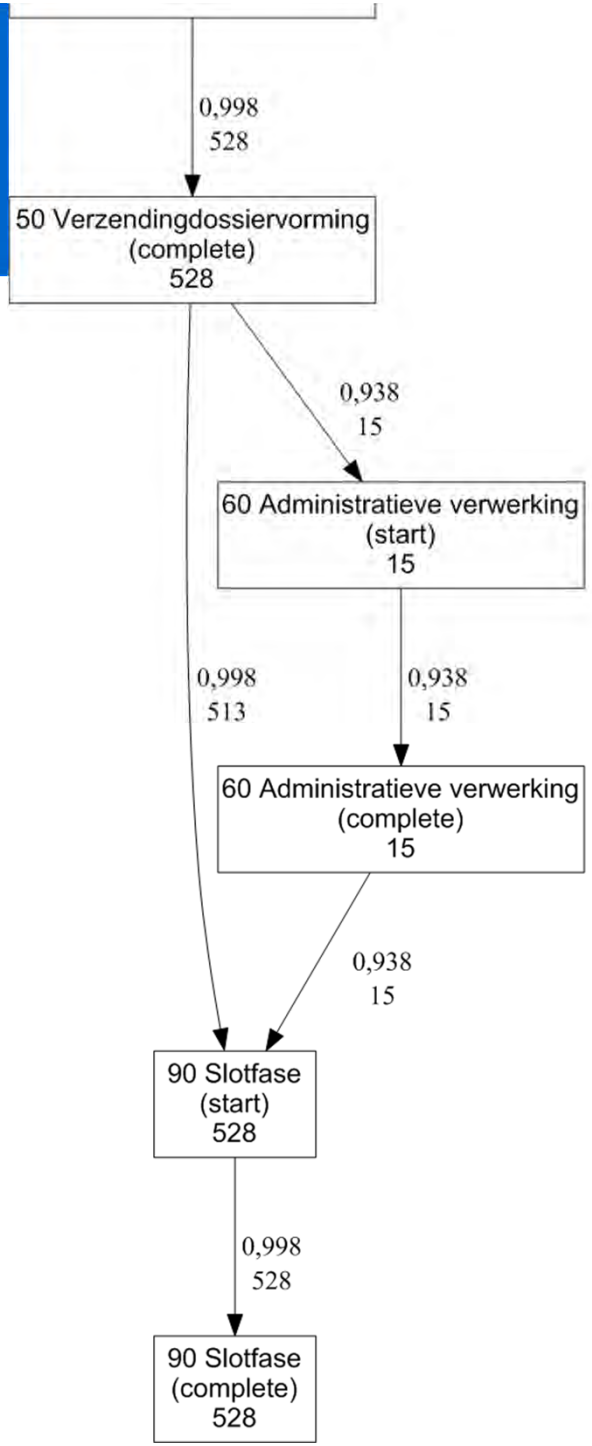
C-net discovered using heuristic miner (2/3)



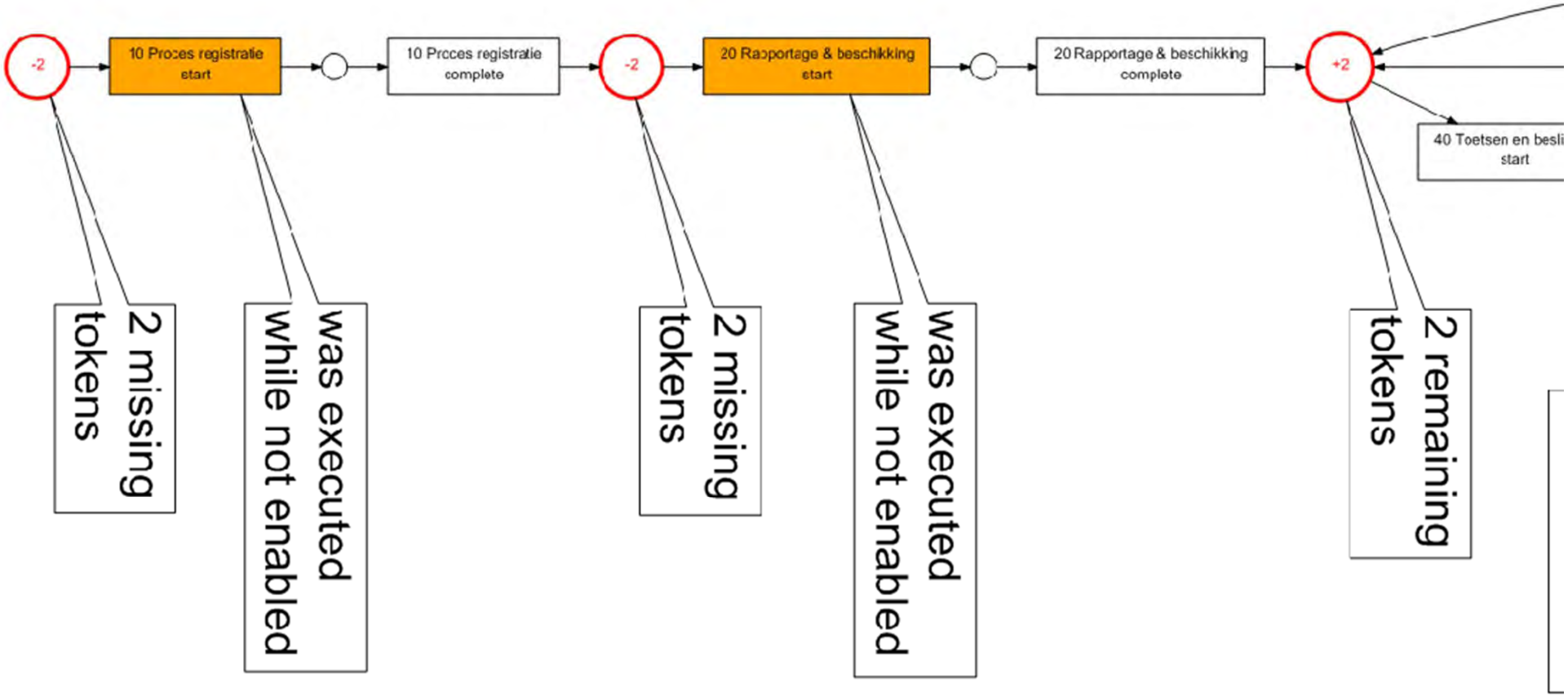


C-net discovered using heuristic miner (3/3)

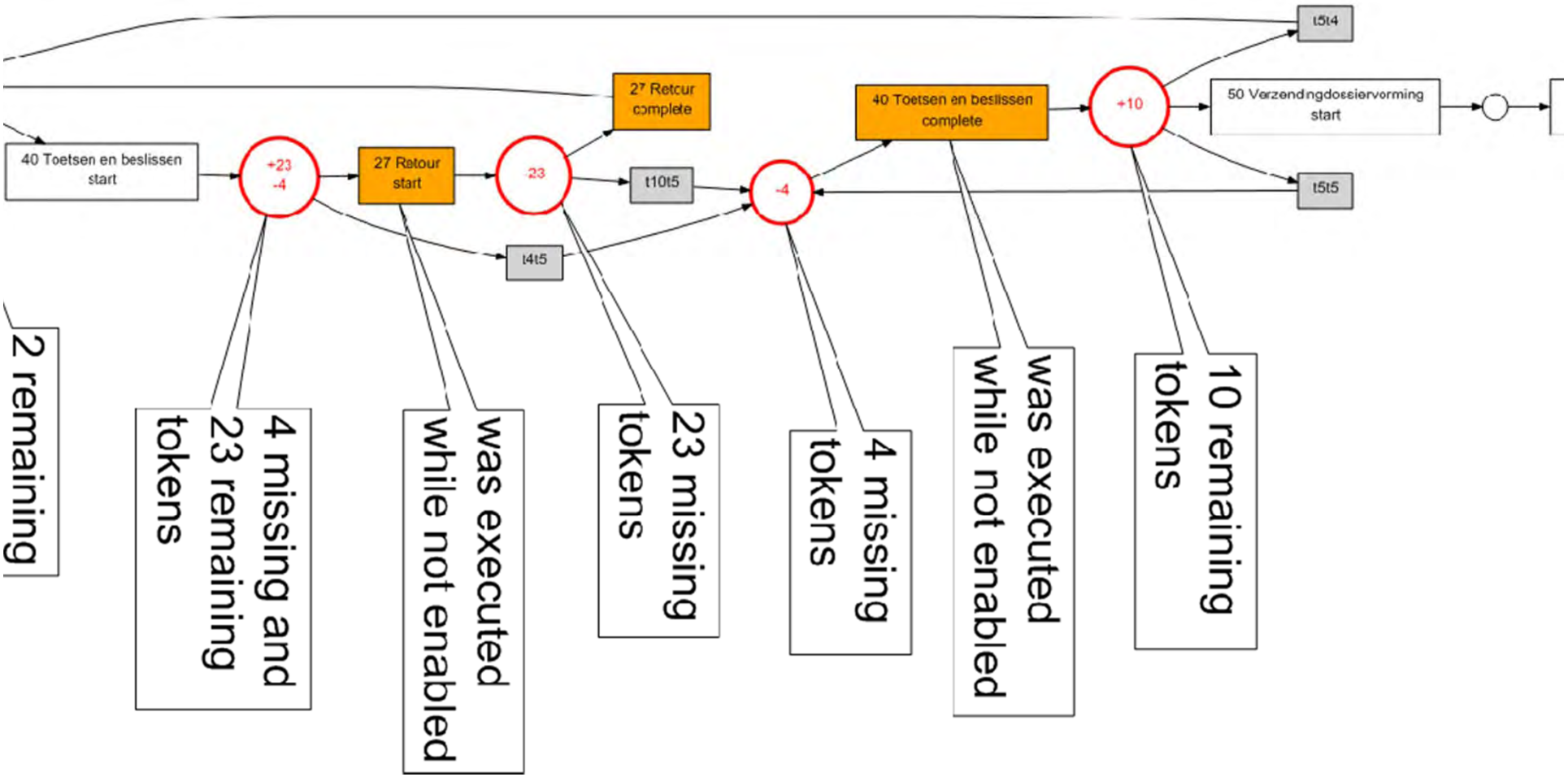




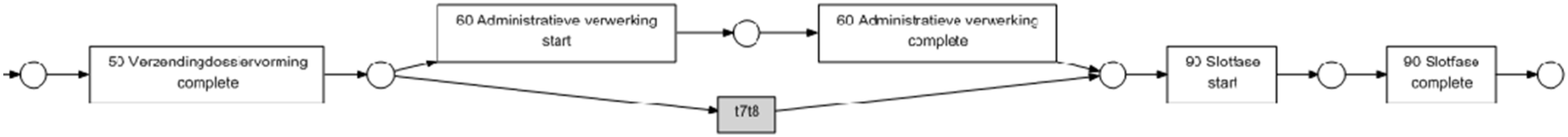
Conformance check WMO process (1/3)



Conformance check WMO process (2/3)

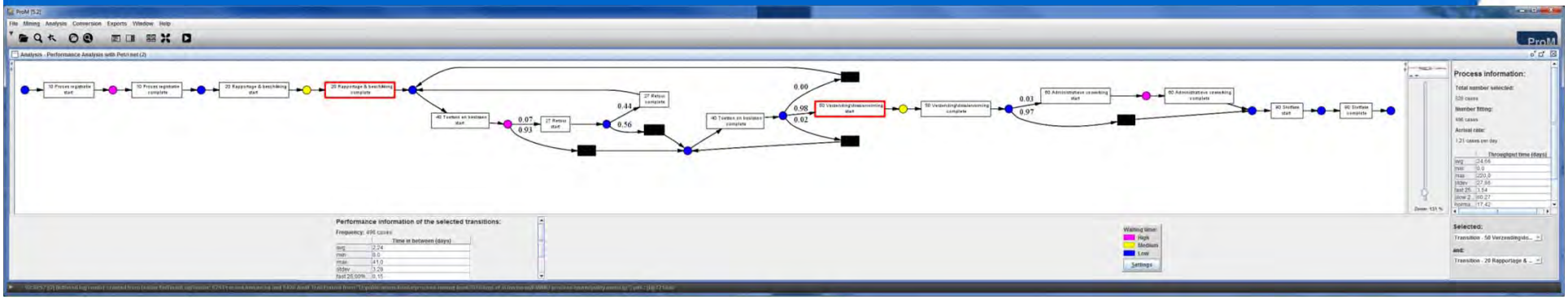


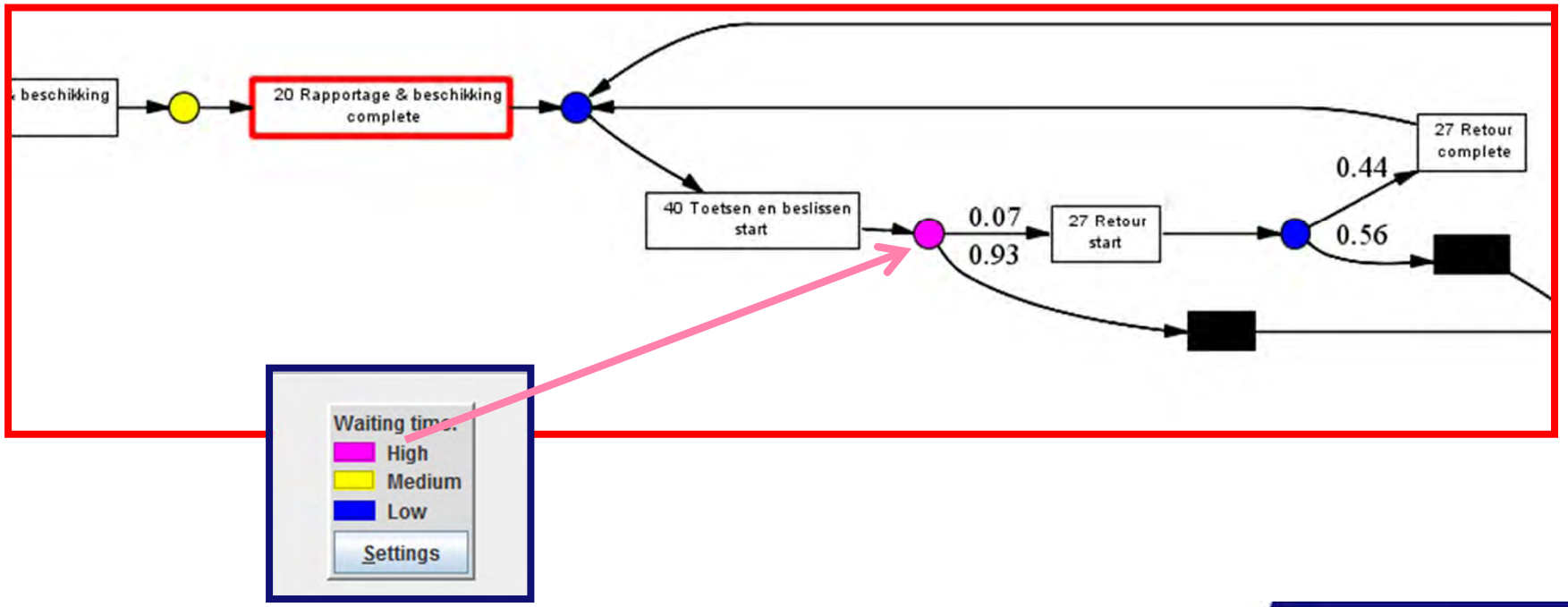
Conformance check WMO process (3/3)



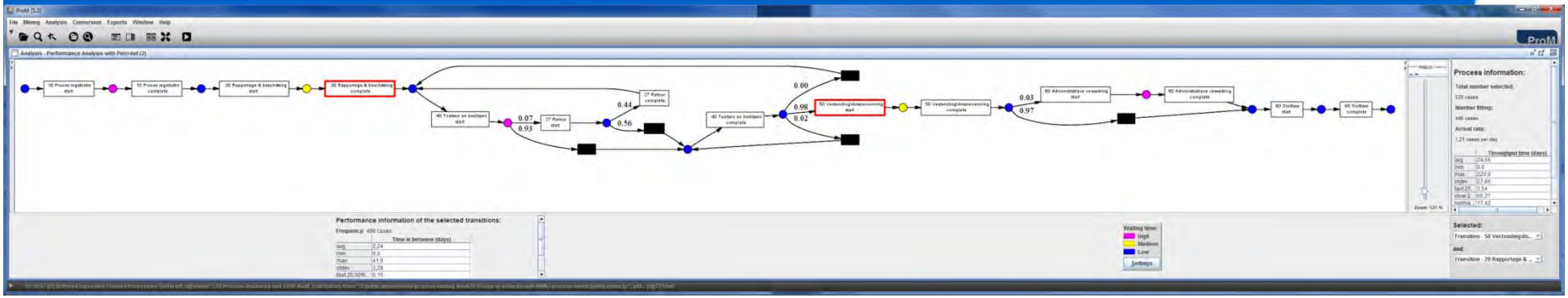
The fitness of the discovered process is 0.99521667. Of the 528 cases, 496 cases fit perfectly whereas for 32 cases there are missing or remaining tokens.

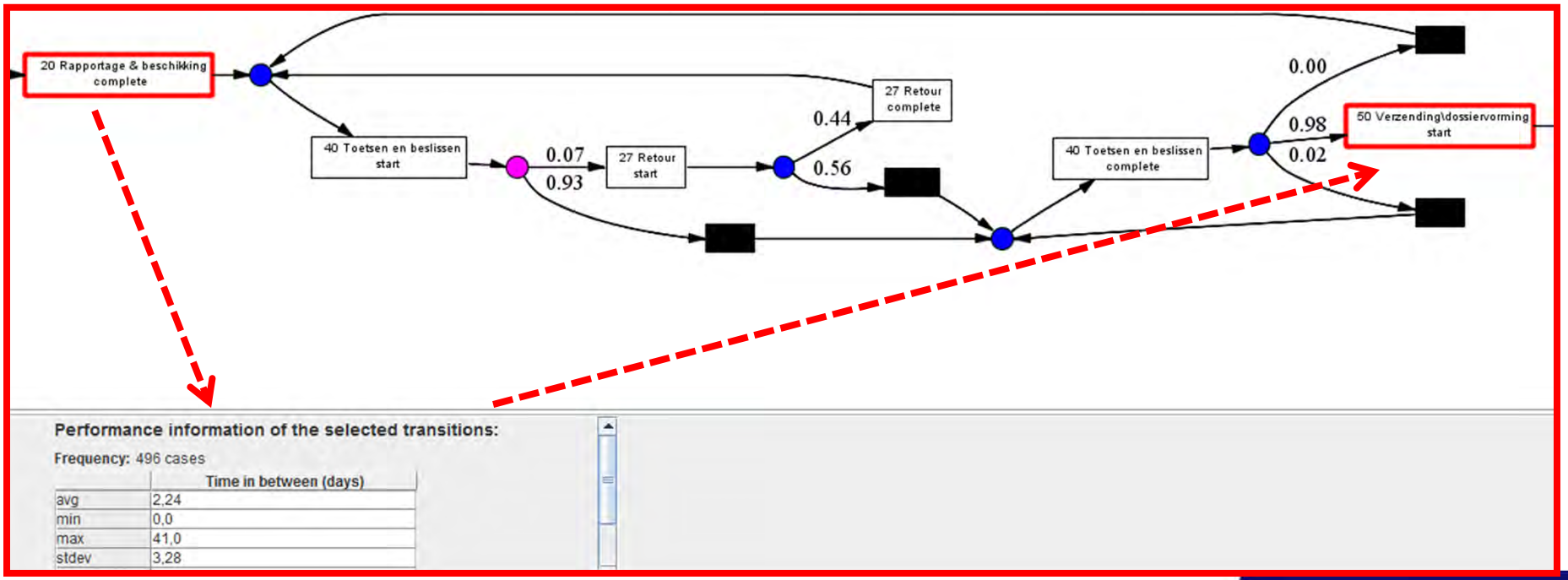
Bottleneck analysis WMO process (1/3)



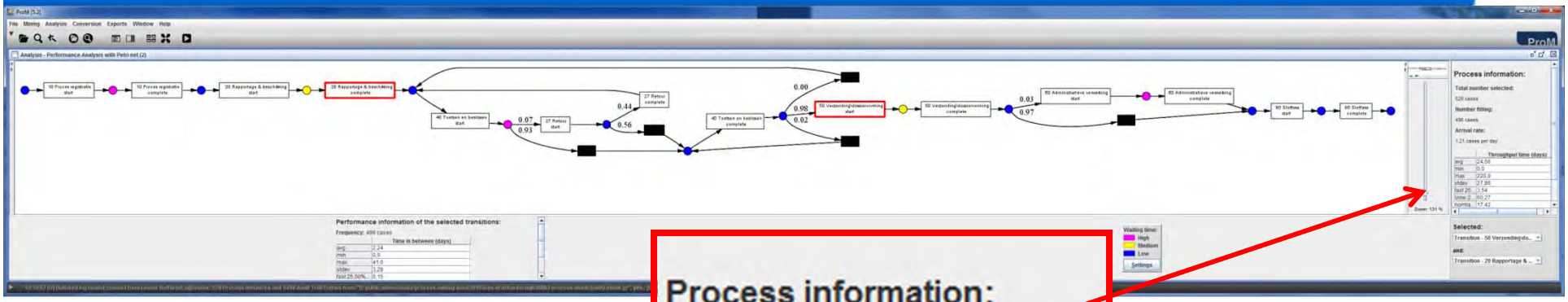


Bottleneck analysis WMO process (2/3)





Bottleneck analysis WMO process (3/3)



Process information:

Total number selected:

528 cases

Number fitting:

496 cases

Arrival rate:

1,21 cases per day

	Throughput time (days)
avg	24,66
min	0,0
max	220,0
stdev	27,86
fast 25...	3,54
slow 2...	60,27
norma...	17,42

flow time of approx. 25 days with a standard deviation of approx. 28

Two additional Lasagna processes

RWS (“Rijkswaterstaat”) process

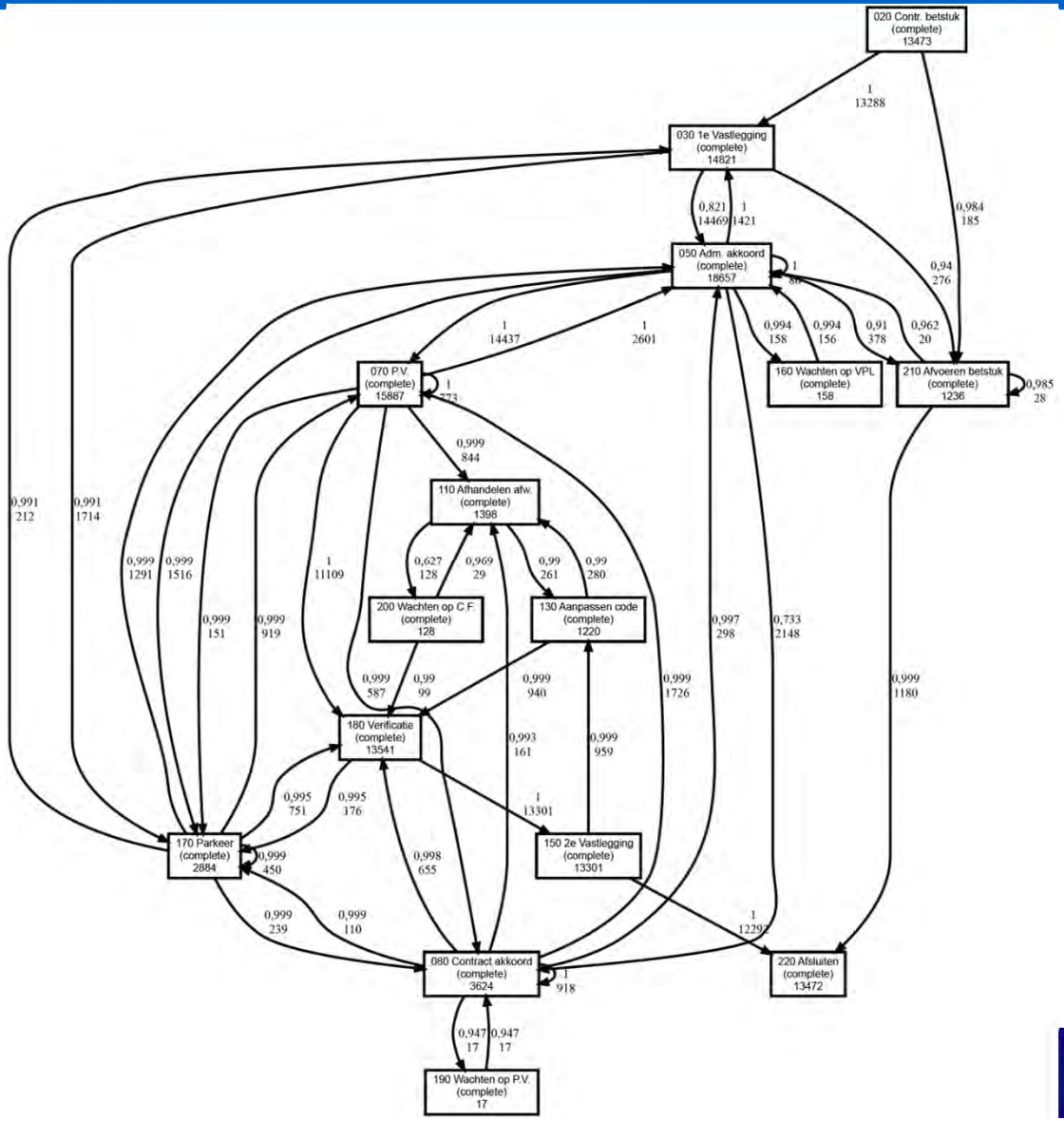
WOZ (“Waardering Onroerende Zaken”) process

RWS Process

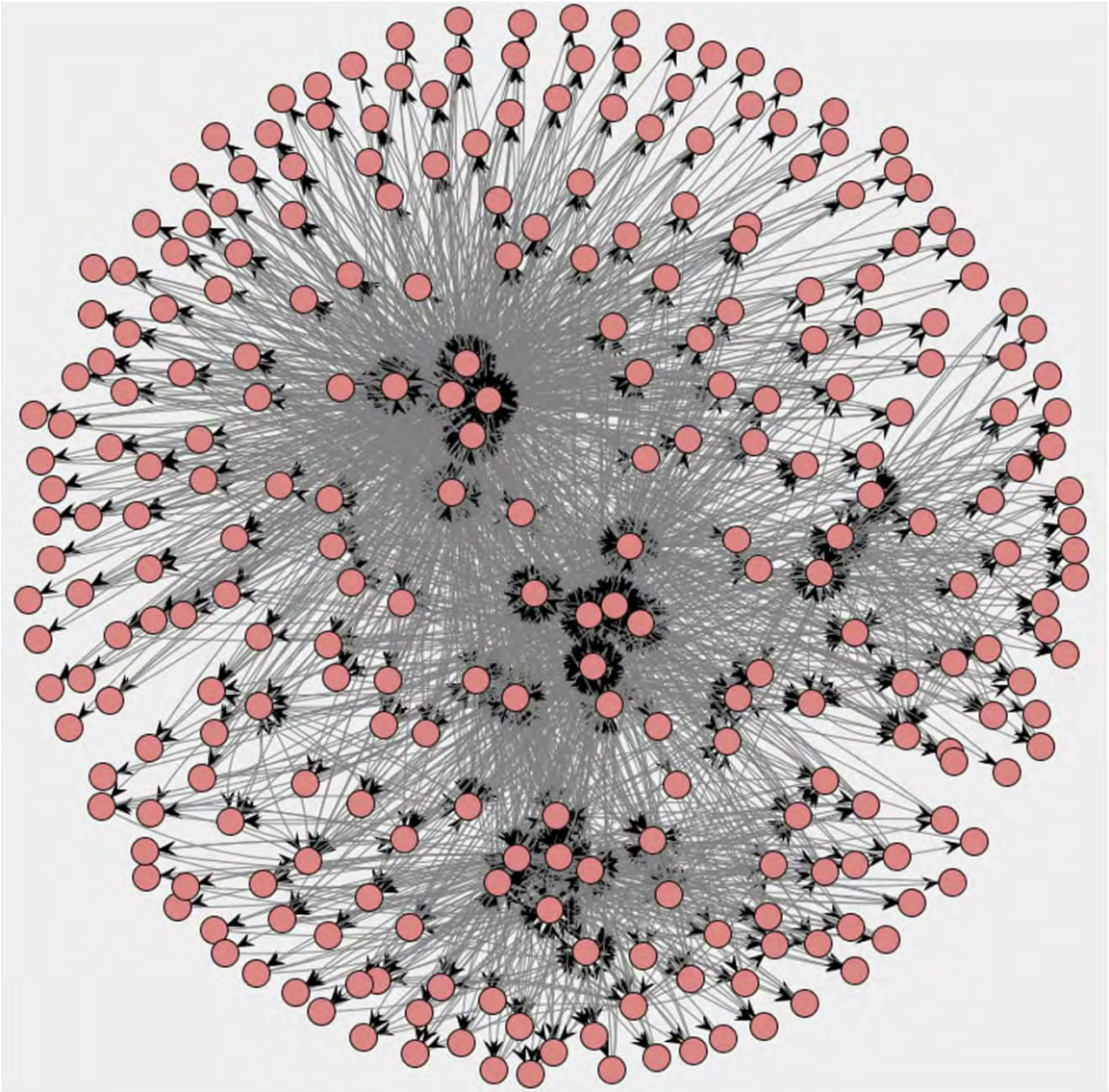
Rijkswaterstaat

- The Dutch national public works department, called “Rijkswaterstaat” (RWS), has twelve provincial offices. We analyzed the handling of invoices in one of these offices.
- The office employs about 1,000 civil servants and is primarily responsible for the construction and maintenance of the road and water infrastructure in its province.
- To perform its functions, the RWS office subcontracts various parties such as road construction companies, cleaning companies, and environmental bureaus. Also, it purchases services and products to support its construction, maintenance, and administrative activities.

C-net discovered using heuristic miner

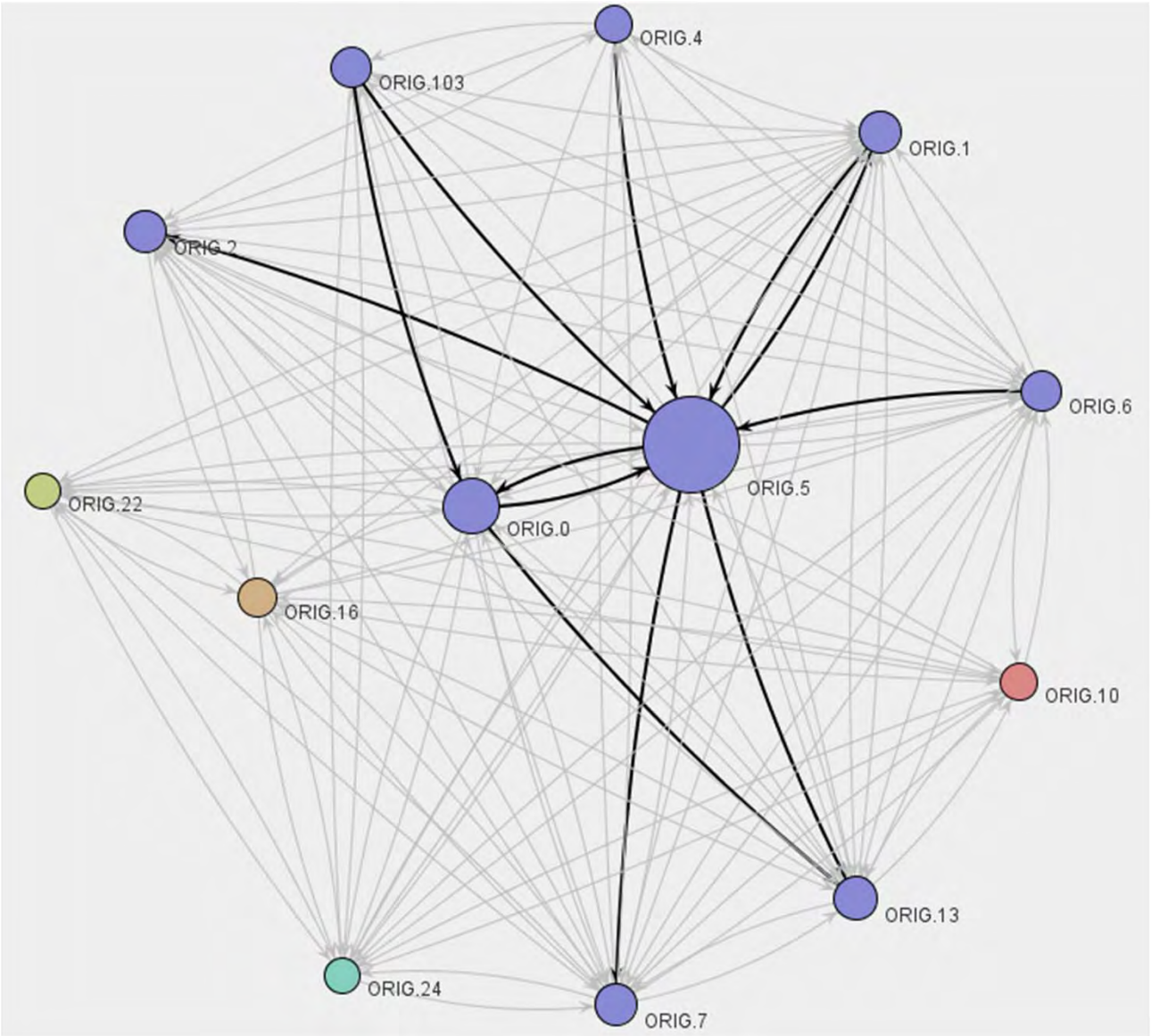


Social network constructed based on handovers of work



Each of the 271 nodes corresponds to a civil servant. Two civil servants are connected if one executed an activity causally following an activity executed by the other civil servant

Social network consisting of civil servants that executed more than 2000 activities in a 9 month period.

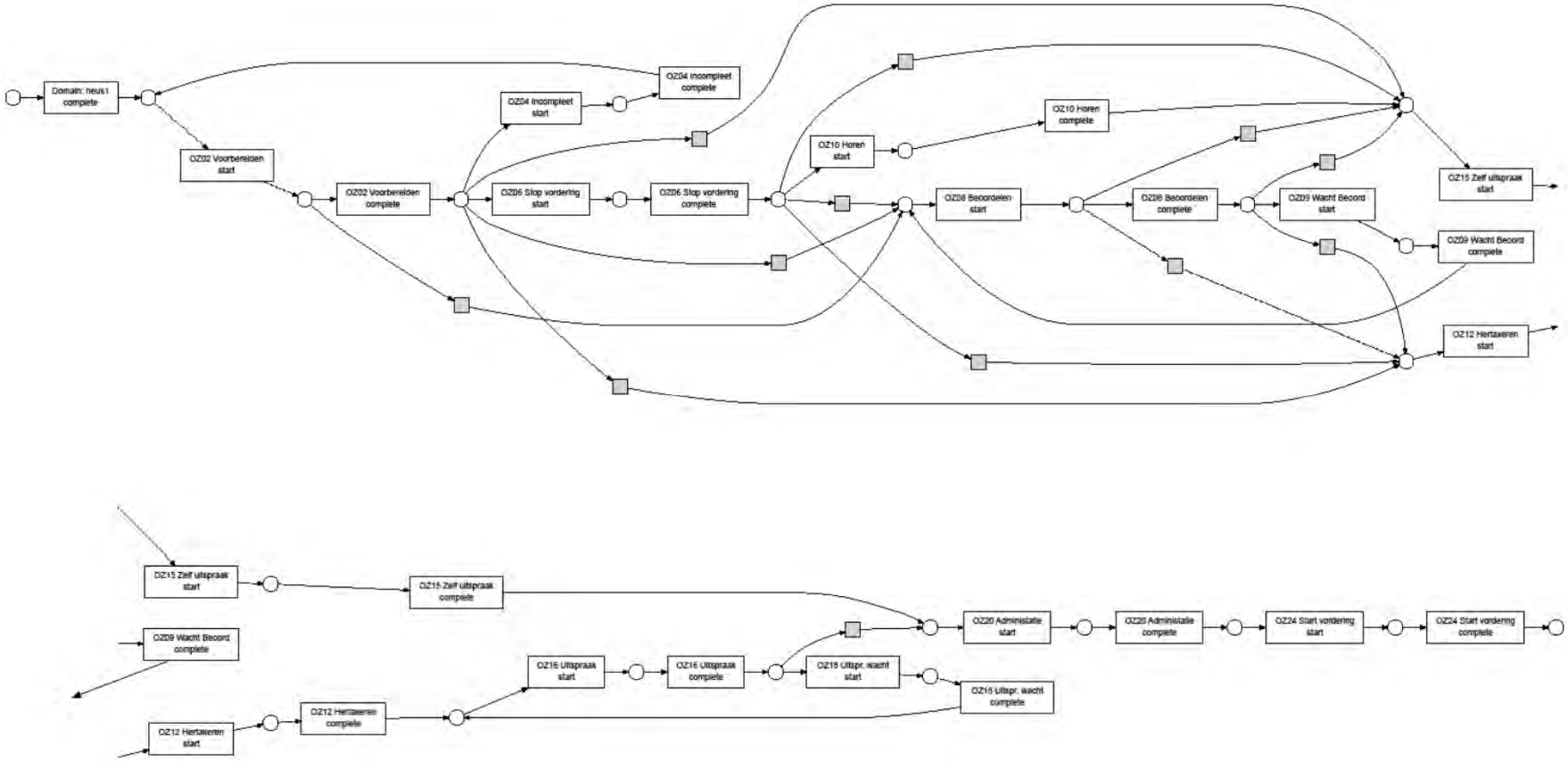


The darker arcs indicate the strongest relationships in the social network. Nodes having the same color belong to the same clique.

WOZ process

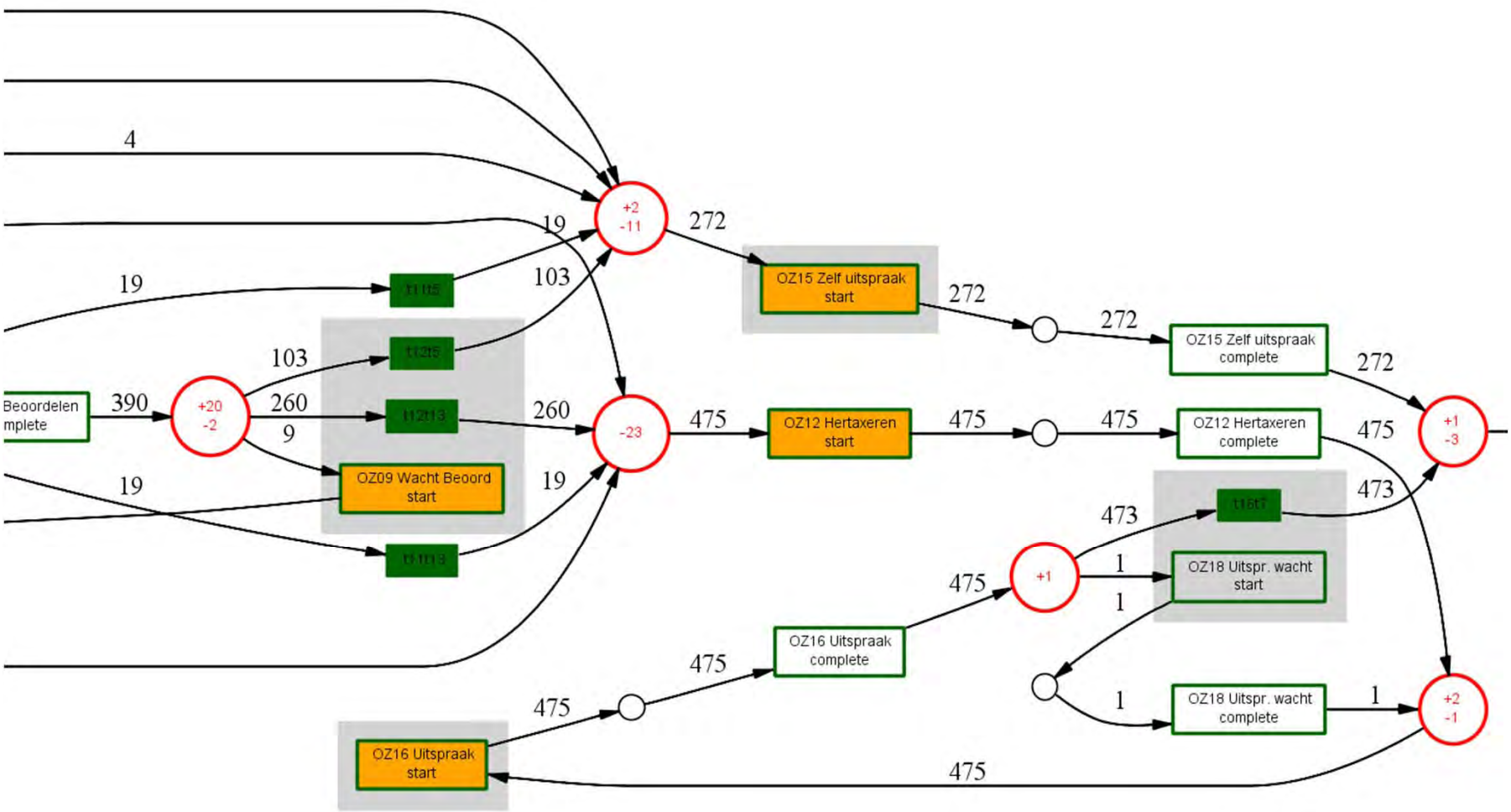
- **Event log containing information about 745 objections against the so-called WOZ (“Waardering Onroerende Zaken”) valuation.**
- **Dutch municipalities need to estimate the value of houses and apartments. The WOZ value is used as a basis for determining the real-estate property tax.**
- **The higher the WOZ value, the more tax the owner needs to pay. Therefore, there are many objections (i.e., appeals) of citizens that assert that the WOZ value is too high.**
- **“WOZ process” discovered for another municipality (i.e., different from the one for which we analyzed the WMO process).**

Discovered process model

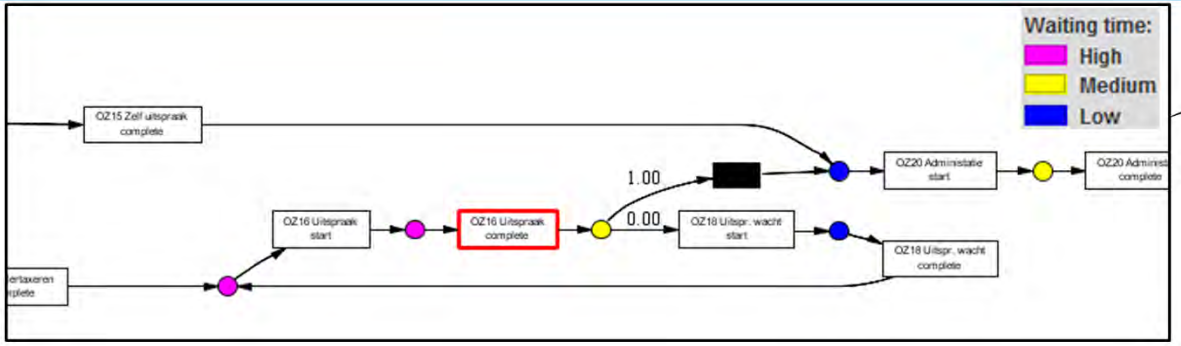


The log contains events related to 745 objections against the so-called WOZ valuation. These 745 objections generated 9583 events. There are 13 activities. For 12 of these activities both start and complete events are recorded. Hence, the WF-net has 25 transitions.

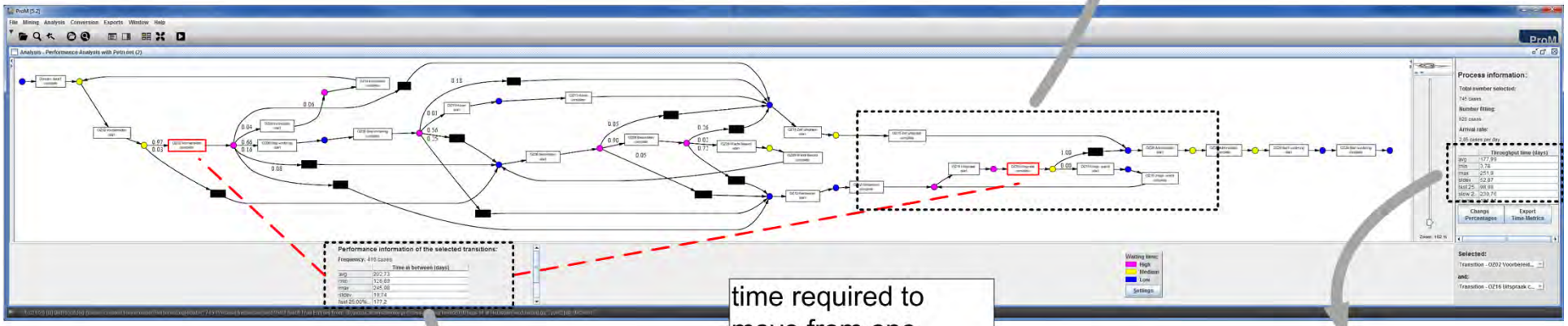
Conformance checker:
(fitness is 0.98876214)



Performance analysis



bottleneck detection: places are colored based on average durations



time required to move from one activity to another

Performance information of the selected transitions:

Frequency: 416 cases

	Time in between (days)
avg	202,73
min	126,89
max	245,98
stdev	19,74
fast 25,00%...	177,2

information on total flow time

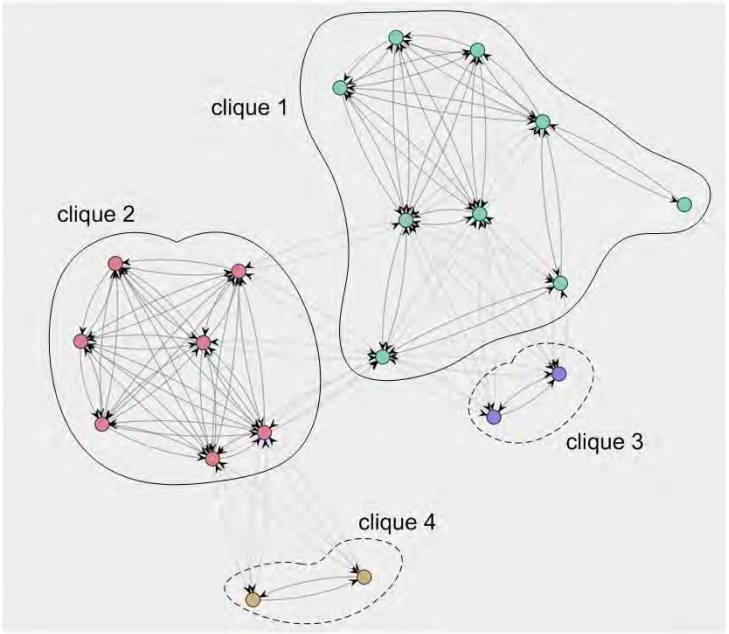
Arrival rate:

2,85 cases per day

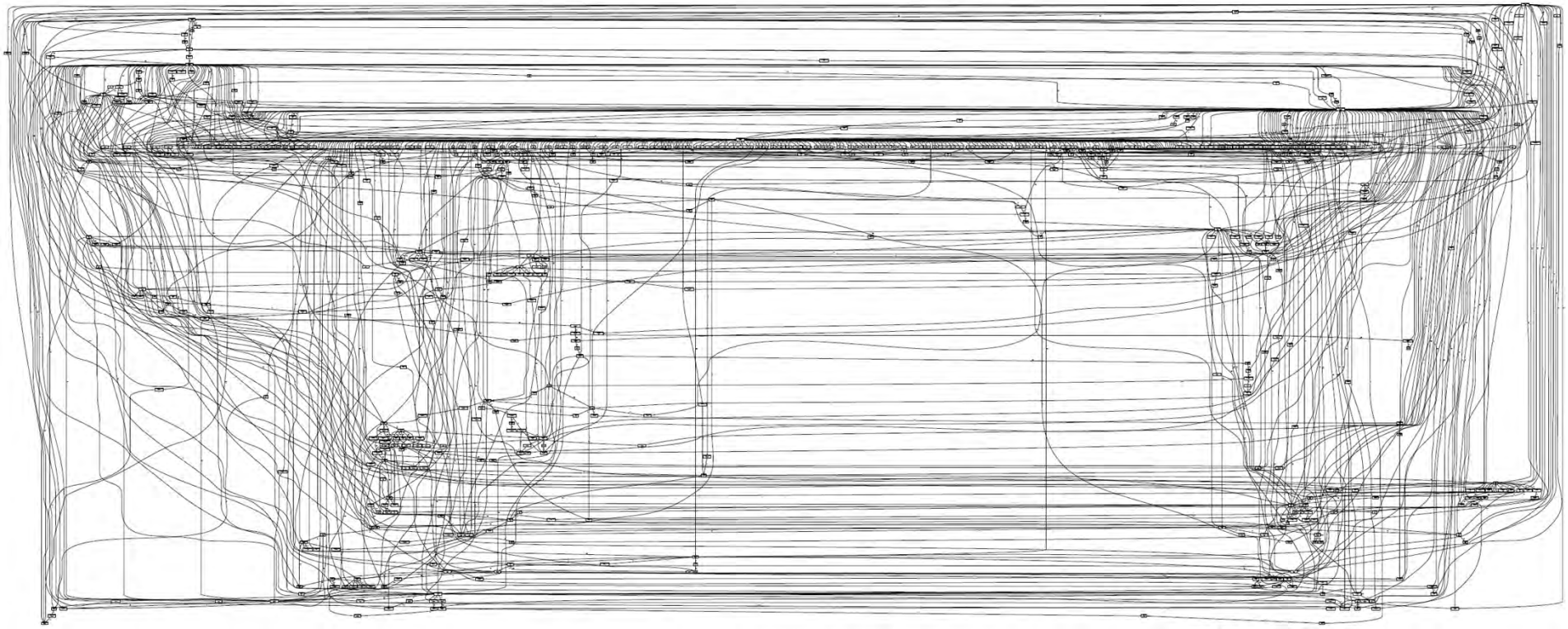
	Throughput time (days)
avg	177,99
min	3,78
max	251,9
stdev	52,87
fast 25...	98,98
slow 2...	230,76
norma...	191,11

Resource-activity matrix (four groups discovered)

user	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
user 1	0	0	51	0	0	0	0	0	0	0	0	0	0
user 2	1	2	0	0	2	0	0	0	0	38	0	69	0
user 3	0	9	0	0	0	0	0	0	0	0	0	0	0
user 4	2	0	0	0	0	0	0	0	0	0	0	0	0
user 5	117	0	4	0	3	0	0	0	0	1	0	20	6
user 6	172	6	14	0	7	3	0	0	1	2	0	48	53
user 7	1	41	8	14	275	8	8	865	55	180	0	128	5
user 8	2	868	7	6	105	0	0	79	266	441	0	844	3
user 9	90	0	2	0	1	2	0	0	1	2	0	27	28
user 10	0	0	0	899	0	0	0	0	0	0	0	0	1019
user 11	336	1	3	1	4	2	0	0	0	1	0	18	23
user 12	1	645	13	21	419	3	0	3	217	281	1	334	9
user 13	0	1	0	0	0	0	0	0	0	0	0	0	0
user 14	0	0	0	0	0	0	0	0	0	1	0	0	0
user 15	0	0	0	0	0	0	0	2	2	0	0	2	0
user 16	1	3	3	2	1	0	0	1	2	3	1	0	0
user 17	0	4	0	0	0	0	0	0	0	0	0	0	0
user 18	9	0	0	0	0	0	0	0	0	0	0	0	0
user 19	13	1	0	0	1	0	0	0	0	0	0	4	0
user 20	0	0	0	21	0	0	0	0	0	0	0	0	258

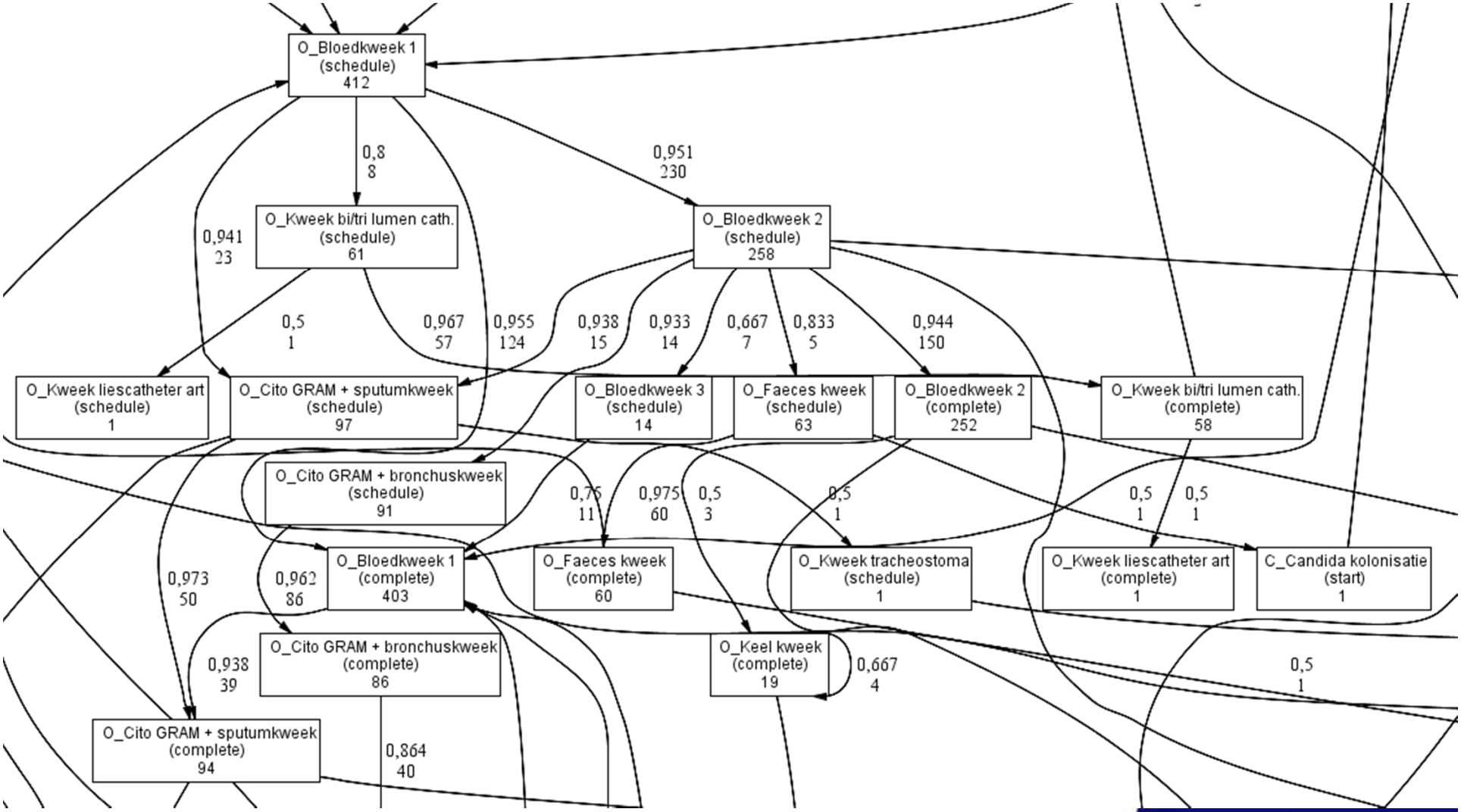


Example of a Spaghetti process

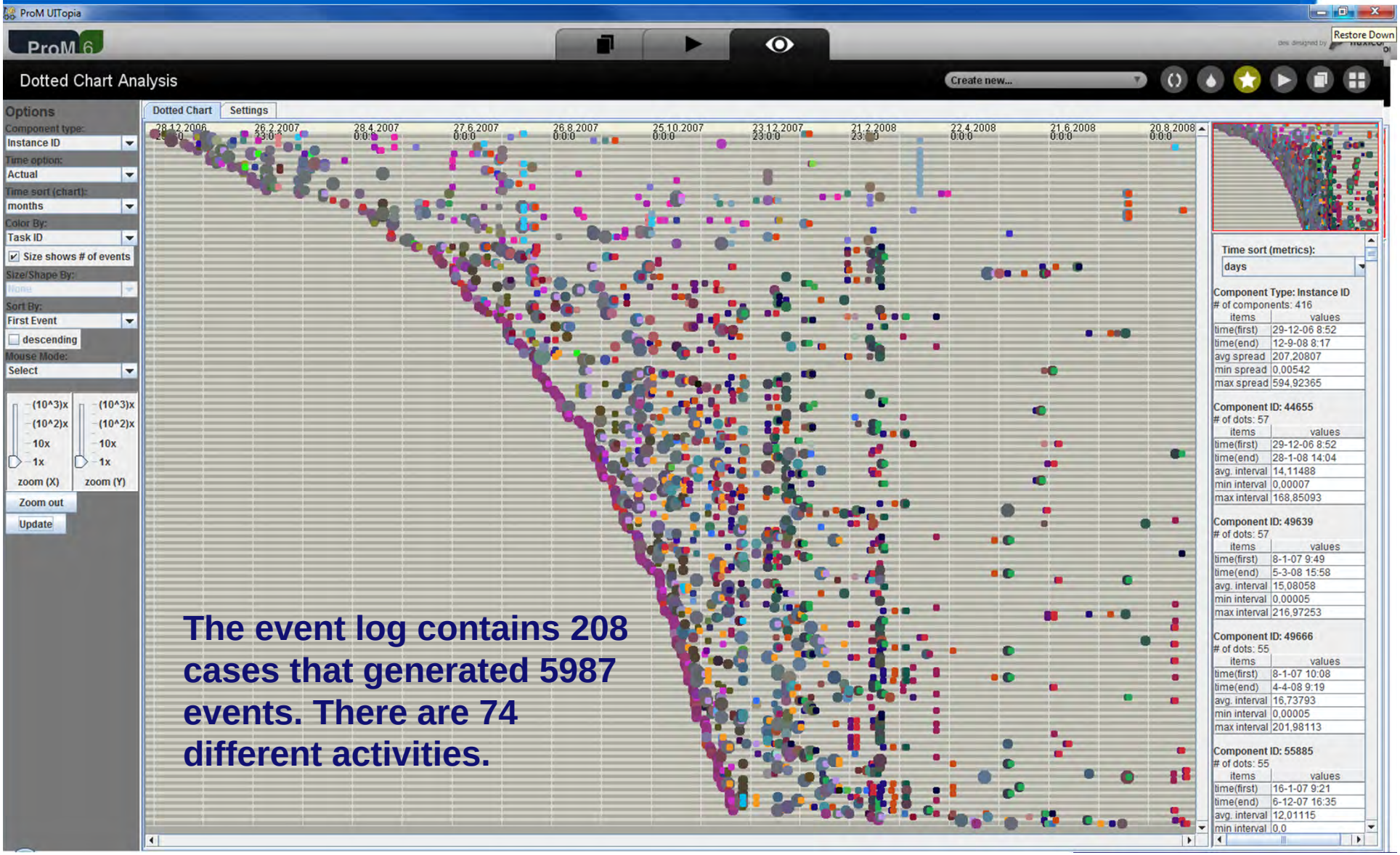


Spaghetti process describing the diagnosis and treatment of 2765 patients in a Dutch hospital. The process model was constructed based on an event log containing 114,592 events. There are 619 different activities (taking event types into account) executed by 266 different individuals (doctors, nurses, etc.).

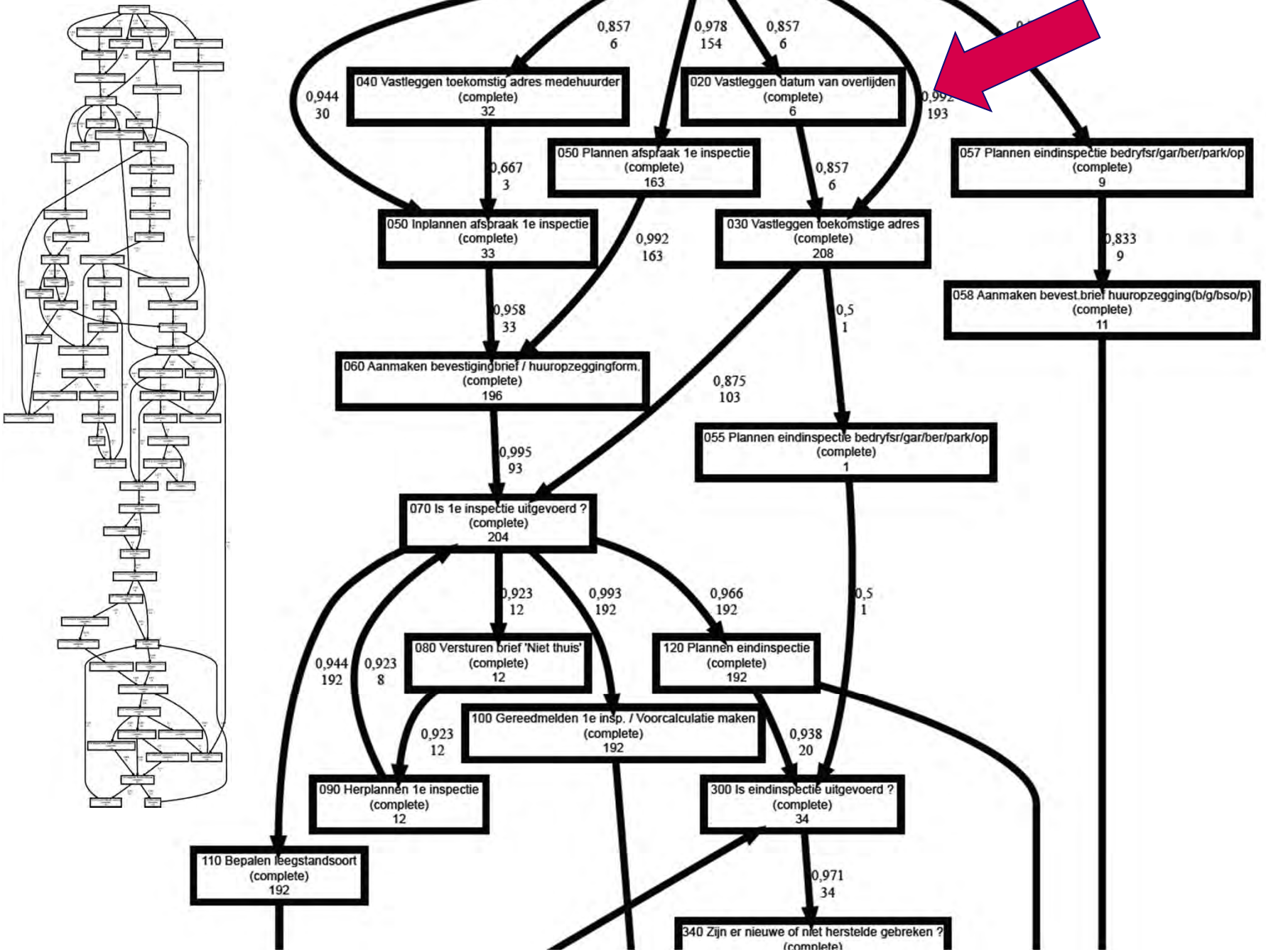
Fragment
18 activities of the 619 activities (2.9%)



Another example
(event log of Dutch housing agency)



The event log contains 208 cases that generated 5987 events. There are 74 different activities.

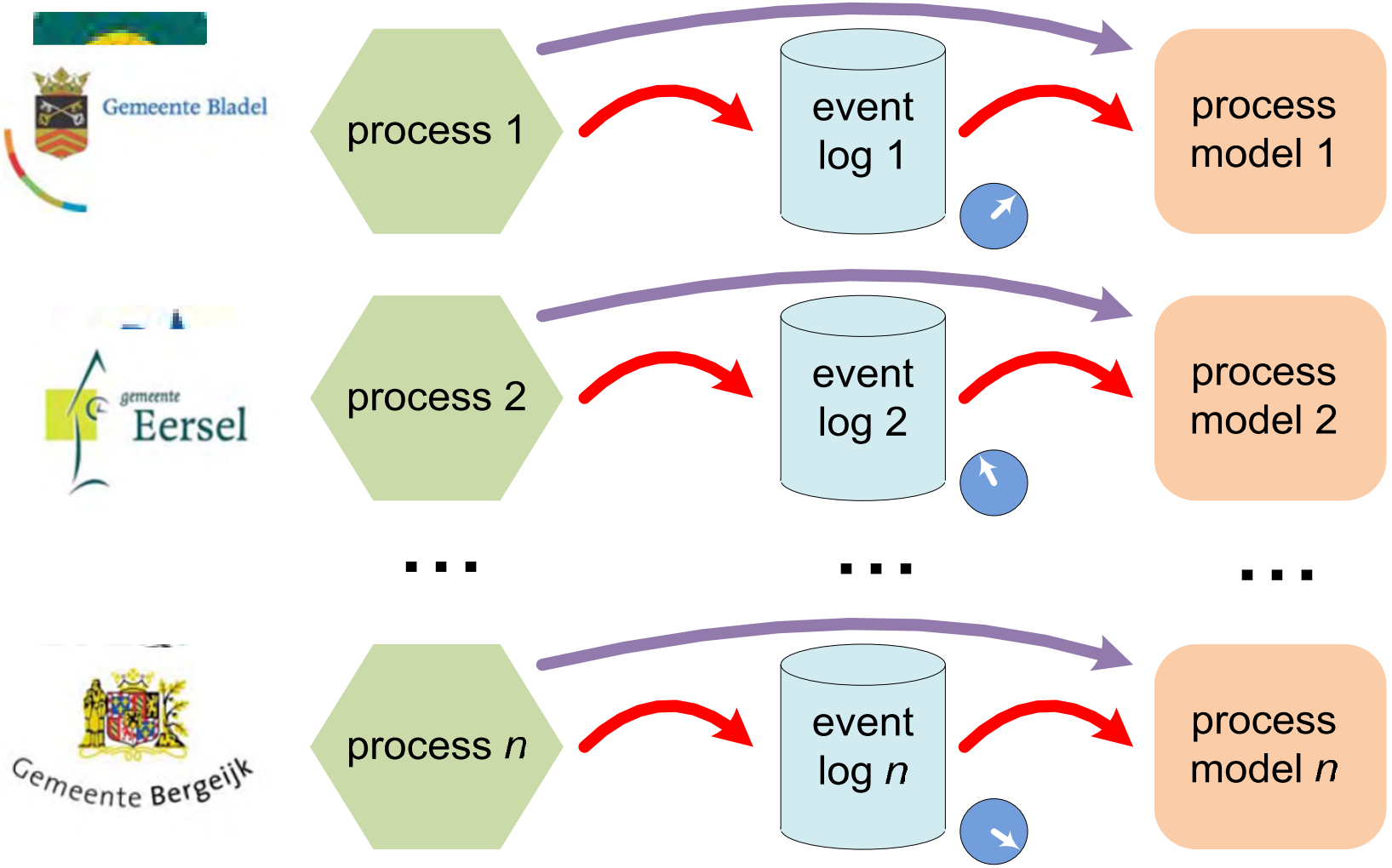

010 Registreren huuropzegging
(complete)
208
0,857
6
0,978
154
0,857
6
040 Vastleggen toekomstig adres medehuurder
(complete)
32
020 Vastleggen datum van overlijden
(complete)
6
0,944
30
193
057 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
9
050 Plannen afspraak 1e inspectie
(complete)
163
0,667
3
0,857
6
050 Inplannen afspraak 1e inspectie
(complete)
33
0,992
163
030 Vastleggen toekomstige adres
(complete)
208
0,833
9
058 Aanmaken bevest.brief huuropzegging(b/g/bso/p)
(complete)
11
0,958
33
0,5
1
060 Aanmaken bevestigingbrief / huuropzeggingform.
(complete)
196
0,875
103
055 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
1
0,995
93
070 Is 1e inspectie uitgevoerd ?
(complete)
204
0,923
12
0,993
192
0,966
192
0,5
1
080 Versturen brief 'Niet thuis'
(complete)
12
120 Plannen eindinspectie
(complete)
192
0,944
192
0,923
8
100 Gereedmelden 1e insp. / Voorcalculatie maken
(complete)
192
0,923
12
0,938
20
090 Herplannen 1e inspectie
(complete)
12
300 Is eindinspectie uitgevoerd ?
(complete)
34
110 Bepalen leegstandsoort
(complete)
192
0,971
34
340 Zijn er nieuwe of niet herstelde gebreken ?
(complete)


Cross-organizational
mining

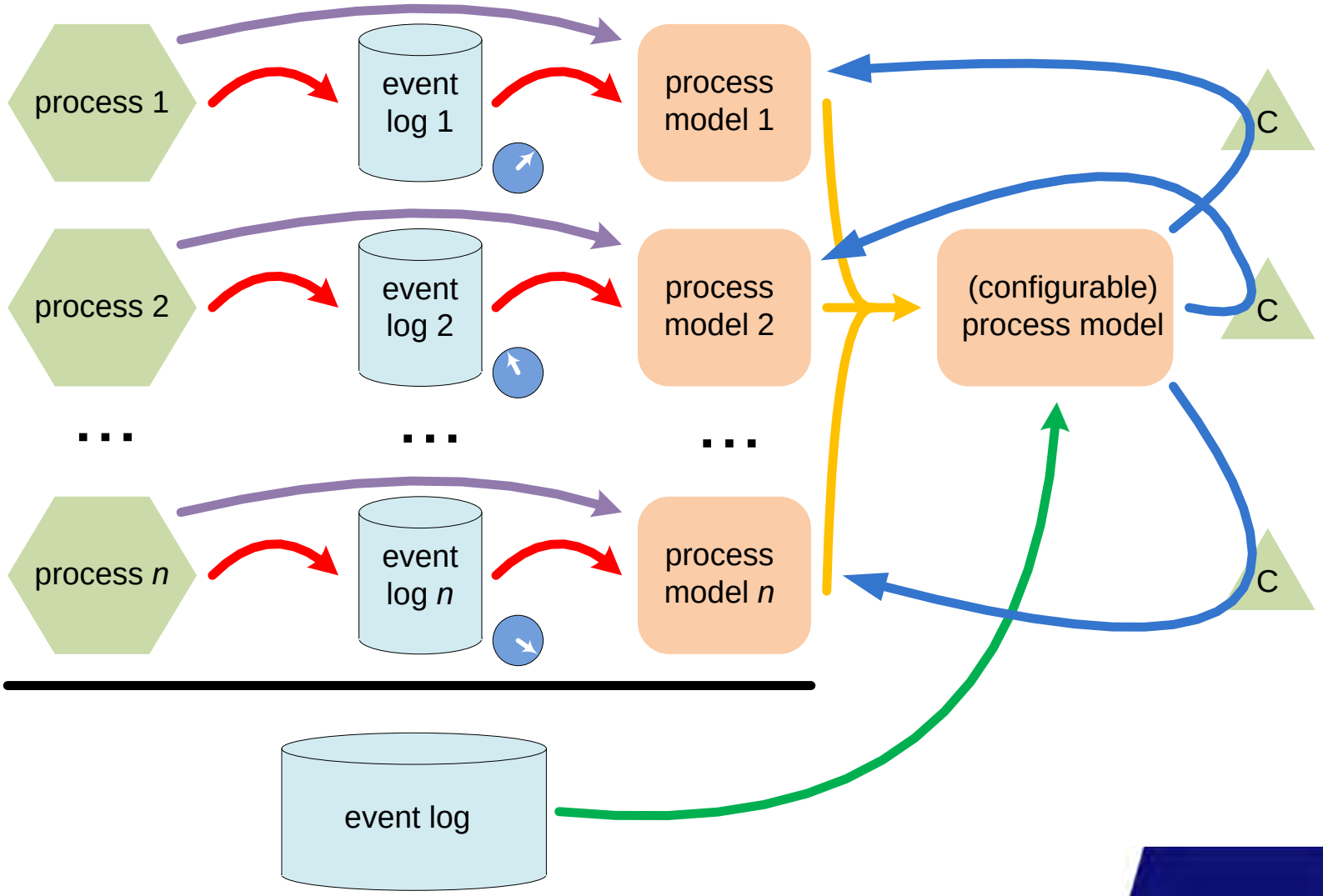
From one to many organizations

- More than 80,000 organizations are using Salesforce
- More than 1 million organizations are using Google Apps
- All 430 Dutch municipalities are implementing the same set of processes
- All 94 U.S. District Courts in the United States share the same set of workflows
- All car-rental offices of Hertz, Avis, …
- …

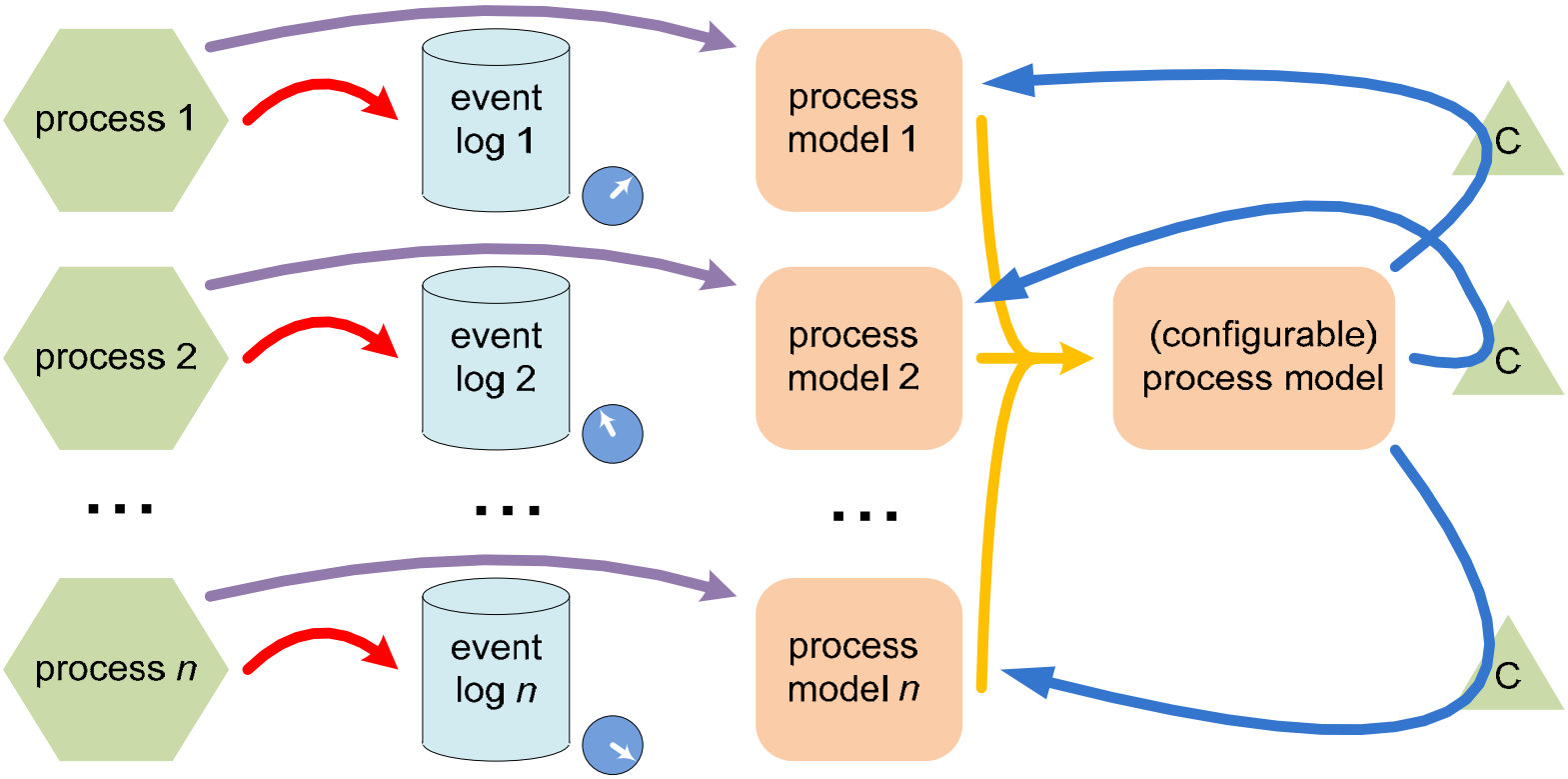
Consider *n* organizations



Cross-organizational process mining



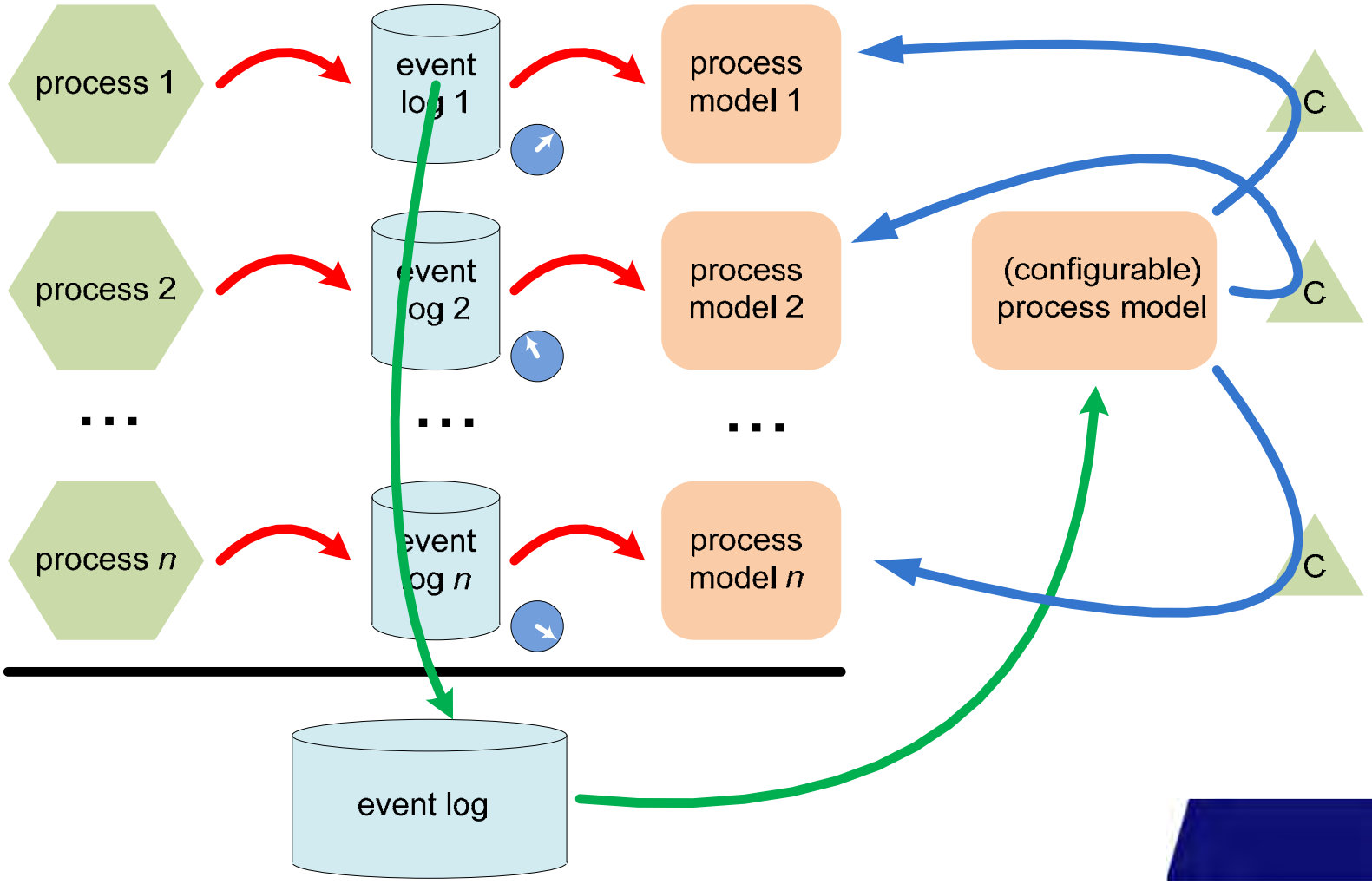
Pure model-based

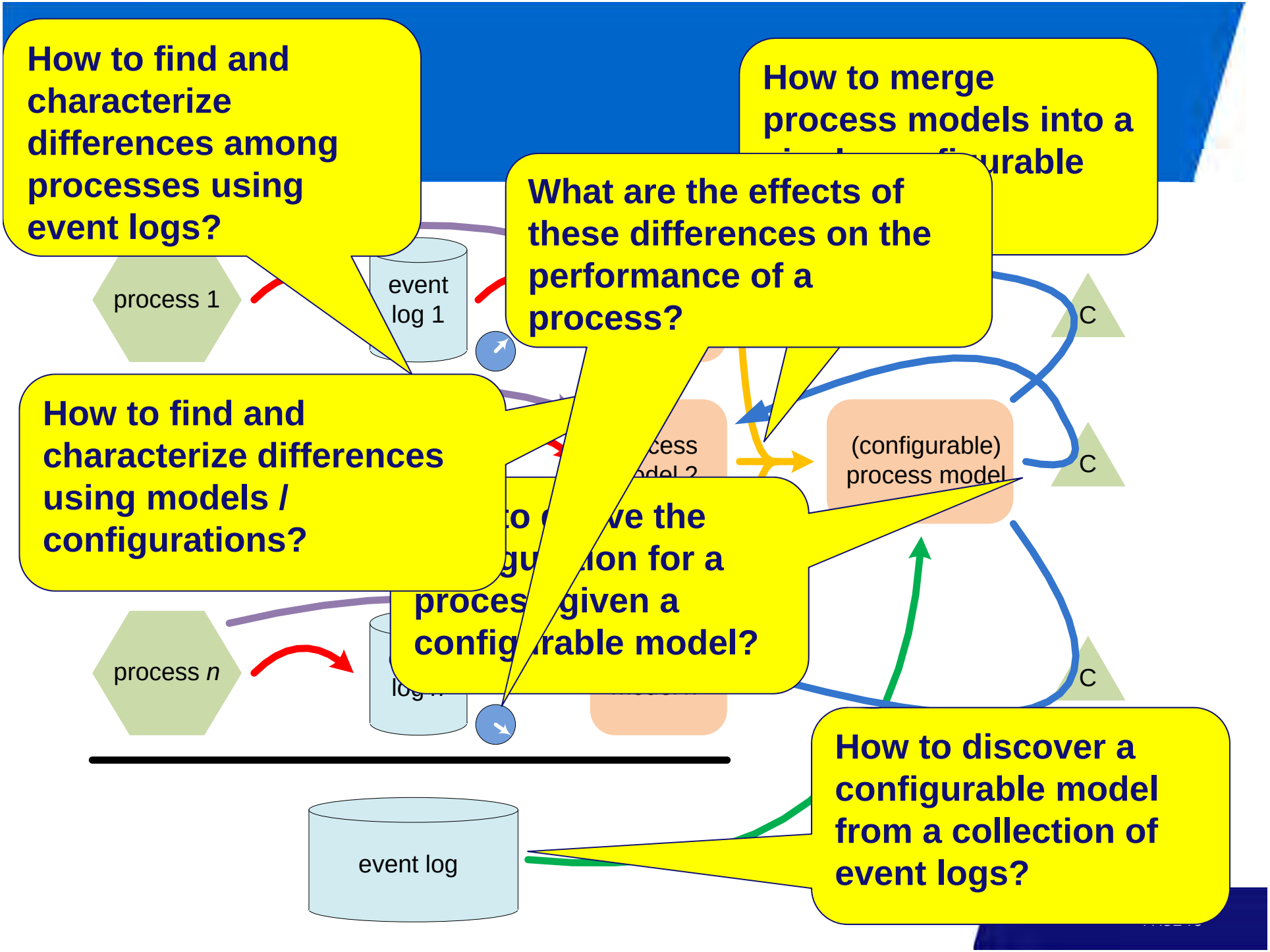


$$PM_1 + PM_2 + \ldots + PM_n = CM$$

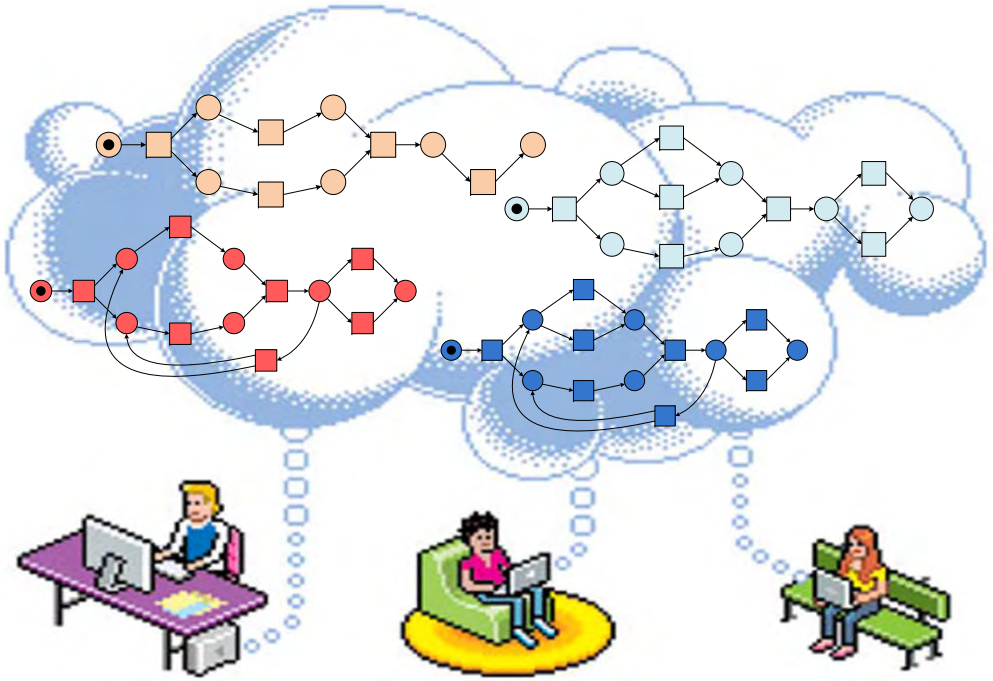
Pure log-based

$$\alpha(EL_1 + EL_2 + \ldots + EL_n) = CM$$




How to find and characterize differences among processes using event logs?
How to merge process models into a
What are the effects of these differences on the performance of a process?
How to find and characterize differences using models / configurations?
process 1
event log 1
process n
(configurable) process model
C
C
C
event log
How to discover a configurable model from a collection of event logs?

Evidence-based “best practices”



- Organizations can learn from each other.
- Configuration support and diagnostics.
- Software vendors/service providers can improve their products/services.

Conclusion

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

ProM
process mining workbench

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

www.processmining.org

www.win.tue.nl/ieeetfpm/

Springer