

Configurable Services in the Cloud

**Supporting variability
while enabling
cross-organizational
process mining**

Wil van der Aalst



TU/e

Technische Universiteit
Eindhoven
University of Technology

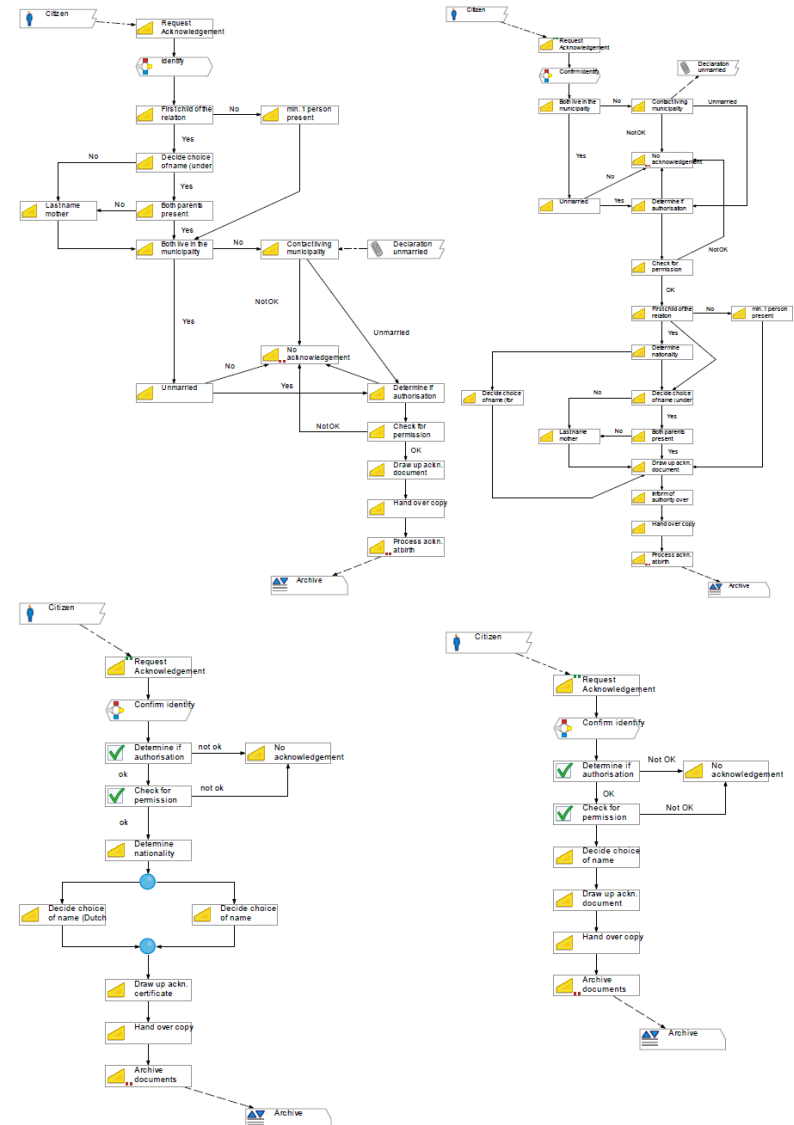
Where innovation starts

Acknowledgements

- **Marcello La Rosa**
- **Florian Gottschalk**
- **CoSeLoG: Joos Buijs, Jan Vogelaar, Boudewijn van Dongen, Eric Verbeek, Hajo Reijers.**
- **Marlon Dumas, Arthur ter Hofstede, Niels Lohmann, Michael Rosemann, Jan Mendling, ...**
- **ProM team (www.processmining.org)**
- **YAWL team (www.yawlfoundation.org)**



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



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

500
steps



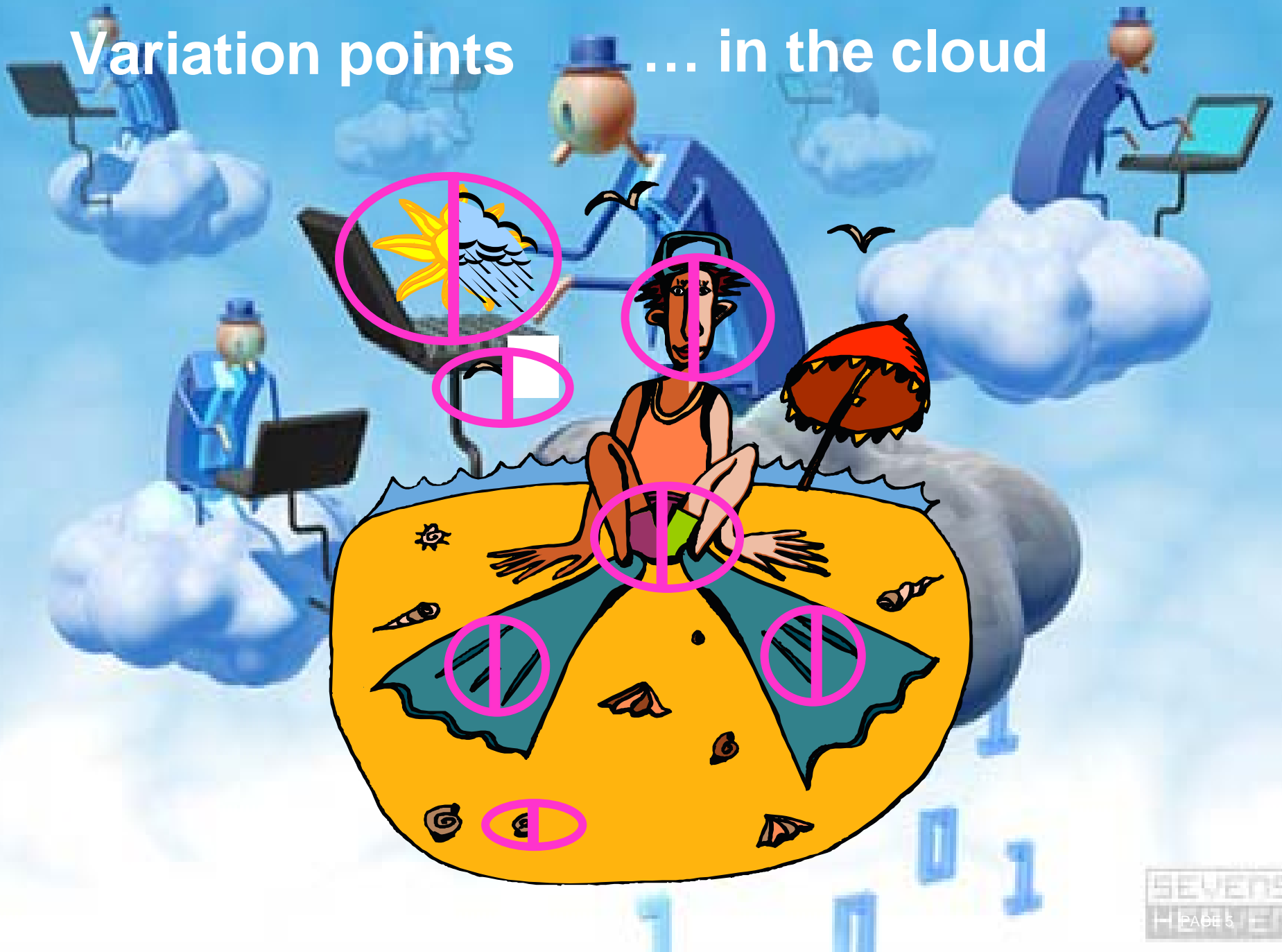
	Home								
	Home	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Motor	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Commercial	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	Liability	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	CTP / WC	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

30
variations

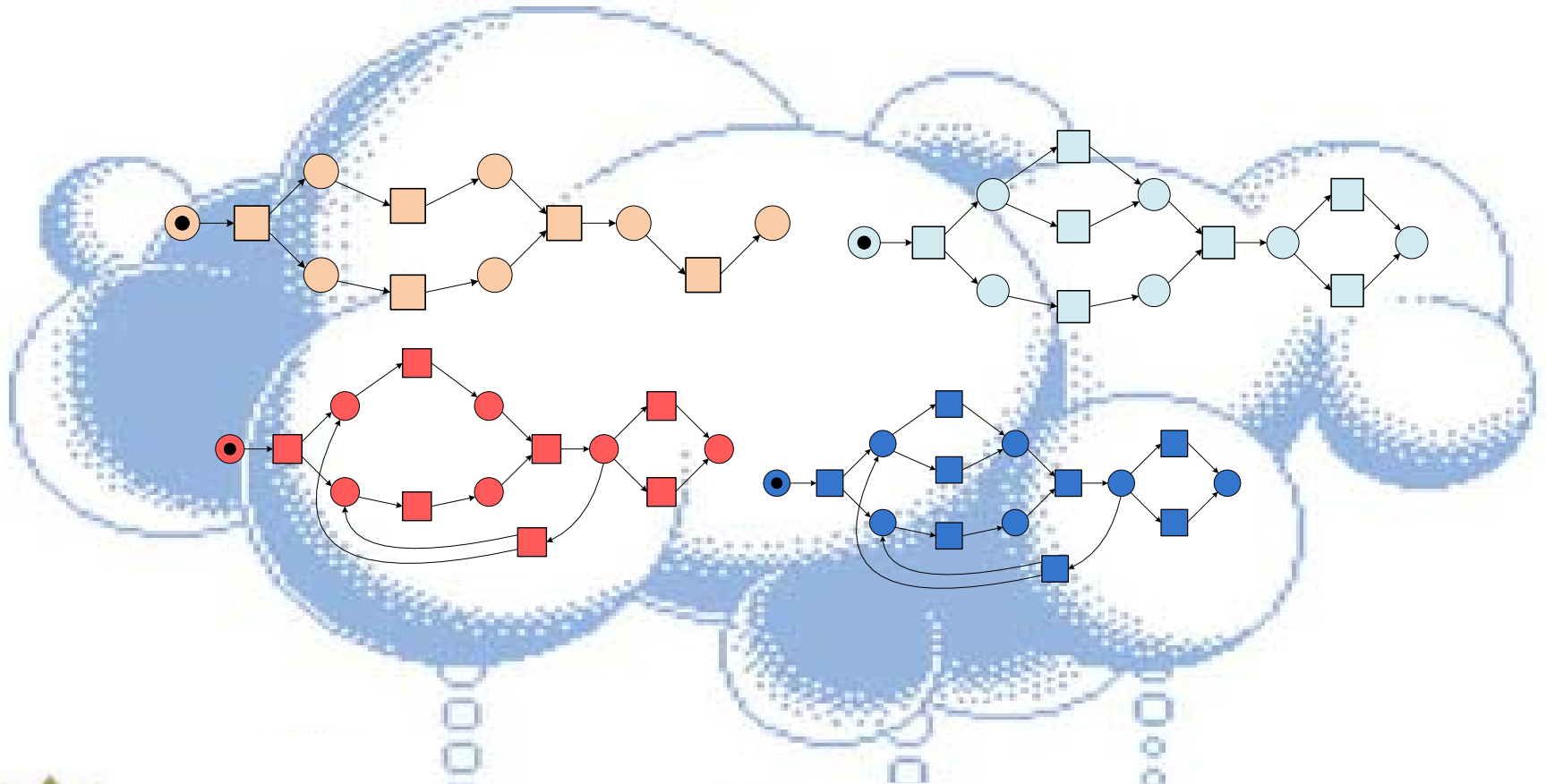
Two variants of the same process ...



Variation points ... in the cloud



Cloud computing



Gemeente Bladel

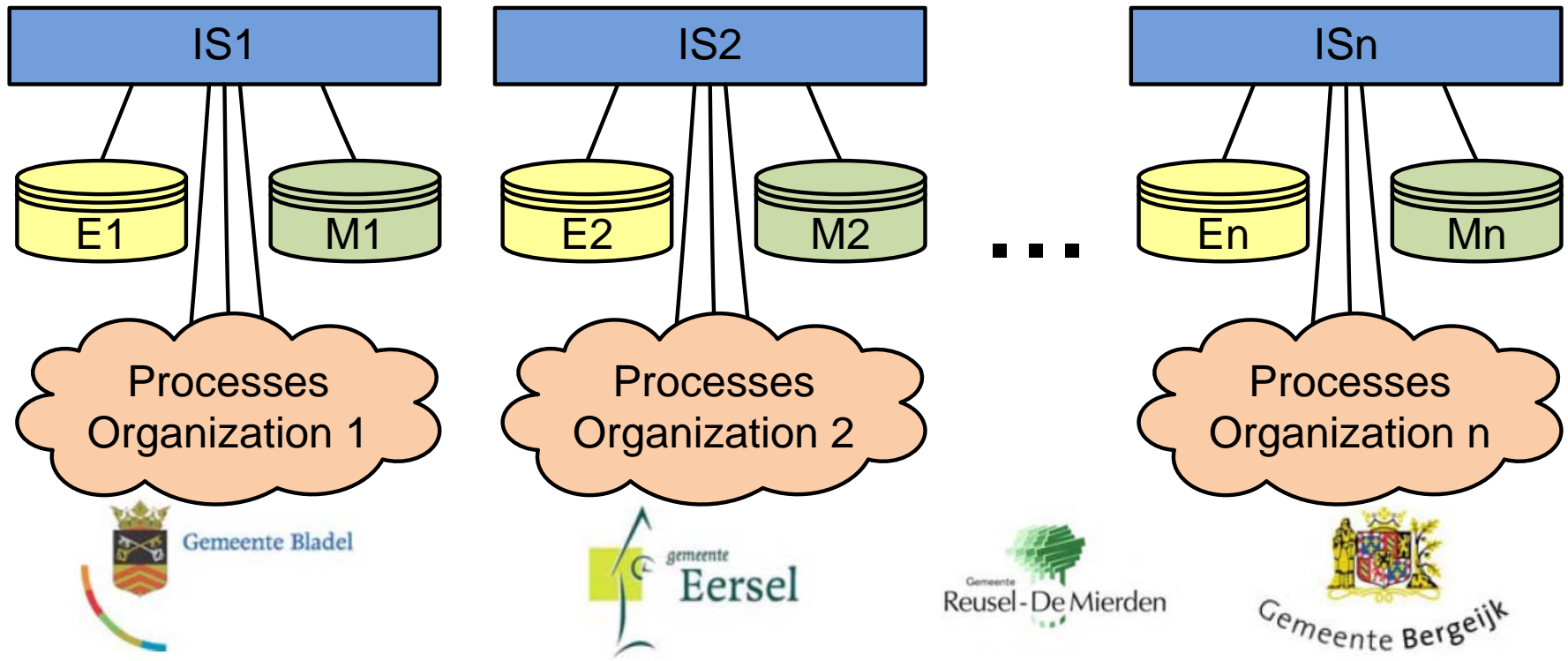


gemeente
Eersel



Gemeente Bergeijk

Traditional Situation



IS = Information System

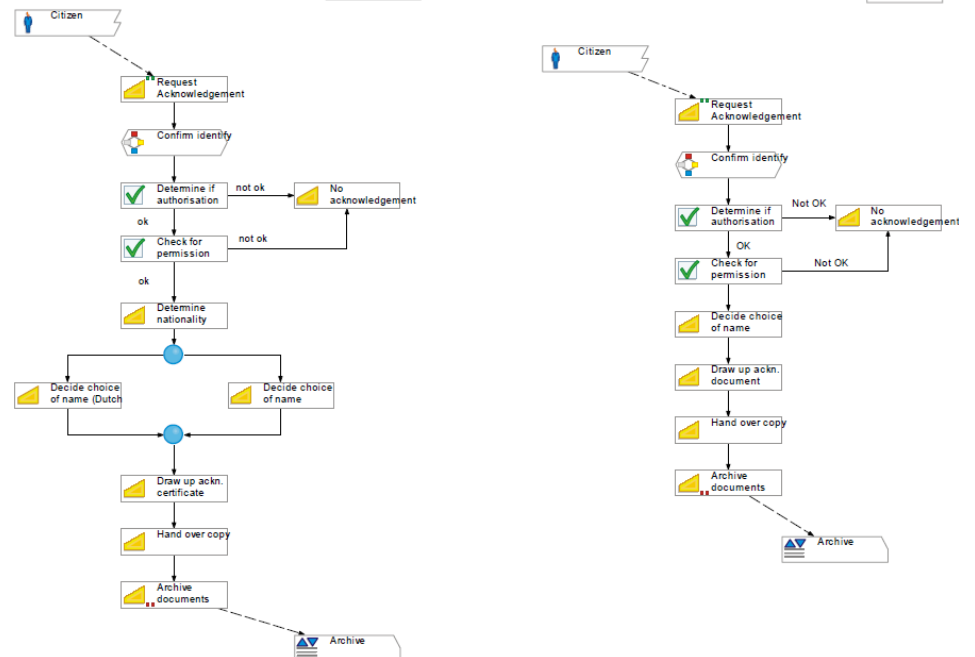
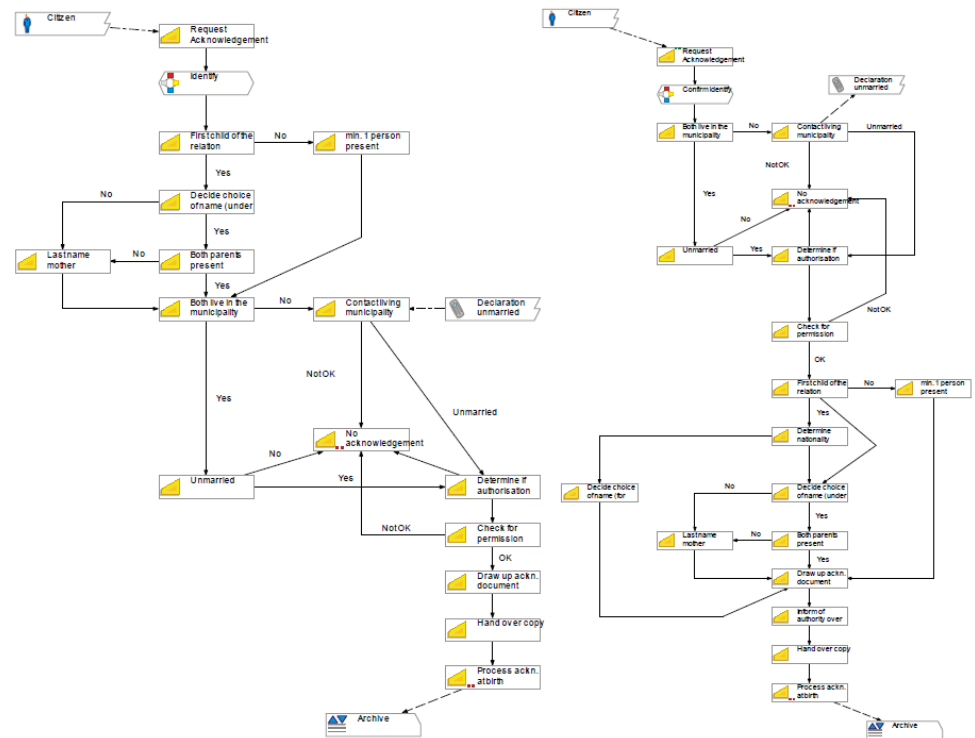
E = Event log

M = Models

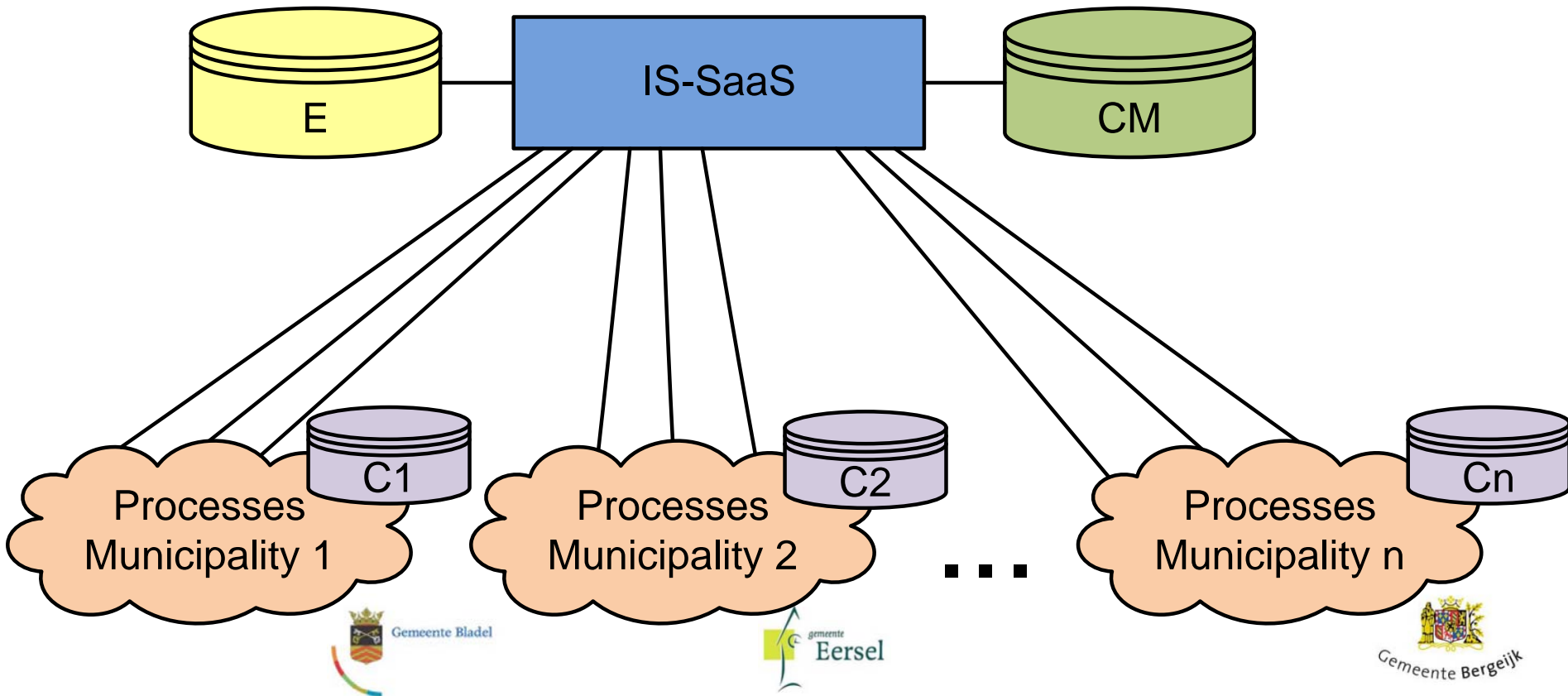
Example

Acknowledgement of an Unborn Child

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



Using SaaS Technology



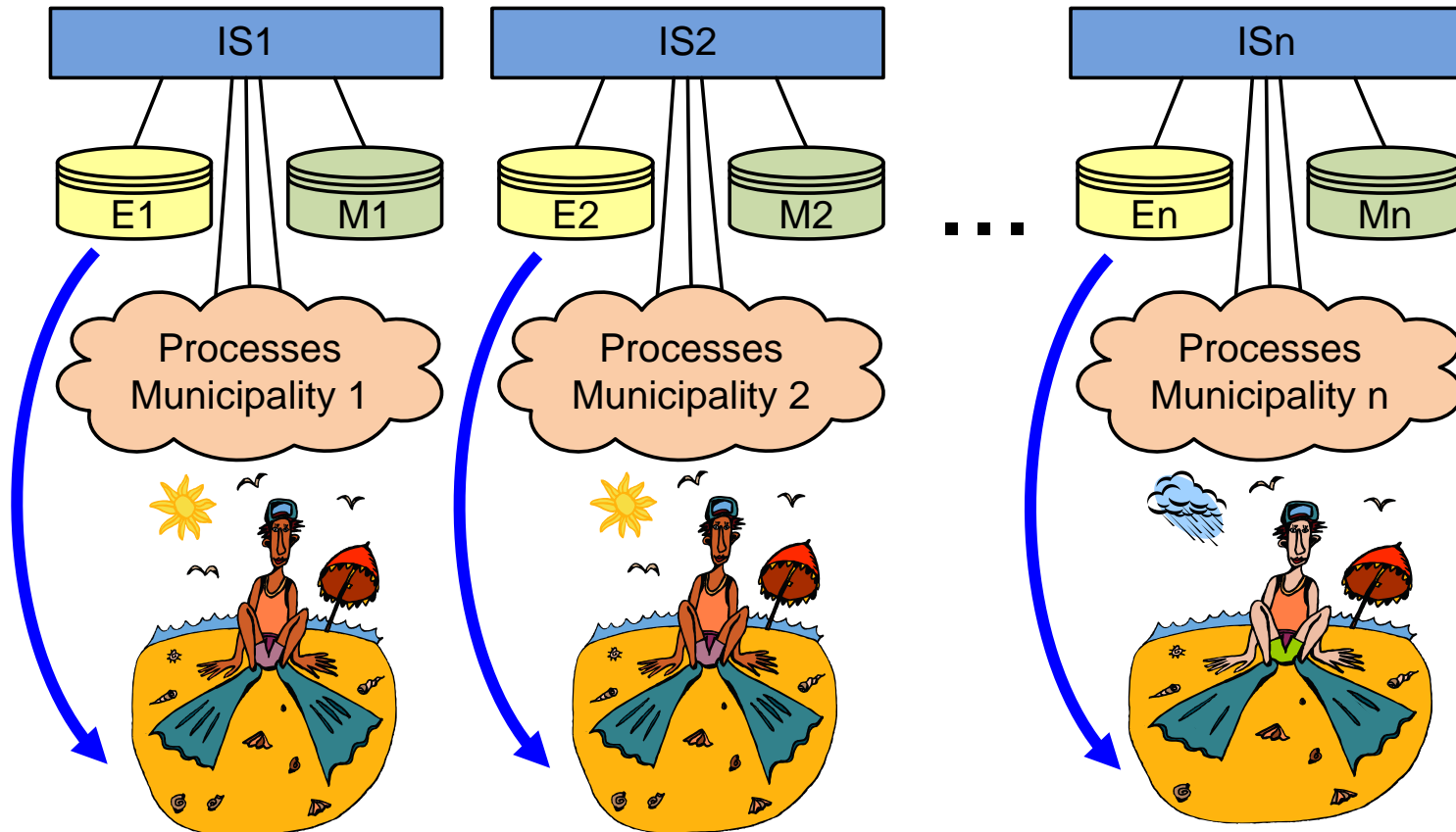
IS-SaaS = Information System (using a SaaS-based BPMS)

E = Event log

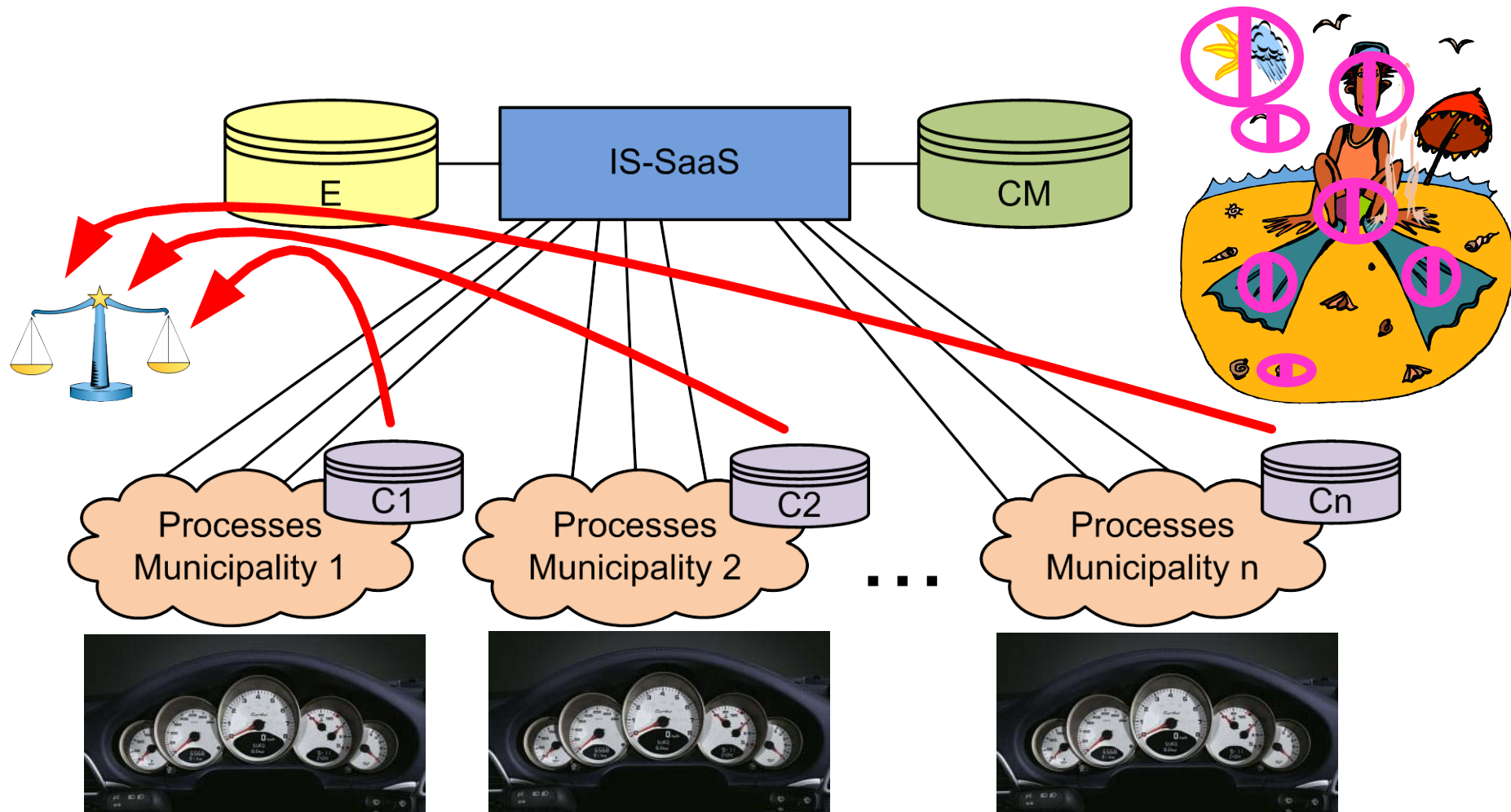
CM = Configurable Models

C = Configuration

Process Mining: Before



Process Mining: After



cross-organizational process mining



Configuration

Positioning of Configuration

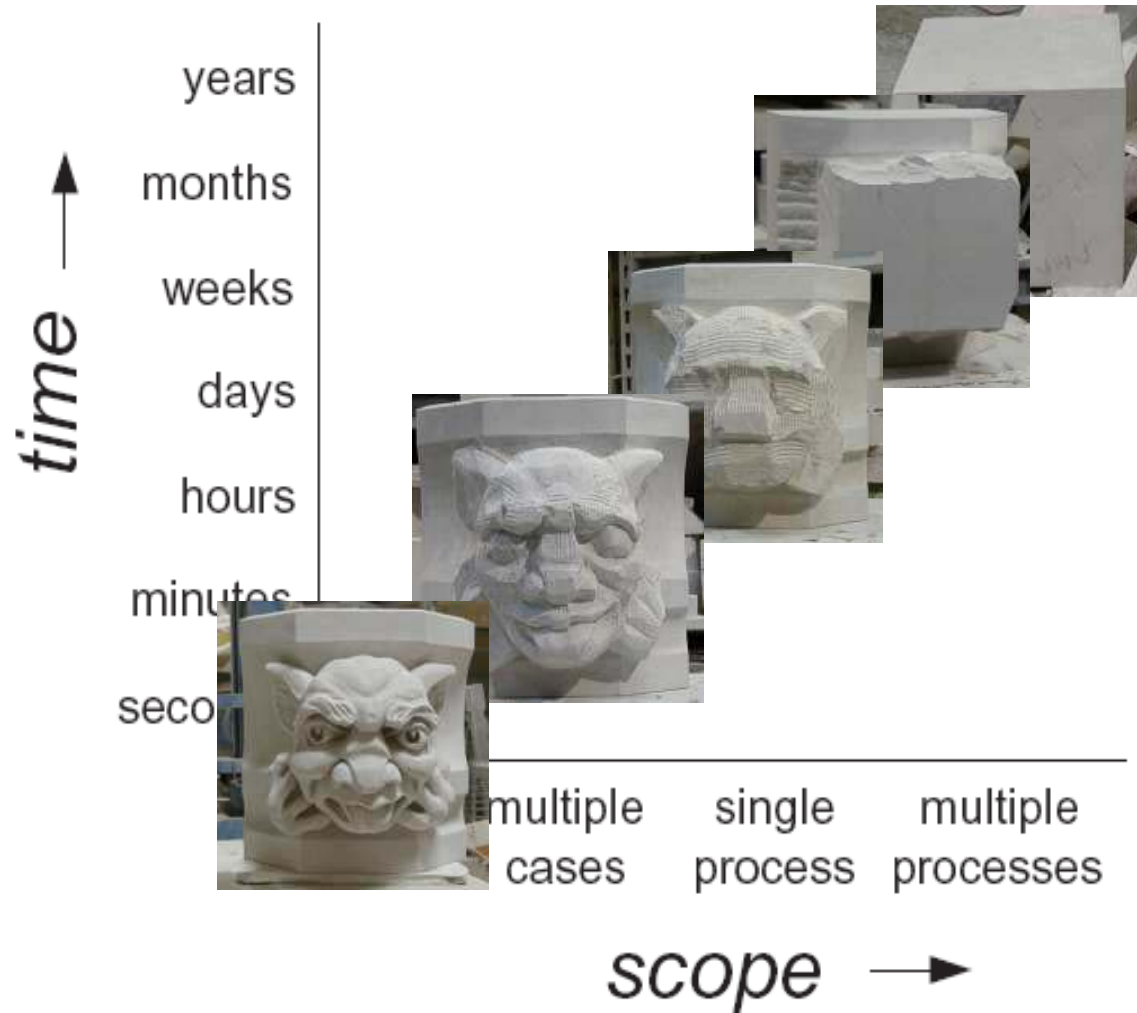
Some quotes from Michelangelo

- “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 ...

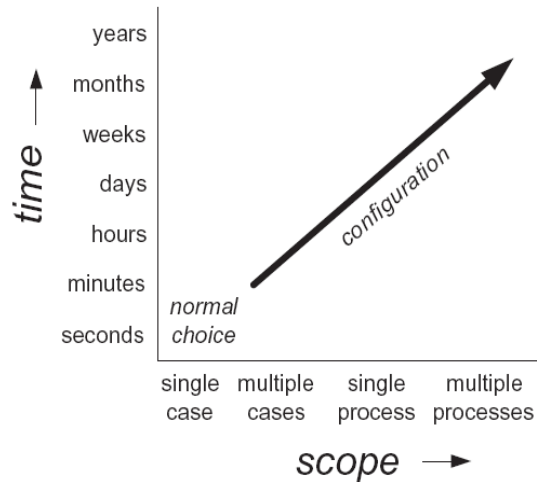


Time and artifacts

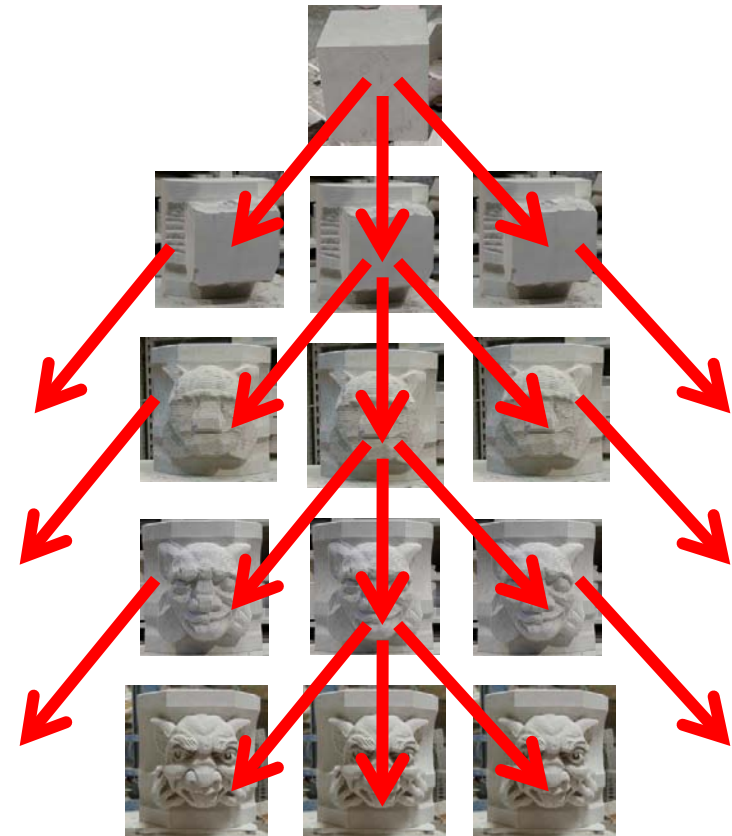
- Design time (**generic model**, i.e., is not released for instantiation)
- Configuration time (**specific model**, i.e., can be instantiated)
- Instantiation time (**specific model + instance**)
- Run time (**specific model + instance + state/partial trace**)
- Auditing time (**specific model + instance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



Continuum

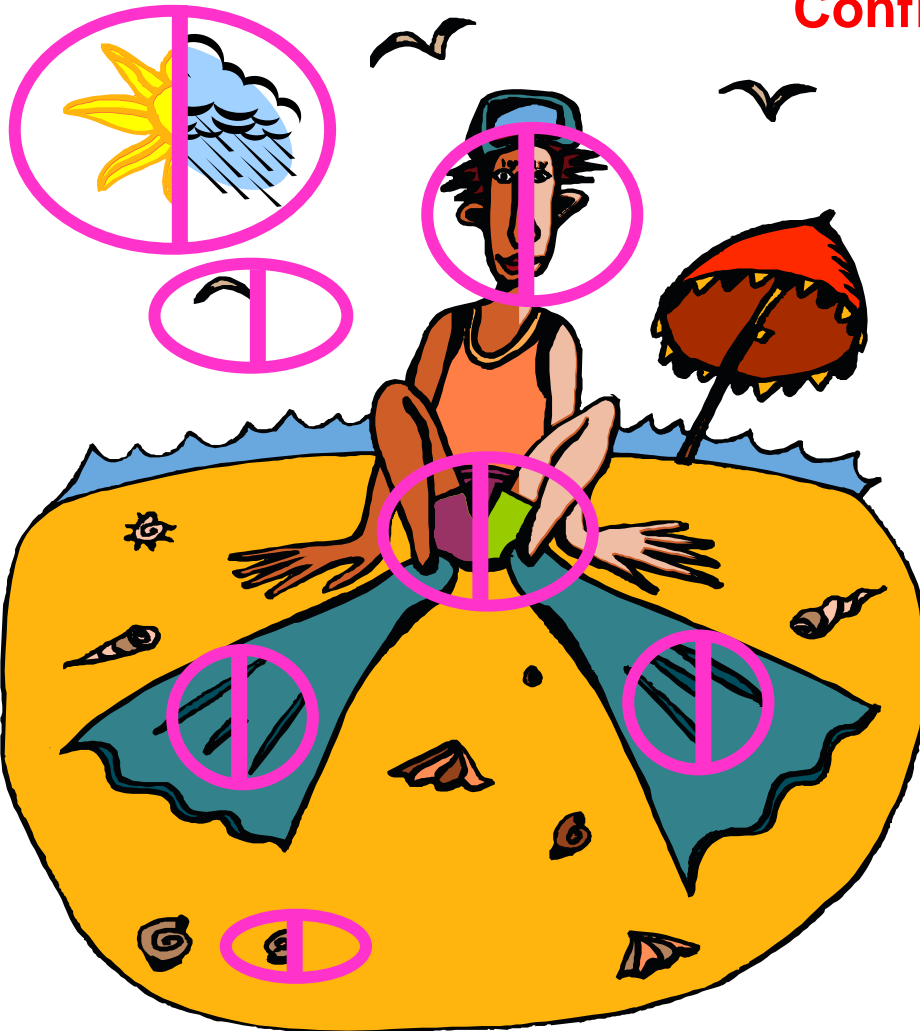


- In The Netherlands, ...
- In Brisbane, ...
- When the sun shines, ...
- On Sunday, ...
- When very busy, ...
- For these customers, ...
- ...



Hiding and blocking

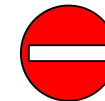
Configuration = limiting behavior !



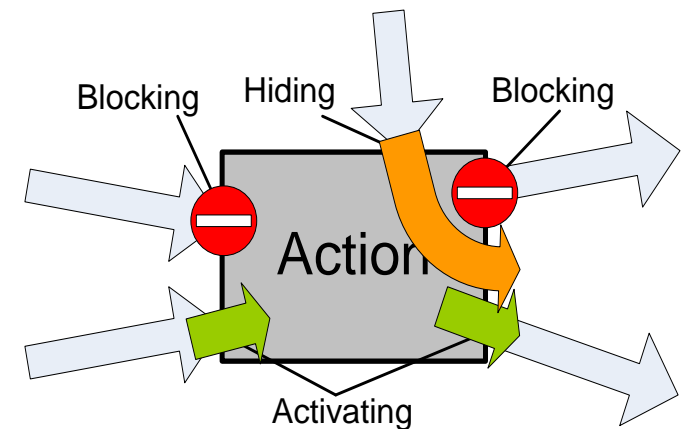
Activate



Hide/skip

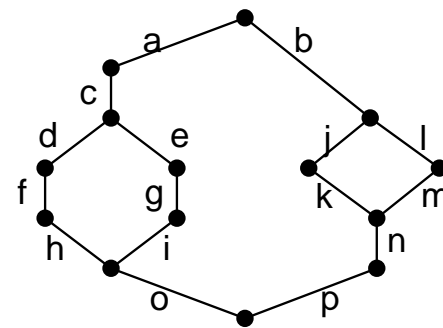
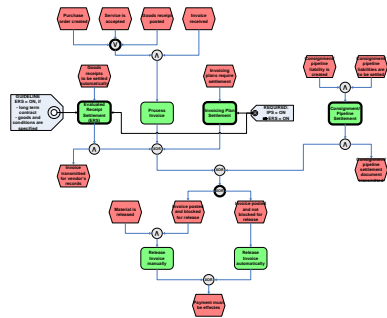


Block

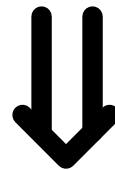


Configurable Process Models

C-EPC
C-Petri Net
C-YAWL
C-BPEL

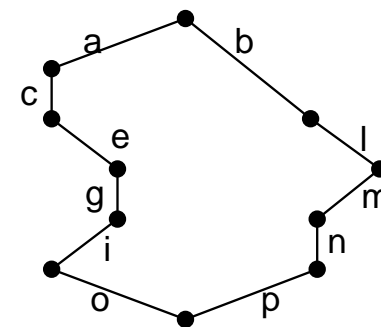
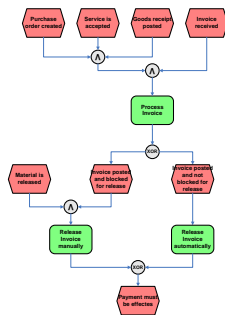


C-LTS



Configuration

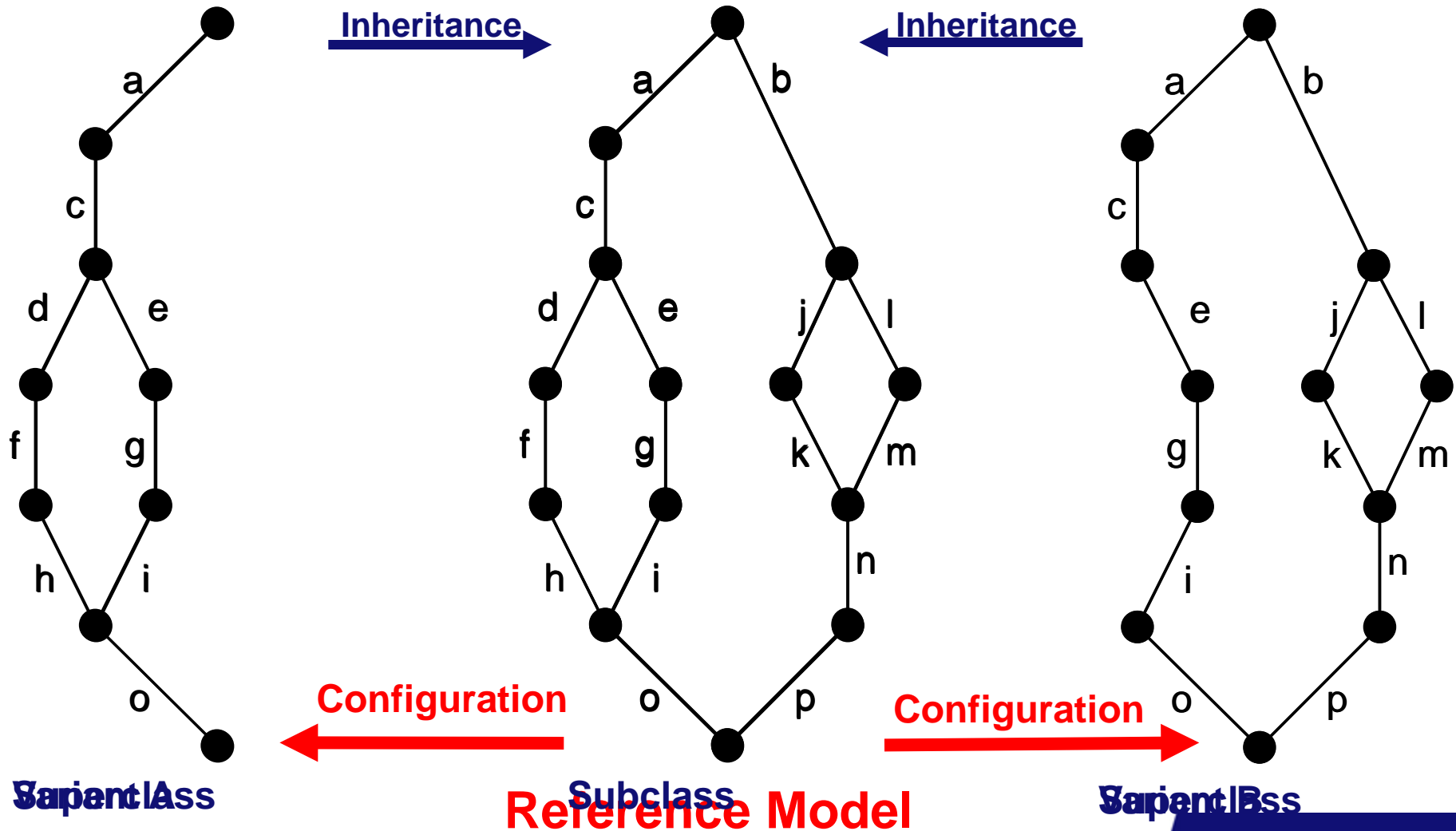
EPC
Petri Net
YAWL
BPEL





LTS

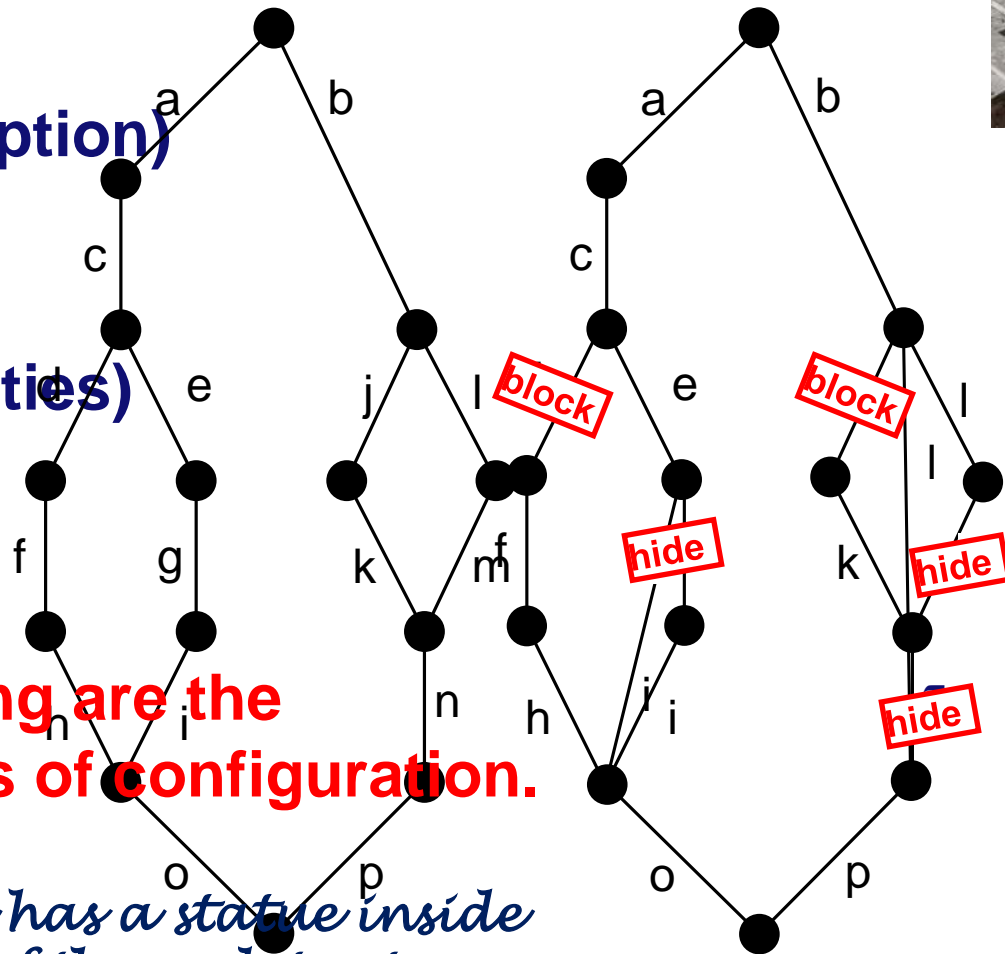
Blocking Hiding

Inheritance of dynamic behavior



Configuration Techniques

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



Blocking and hiding are the essential concepts of configuration.

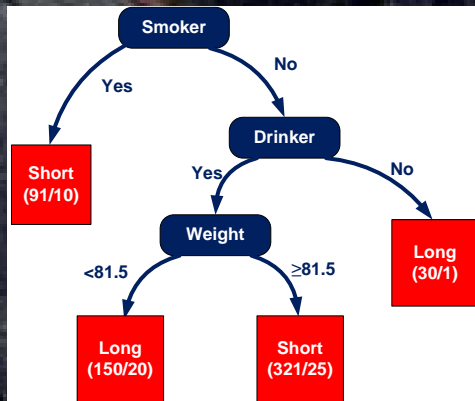
“Every block of stone has a statue inside it and it is the task of the sculptor to discover it.”



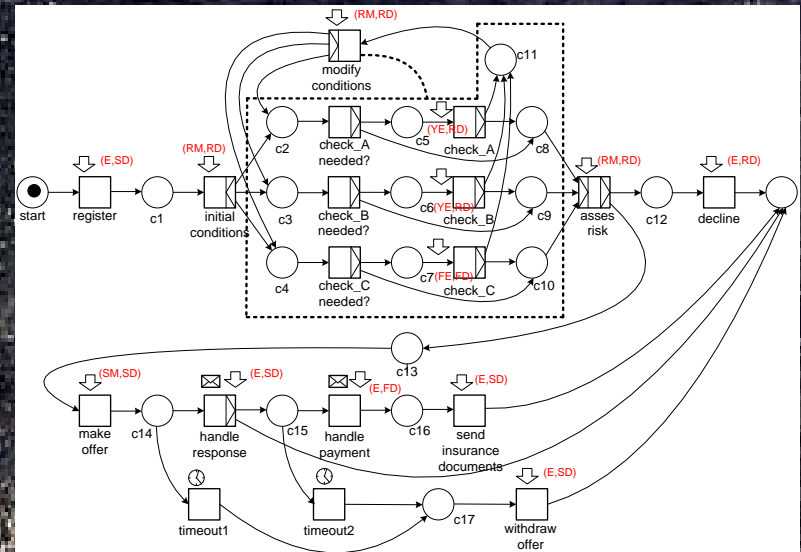


Cross-organizational mining

Process Mining =



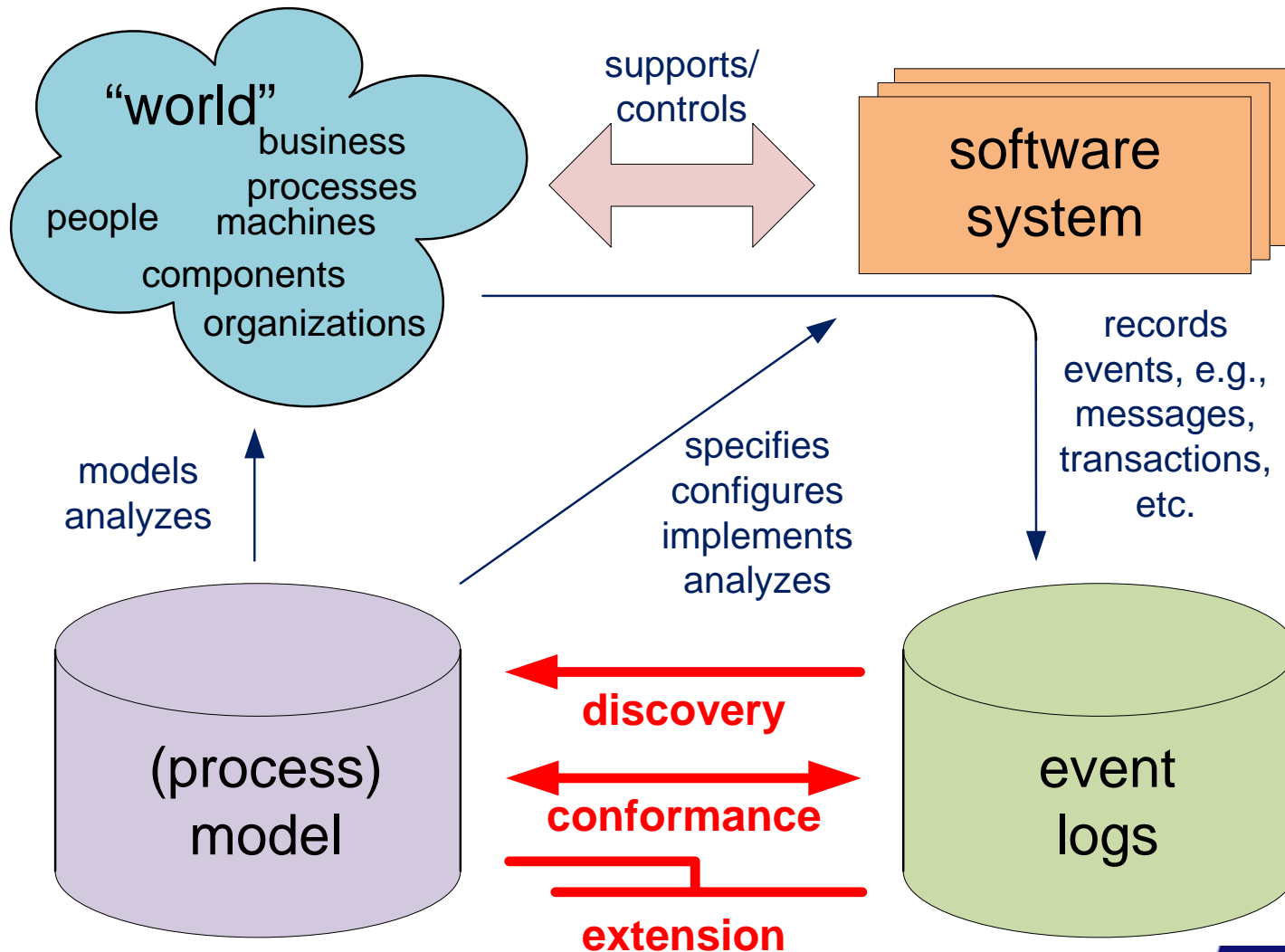
+



Data Mining

Process Analysis

Process mining: Linking events to models

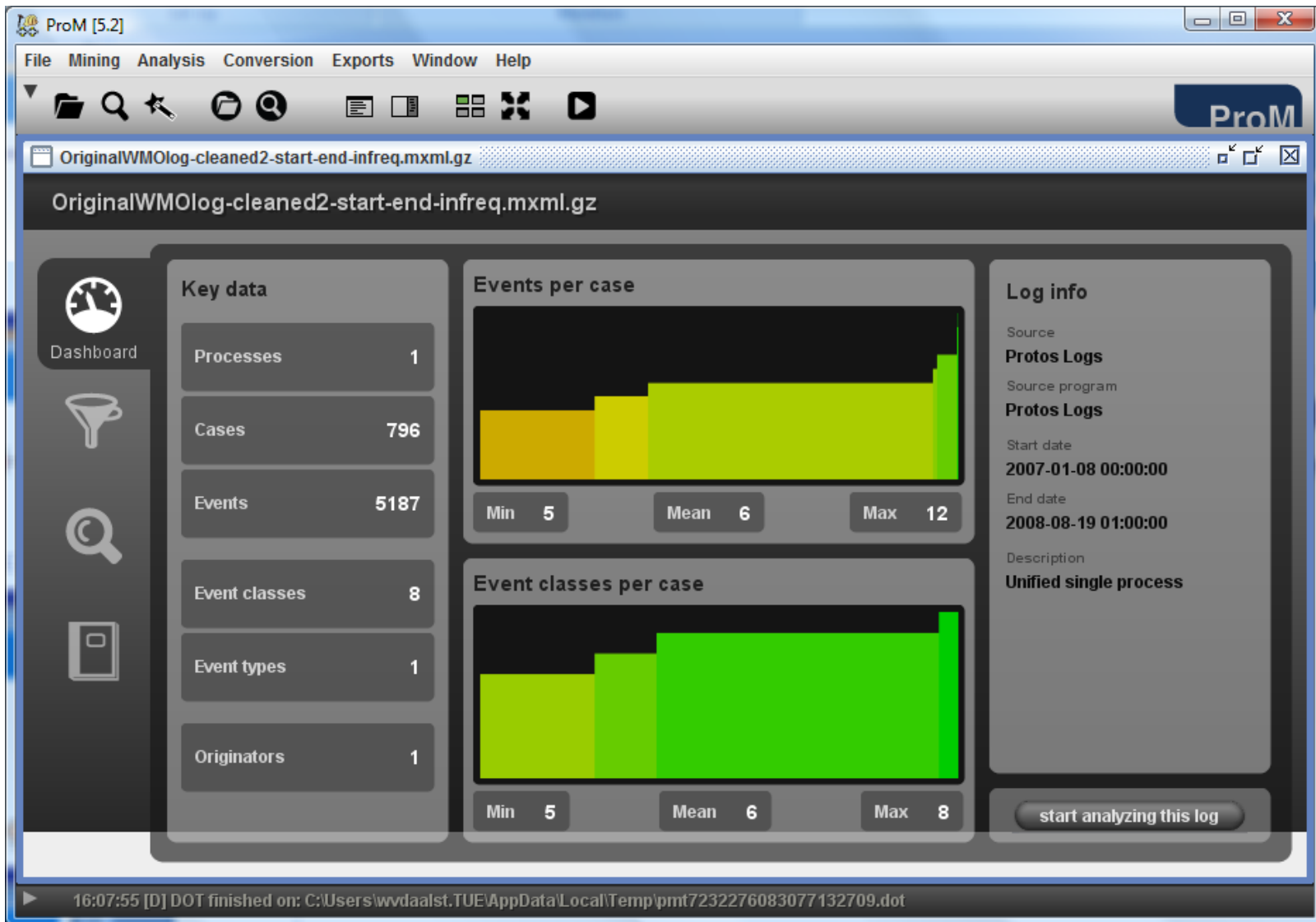


Example: WMO Harderwijk

- Process related to the execution of “Wet Maatschappelijke Ondersteuning” (WMO) Harderwijk
- Handling WMO applications
- WMO: supporting citizens of municipalities (illness, handicaps, elderly, etc.).
- Examples:
 - wheelchair, scootmobiel, ...
 - adaptation of house (elevator), ...
 - household help, ...

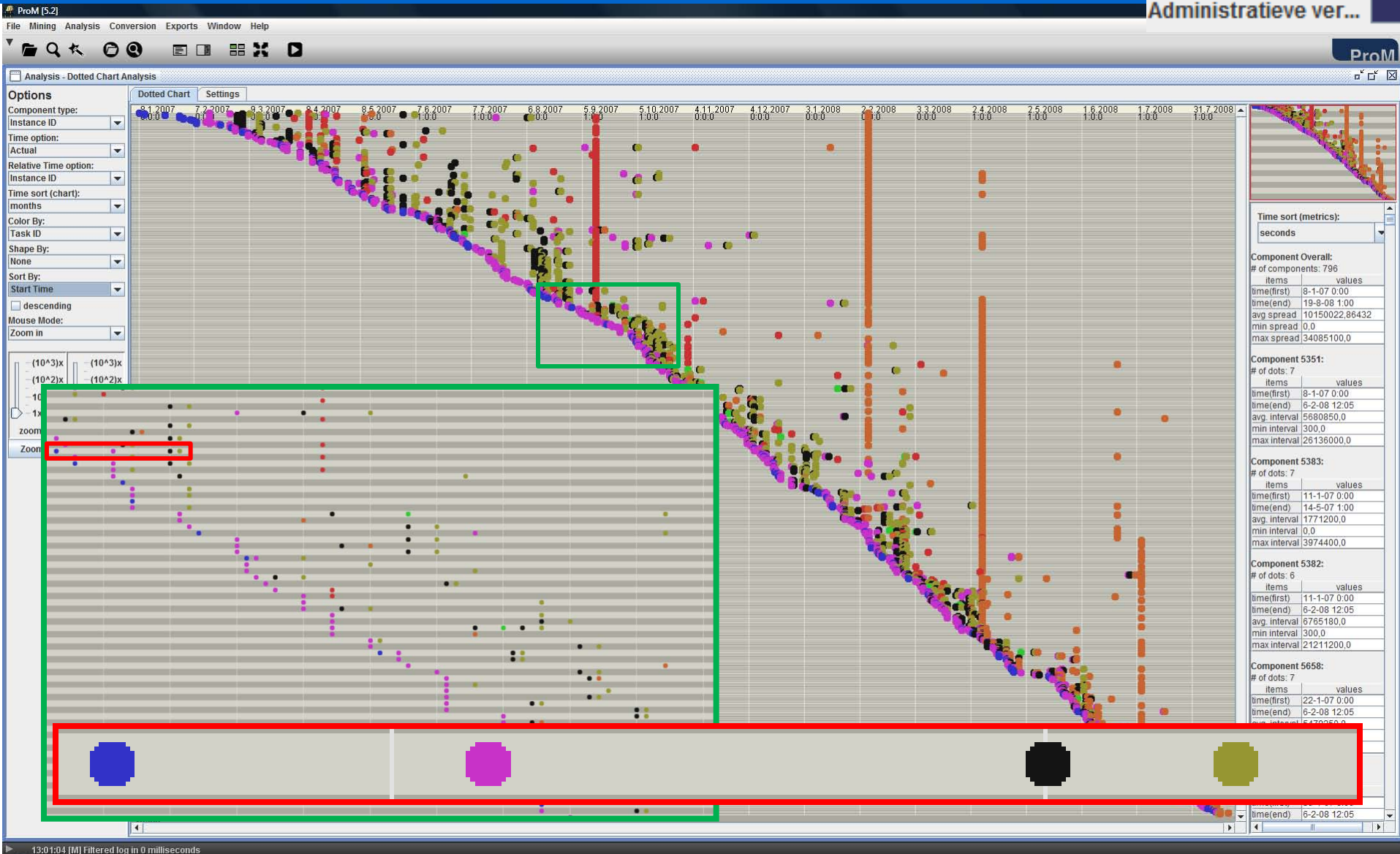
Event log

(796 applications, 5187 events)



Helicopter view of 1.5 years

Wachten terugmeldi...
Retour:
Aanvraag registratie:
Slotfase:
Rapportage & besch...
Verzending\dossierv...
Toetsen en beslissen:
Administratieve ver...



Process discovered using Genetic Miner

The screenshot displays the ProM 5.2 software interface. The main window is titled "Settings for mining Raw OriginalWMOlog-cleaned2-start-end-infreq.mxml.gz (unfiltered) using Genetic algorithm plugin". The interface is divided into two main sections: "Genetic algorithm plugin" on the left and "Results - Genetic algorithm plugin on Raw OriginalWMOlog-cleaned2-start-end-infreq.mxml.gz (unfiltered)" on the right.

Genetic algorithm plugin settings:

- Population size: 100
- Initial population type: No Duplicates
- Max number generations: 1000
- Seed: 1
- Power value: 1
- Elitism rate: 0.02
- Fitness type: ExtraBehaviorPunishment
- ☐ Show Advanced Fitness Parameters
- ☒ Use genetic operators
 - Selection method type: Tournament 5
 - Crossover type: Enhanced
 - Crossover rate: 0.8
 - Mutation type: Enhanced
 - Mutation rate: 0.2

Results - Genetic algorithm plugin on Raw OriginalWMOlog-cleaned2-start-end-infreq.mxml.gz (unfiltered):

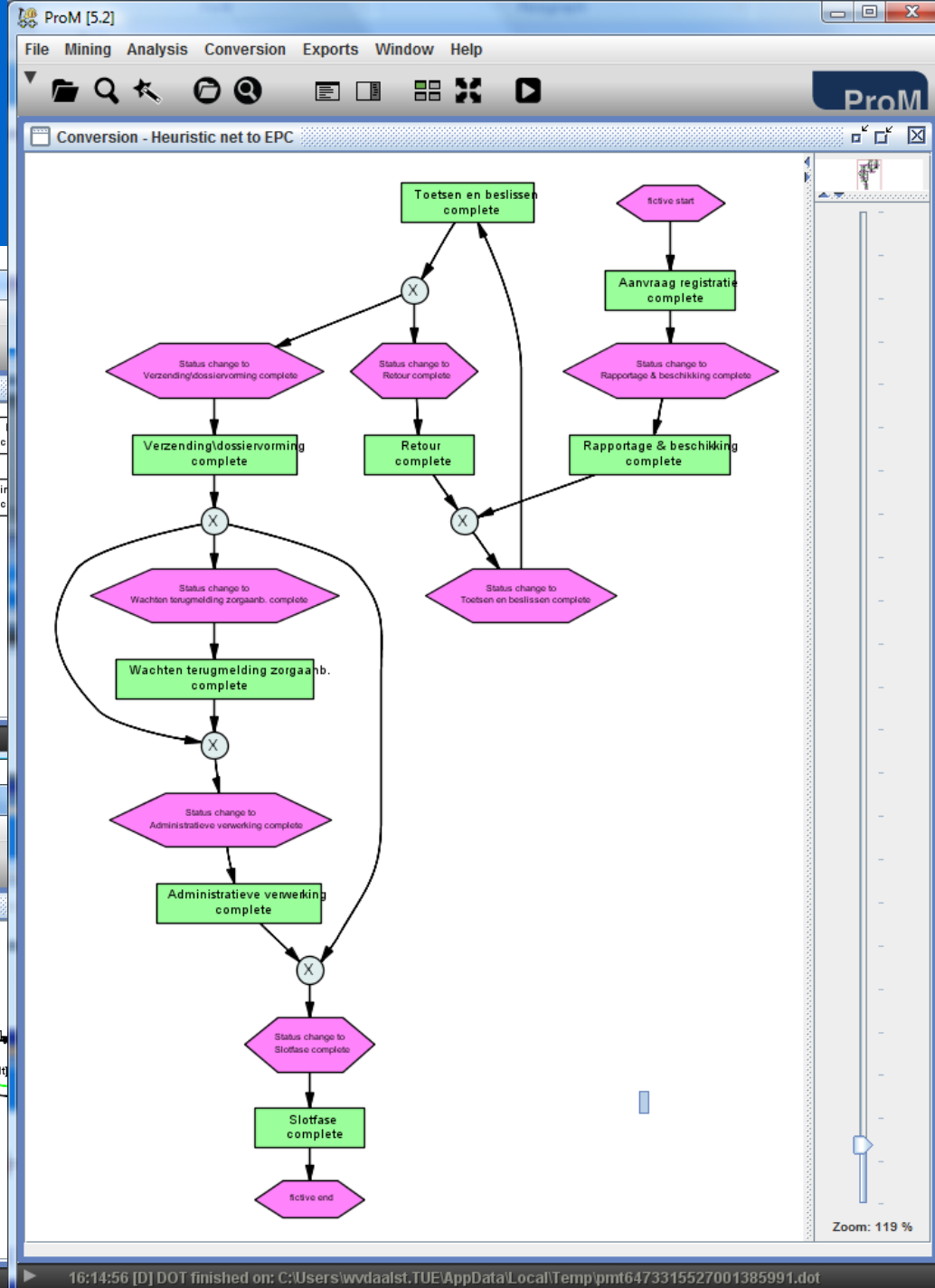
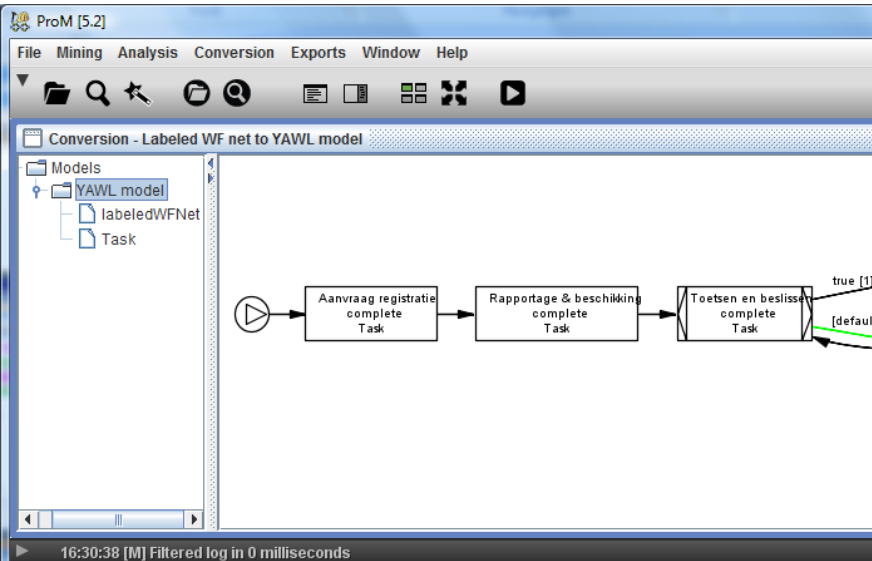
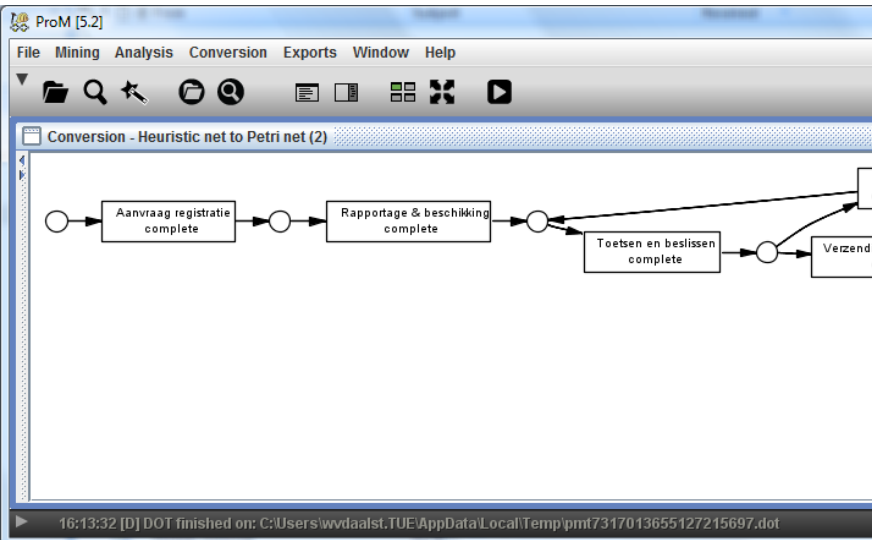
The results section displays a list of 31 individuals, each with a fitness value. The fitness values range from approximately 0.79 to 0.97. The list is as follows:

Individual	Fitness
Individual 0	0.97888213013541...
Individual 1	0.97856574393486...
Individual 2	0.97855591766334...
Individual 3	0.97845372443954...
Individual 4	0.97825748124269...
Individual 5	0.97792505128941...
Individual 6	0.97726671083998...
Individual 7	0.97487539252676...
Individual 8	0.974551407800164...
Individual 9	0.96802919564024...
Individual 10	0.9616994729187...
Individual 11	0.9607247067839...
Individual 12	0.9600781381179...
Individual 13	0.9565621912744...
Individual 14	0.9411785034574...
Individual 15	0.9371611562464...
Individual 16	0.8835435379947...
Individual 17	0.8818051419910...
Individual 18	0.8818051419910...
Individual 19	0.8668926653805...
Individual 20	0.8666966723630...
Individual 21	0.8651239365069...
Individual 22	0.8649025041963...
Individual 23	0.8638451168601...
Individual 24	0.8188783435011...
Individual 25	0.8180922417795...
Individual 26	0.8179035773663...
Individual 27	0.8169288112316...
Individual 28	0.8134705478908...
Individual 29	0.8125026690498...
Individual 30	0.8050562147846...
Individual 31	0.7949839860248...

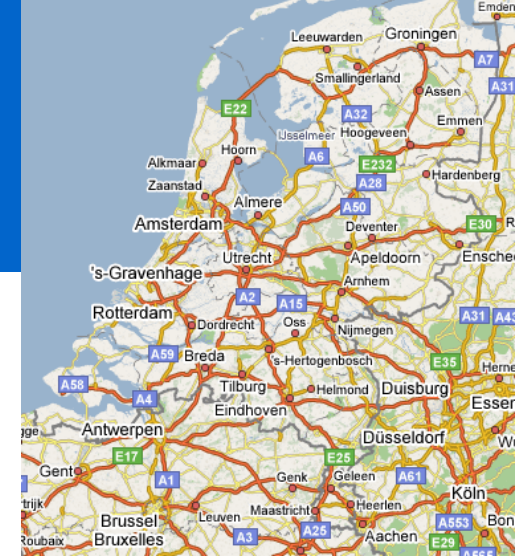
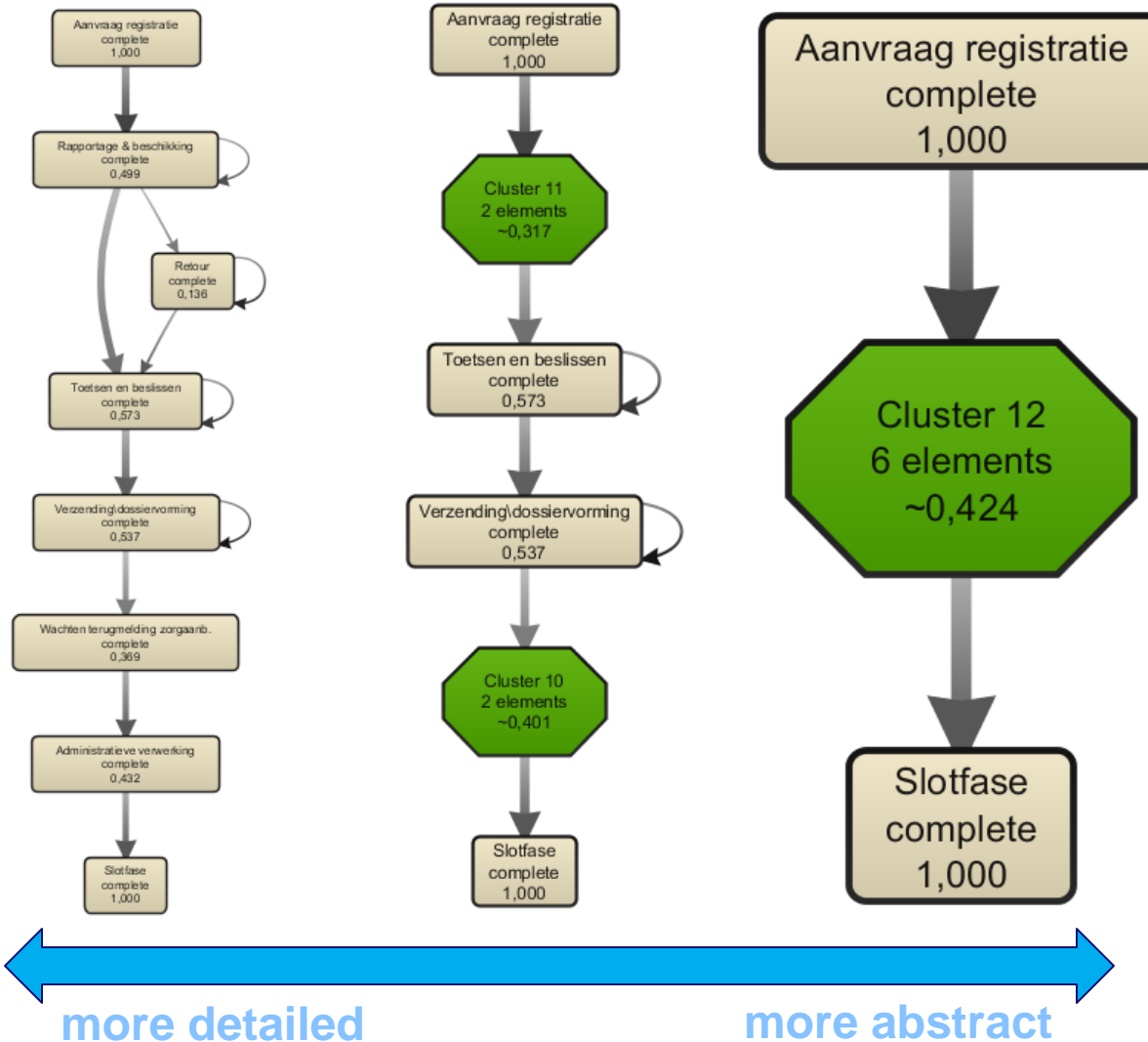
On the right side of the results section, a process diagram is displayed. The diagram shows a sequence of tasks connected by XOR gates. The tasks are: "Aanvraag registratie (complete) 796", "Populatie & beschikking (complete) 801", "Toelaten en besluiten (complete) 802", "Verzendingsaanvraag (complete) 799", "Wachten toelatingseisen (complete) 498", "Administratieve verwerking (complete) 593", and "Skafwissel (complete) 798". The diagram illustrates the flow of the process, with tasks being connected by XOR gates, indicating a choice or parallel execution.

At the bottom of the window, a status bar shows the time "16:07:55 [D]" and the file path "DOT finished on: C:\Users\wvdaalst.TUE\AppData\Local\Temp\pmt7232276083077132709.dot".

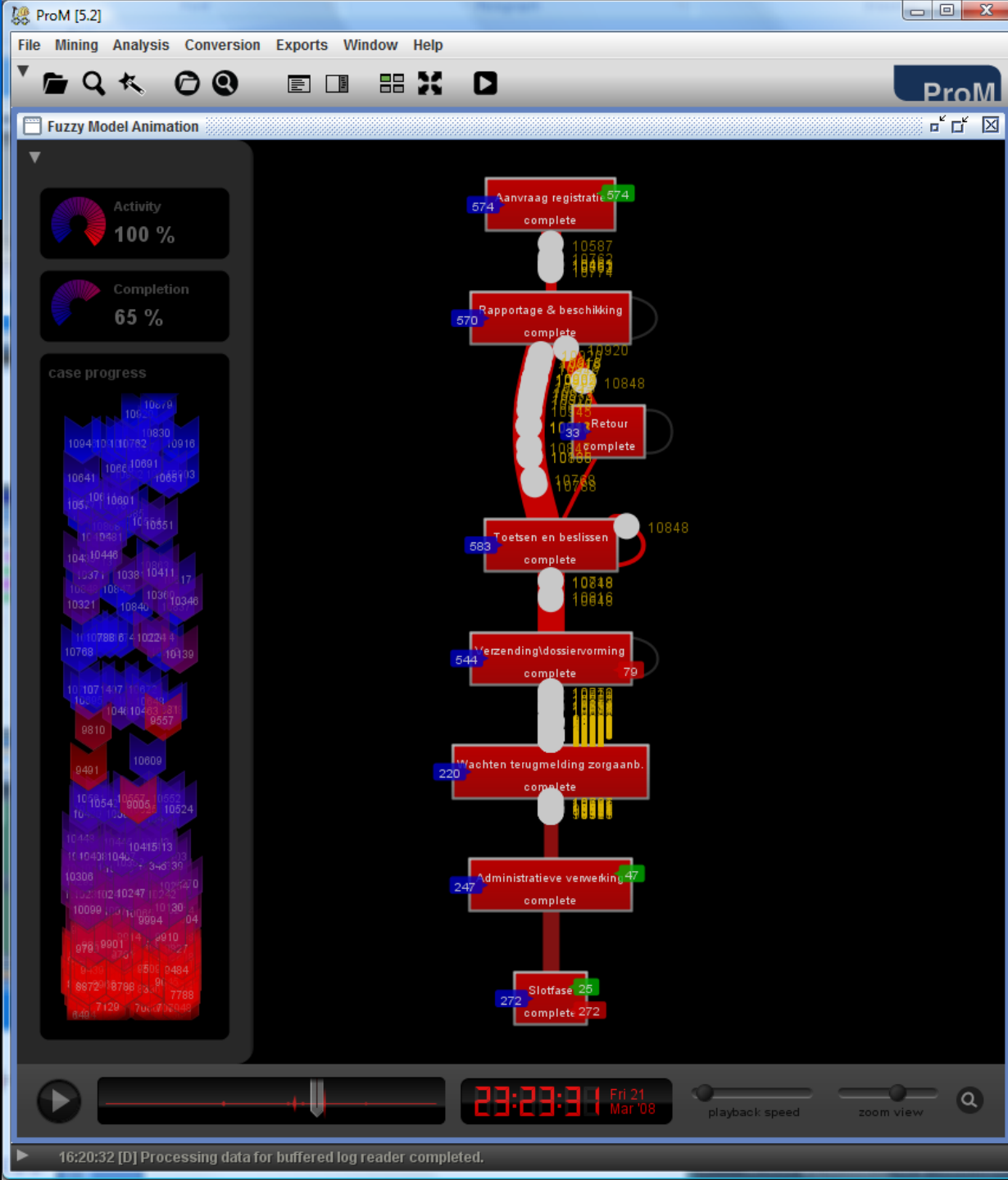
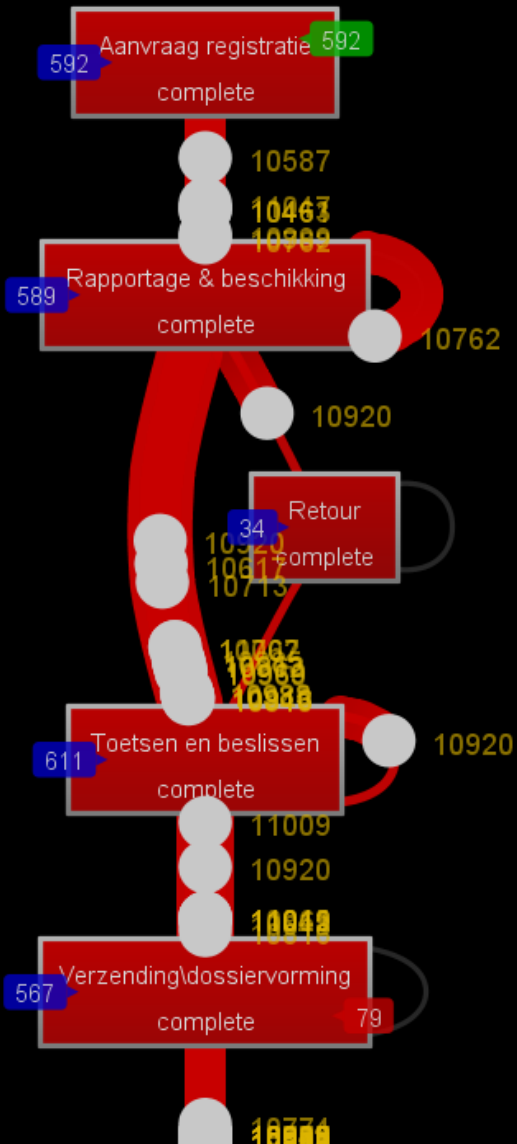
Various representations



Seamless abstraction



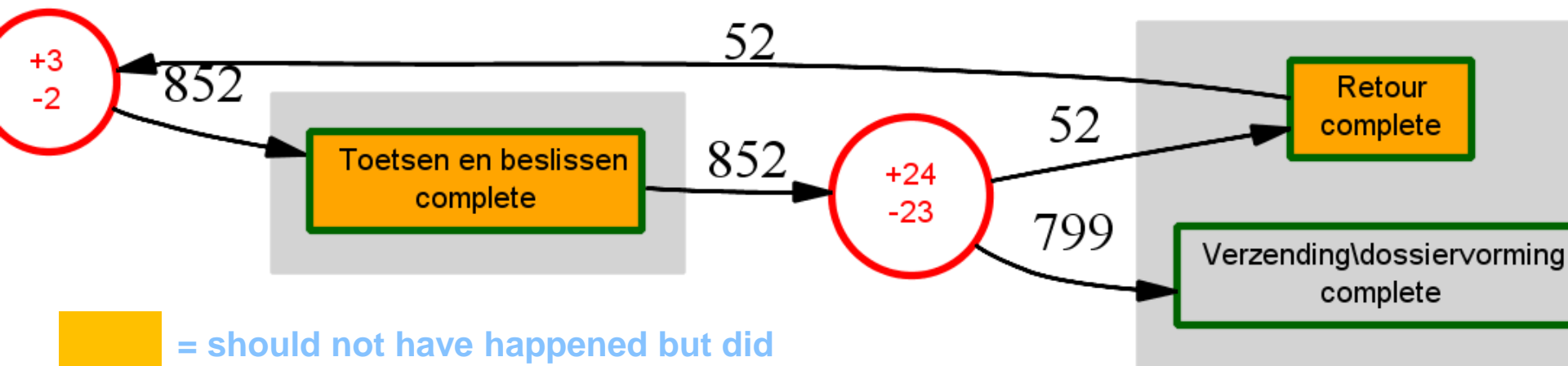
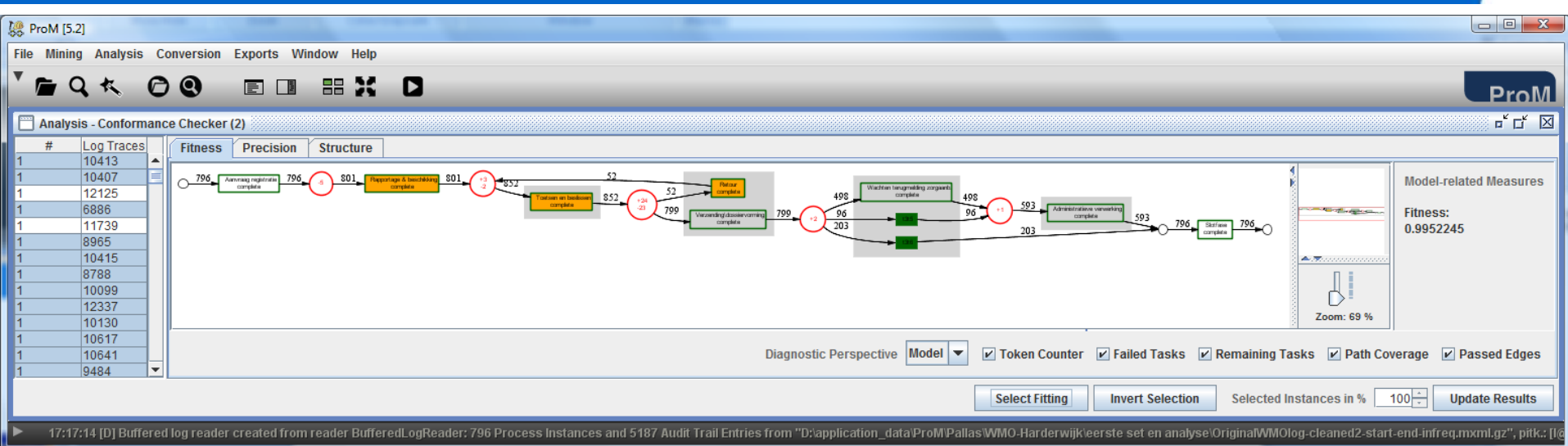
Fuzzy Replay



Conformance checking using Replay

Model-related Measures

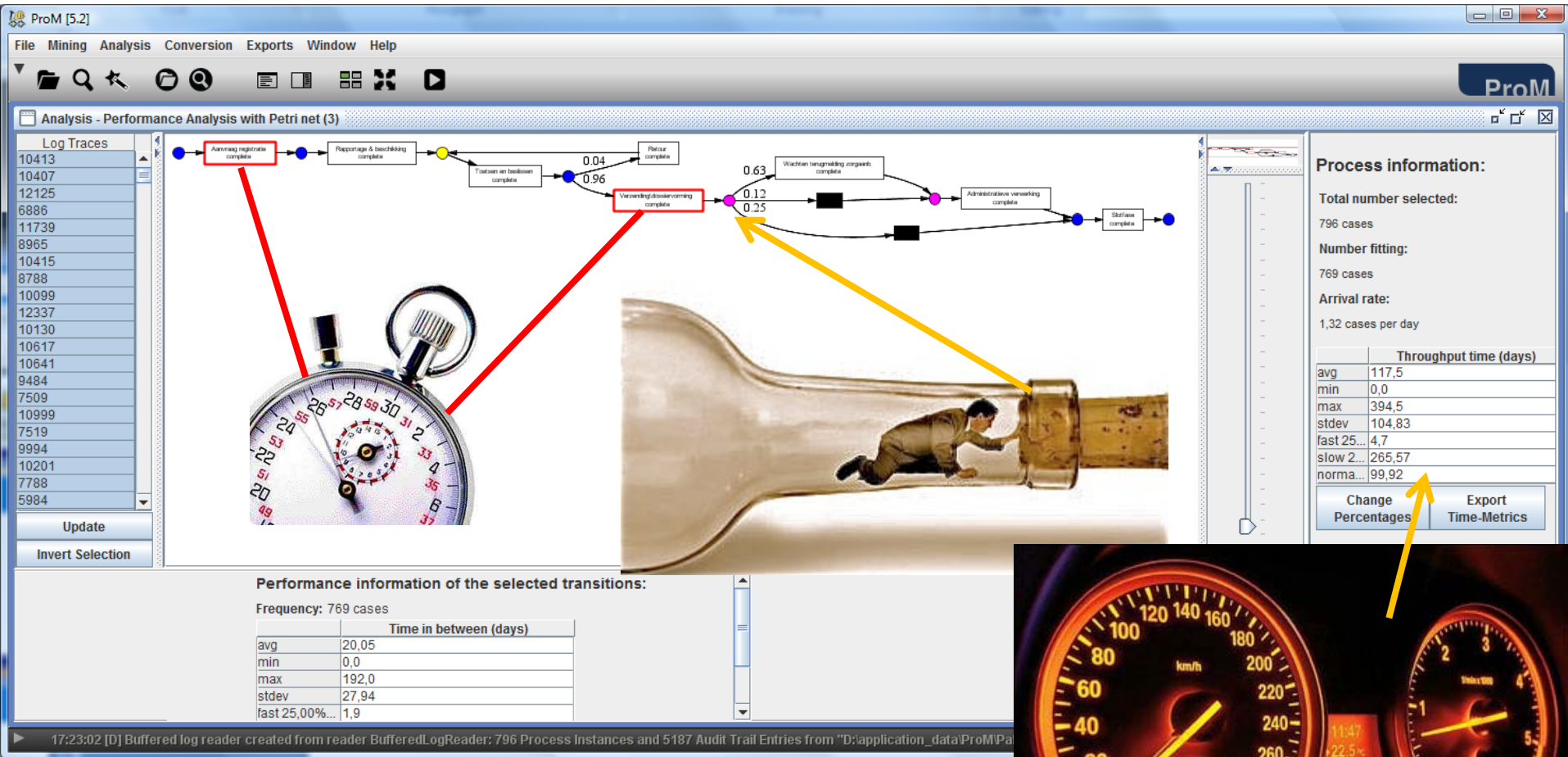
Fitness:
0.9952245



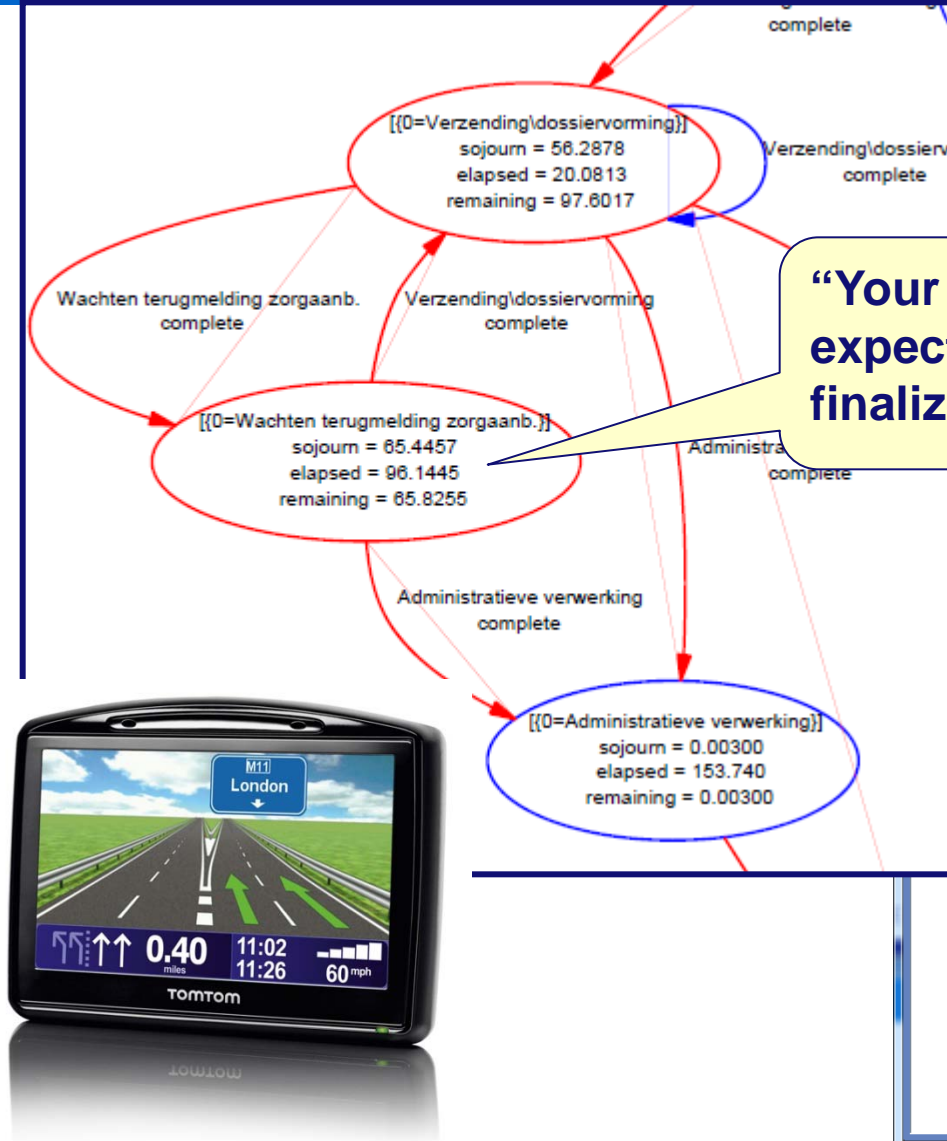
= should not have happened but did

= should have happened but did not

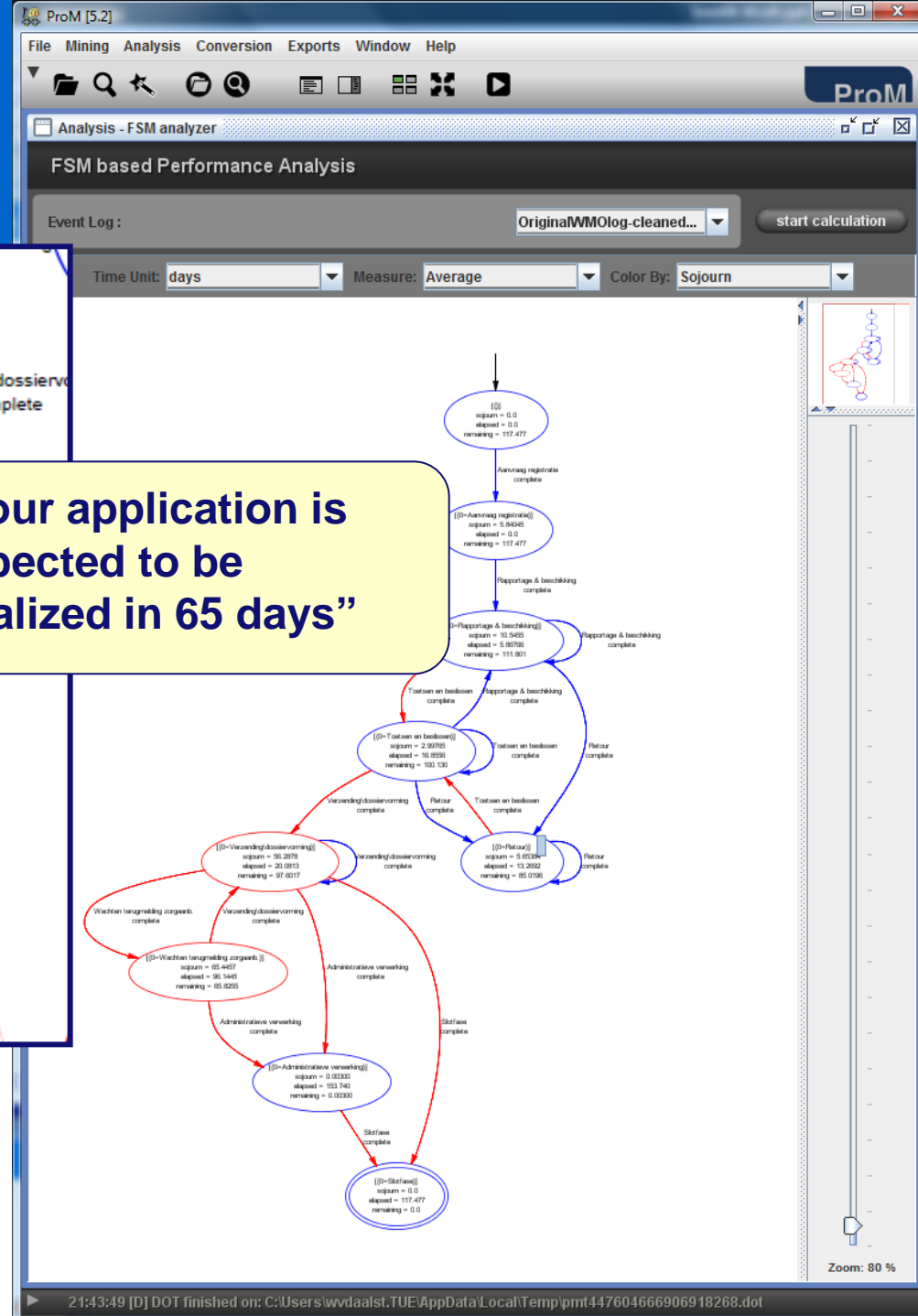
Performance analysis using Replay



Prediction based on Replay



“Your application is expected to be finalized in 65 days”

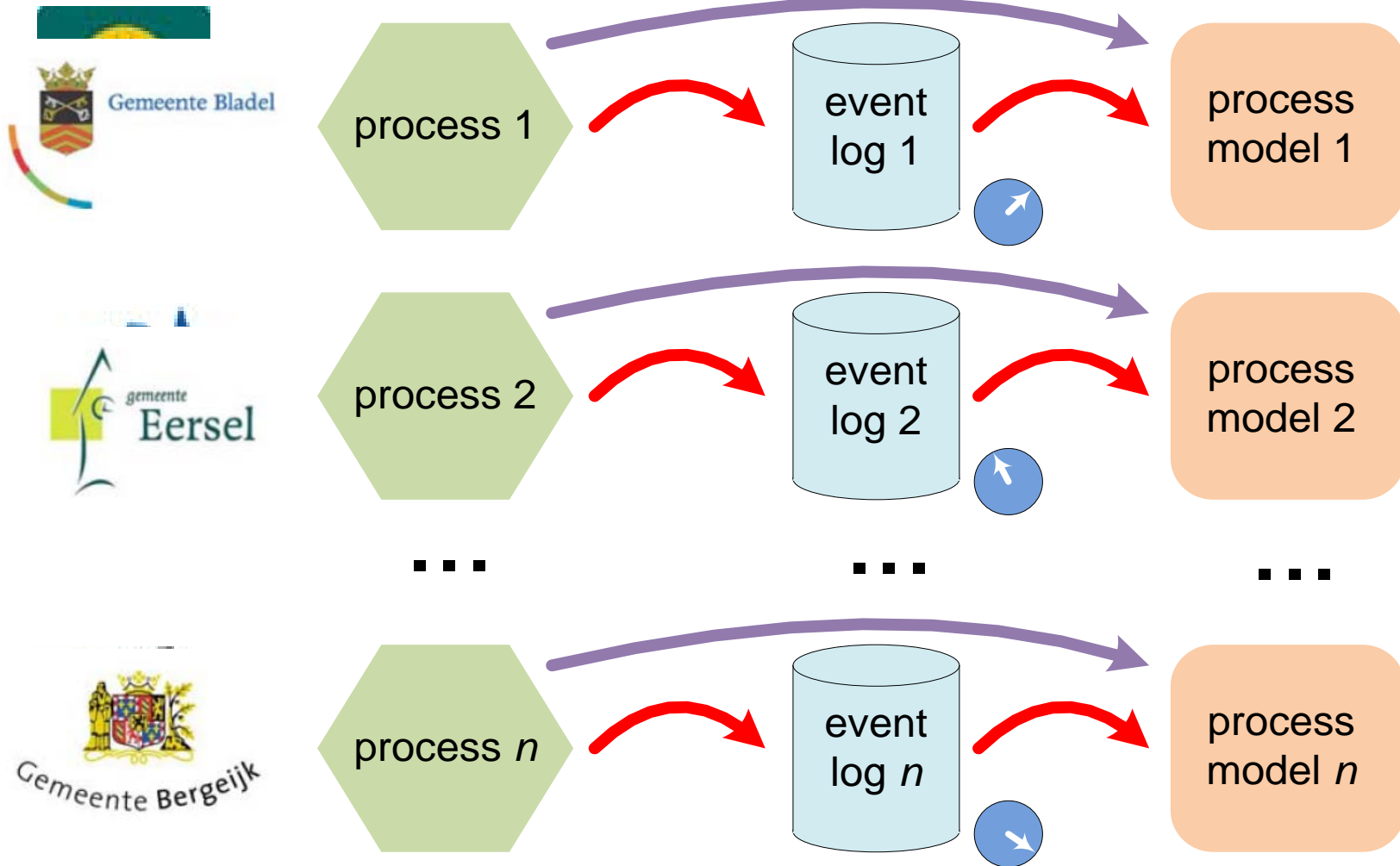


From one to many organizations

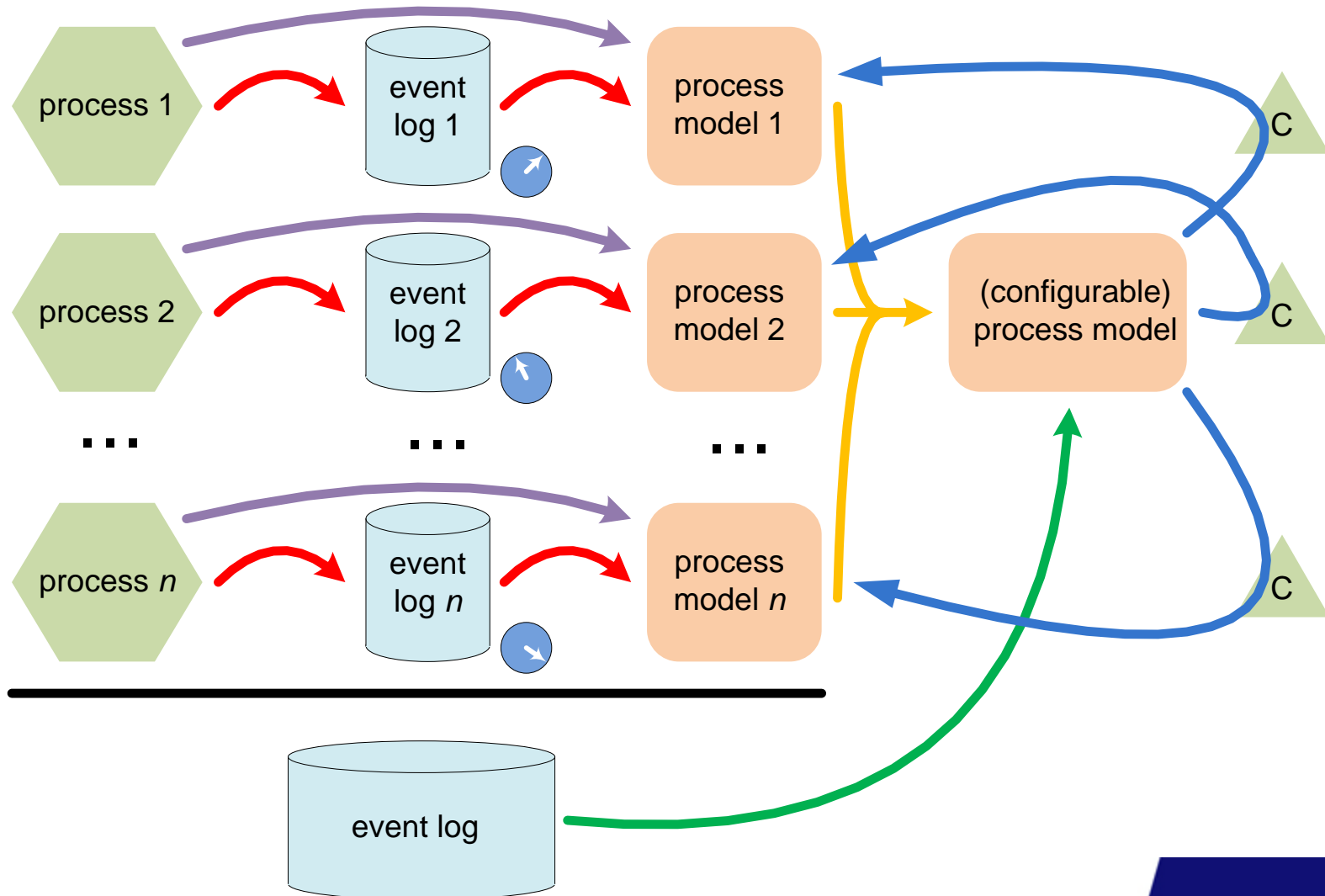
- More than 80,000 organizations are using Salesforce
- More than 1 million organizations are using Google Apps
- All 430 Dutch municipalities are implementing the same set of processes
- All 94 U.S. District Courts in the United States share the same set of workflows
- All car-rental offices of Hertz,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



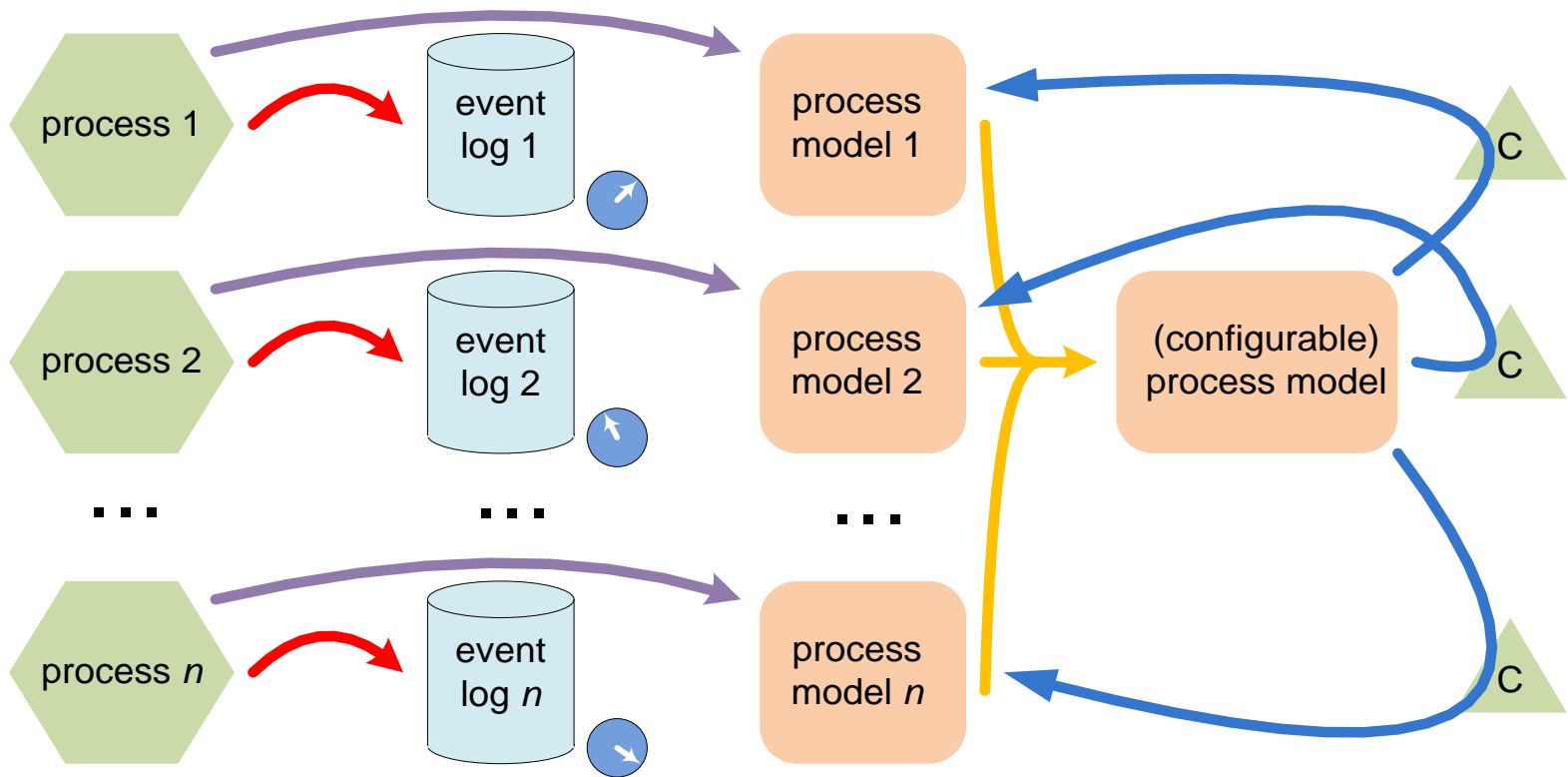
Consider n organizations



Cross-organizational process mining



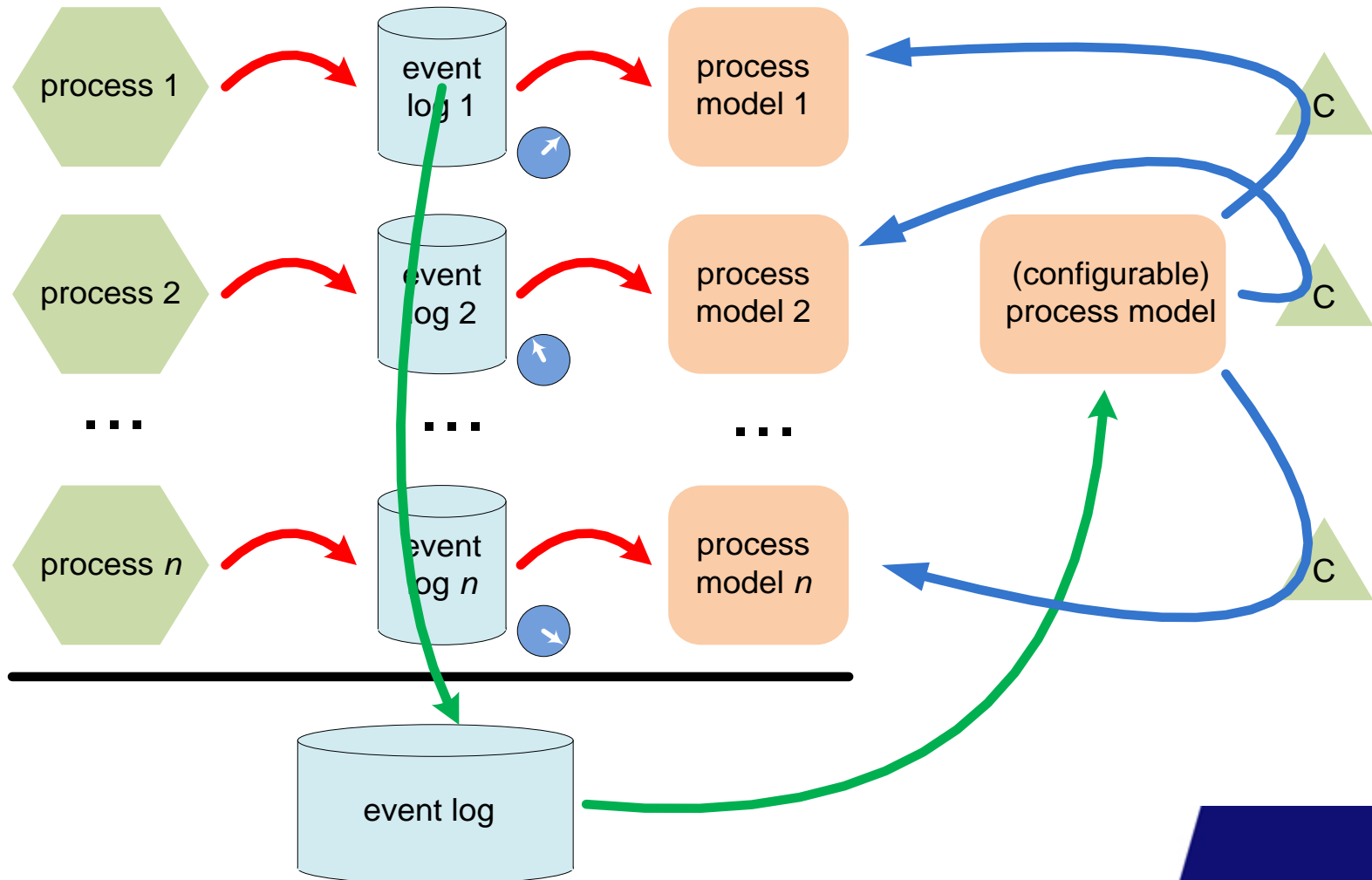
Pure model-based



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

Pure log-based

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



How to find and characterize differences among processes using event logs?

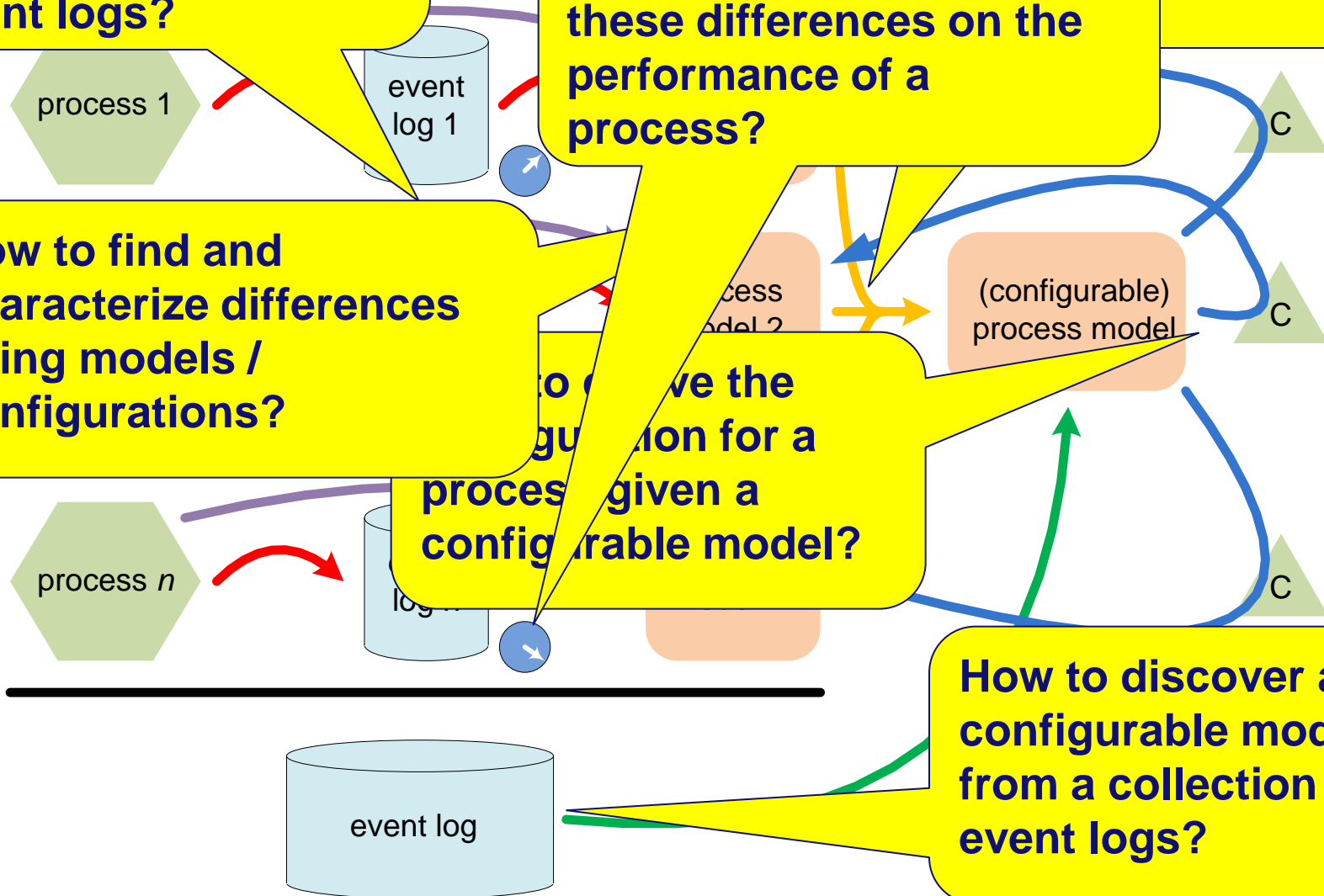
How to merge process models into a configurable model

What are the effects of these differences on the performance of a process?

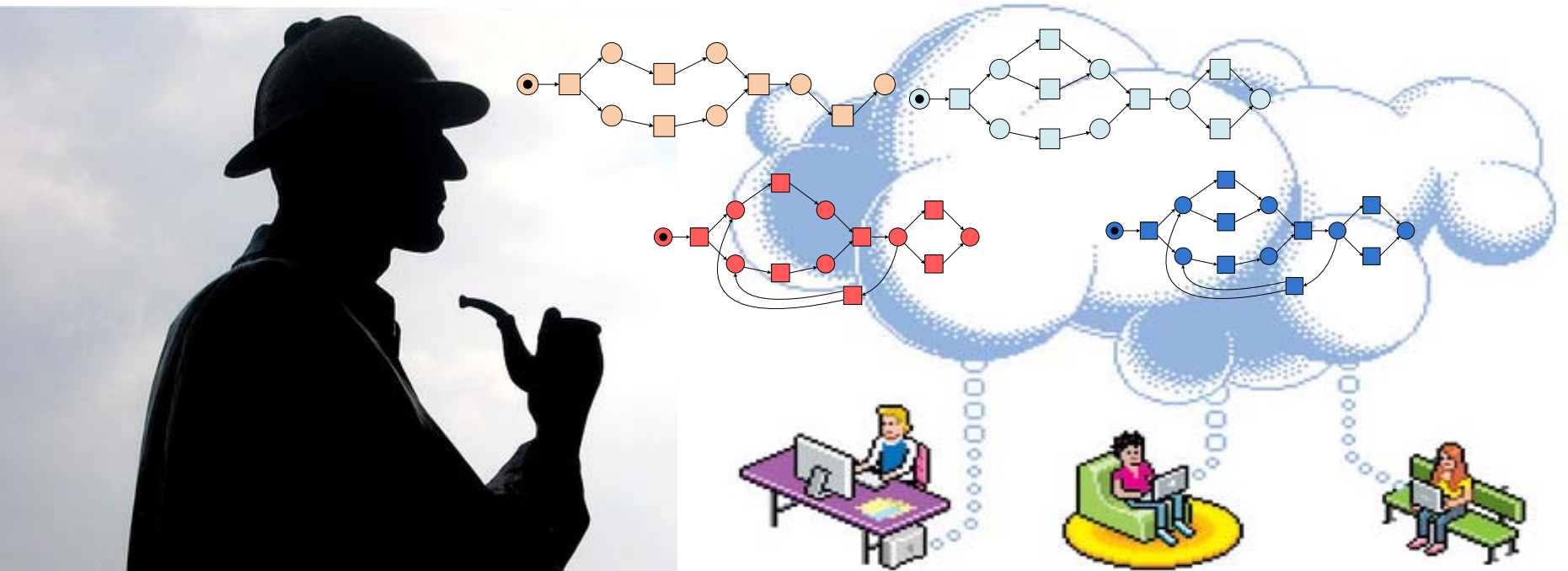
How to find and characterize differences using models / configurations?

How to derive the configuration for a process given a configurable model?

How to discover a configurable model from a collection of event logs?



Evidence-based “best practices”

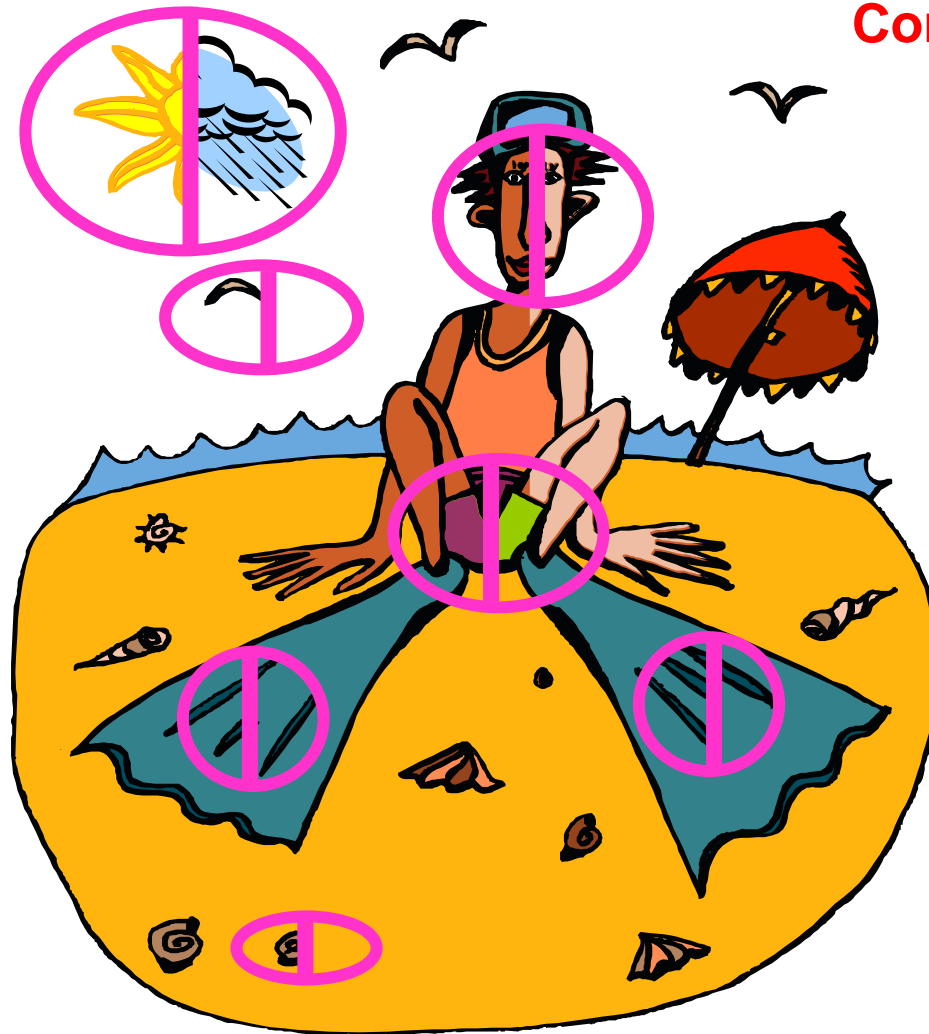


- Organizations can learn from each other.
- Configuration support and diagnostics.
- Software vendors/service providers can improve their products/services.

A black and white photograph of a Terracotta Warrior statue, partially buried in the ground. The statue is shown from the waist up, wearing a detailed scale-like armor. It is positioned in a trench, with its right arm bent and holding a sword. The background is a rough, cracked earthen wall. In the foreground, there are some broken fragments of pottery or stone. Overlaid on the lower half of the image is the word "Correctness" in a large, bold, yellow sans-serif font.

Correctness

Remember ...



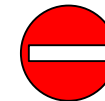
Configuration = limiting behavior !



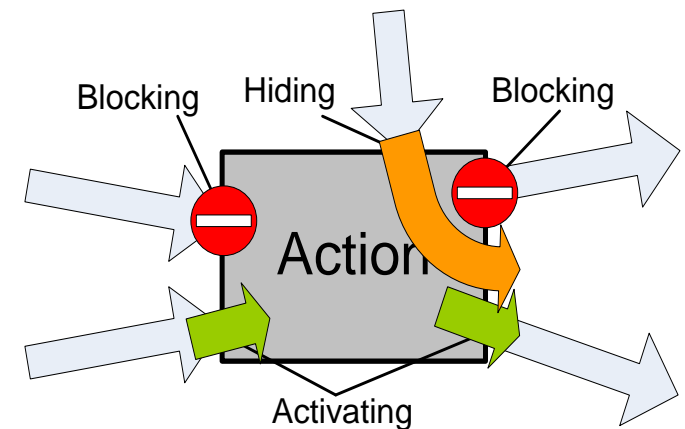
Activate



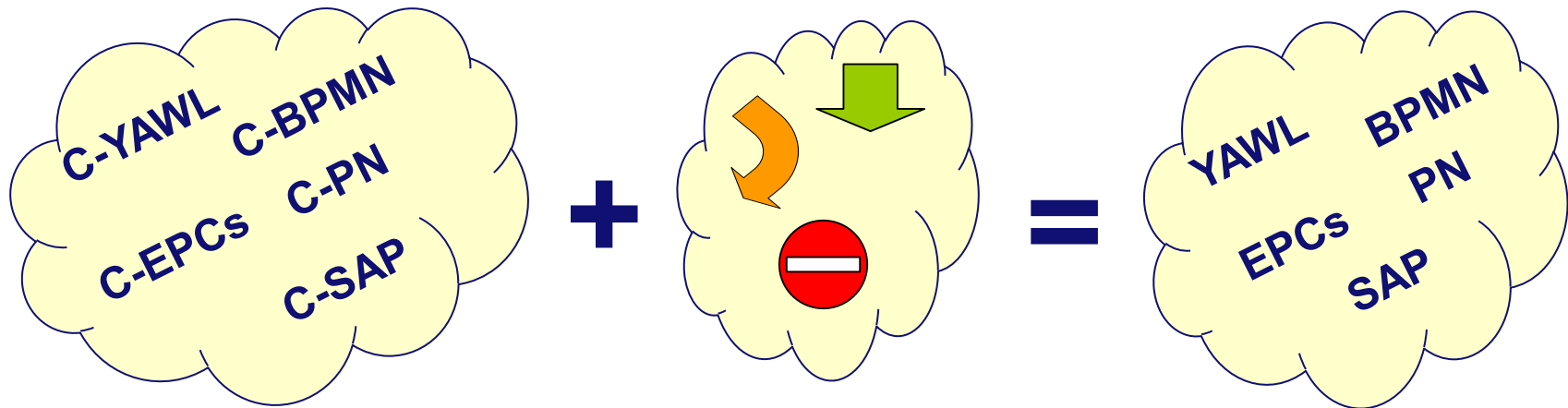
Hide/skip



Block



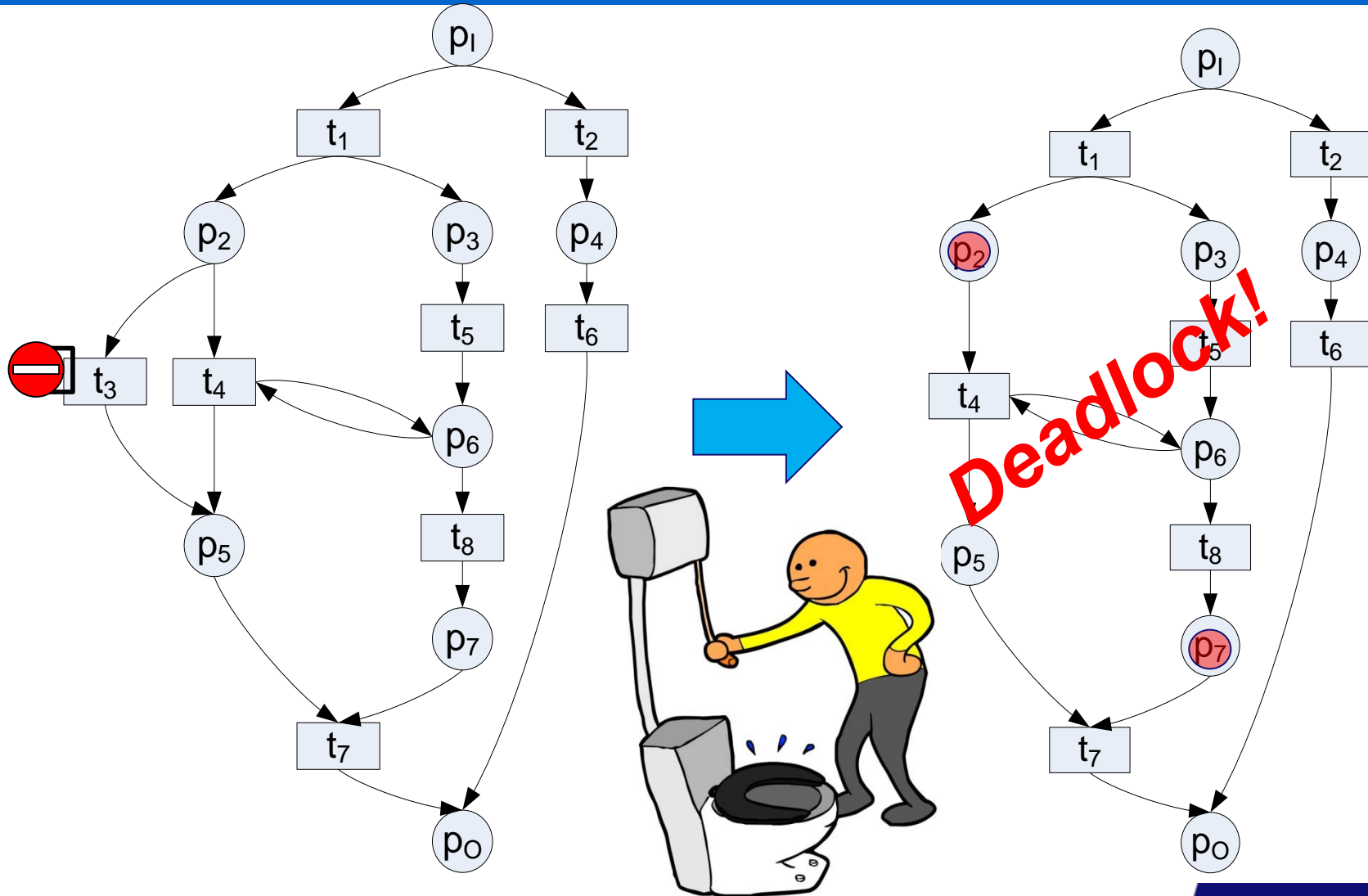
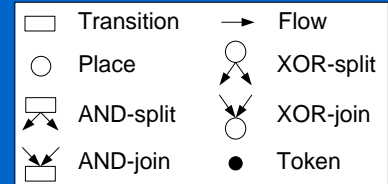
Correctness of configurations



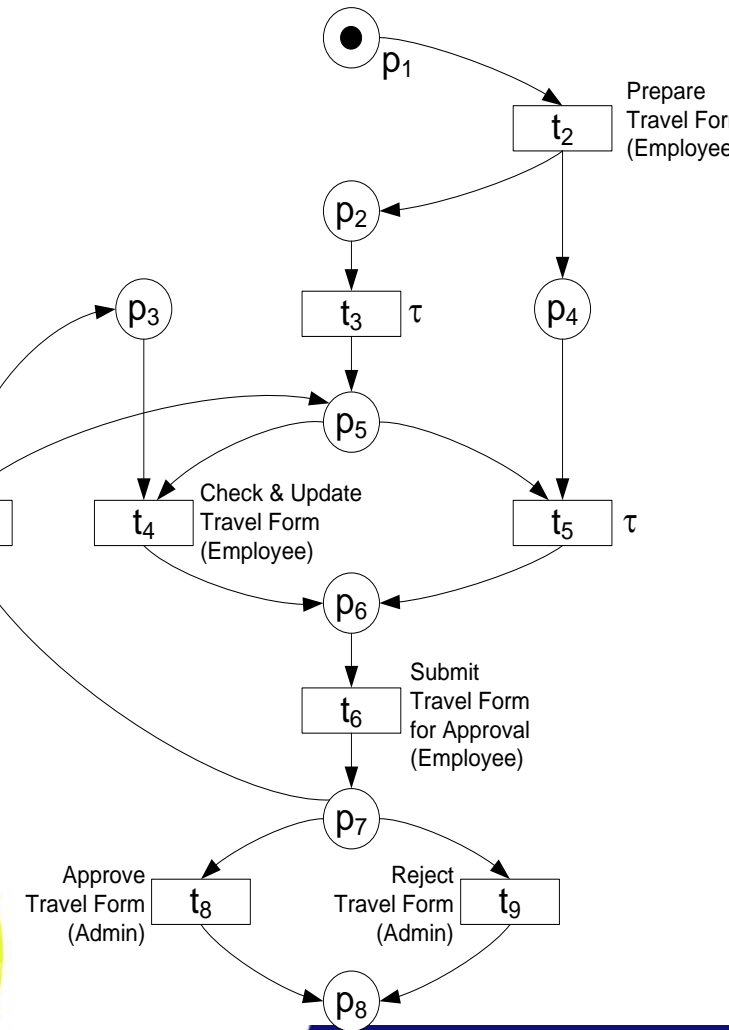
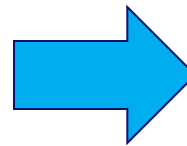
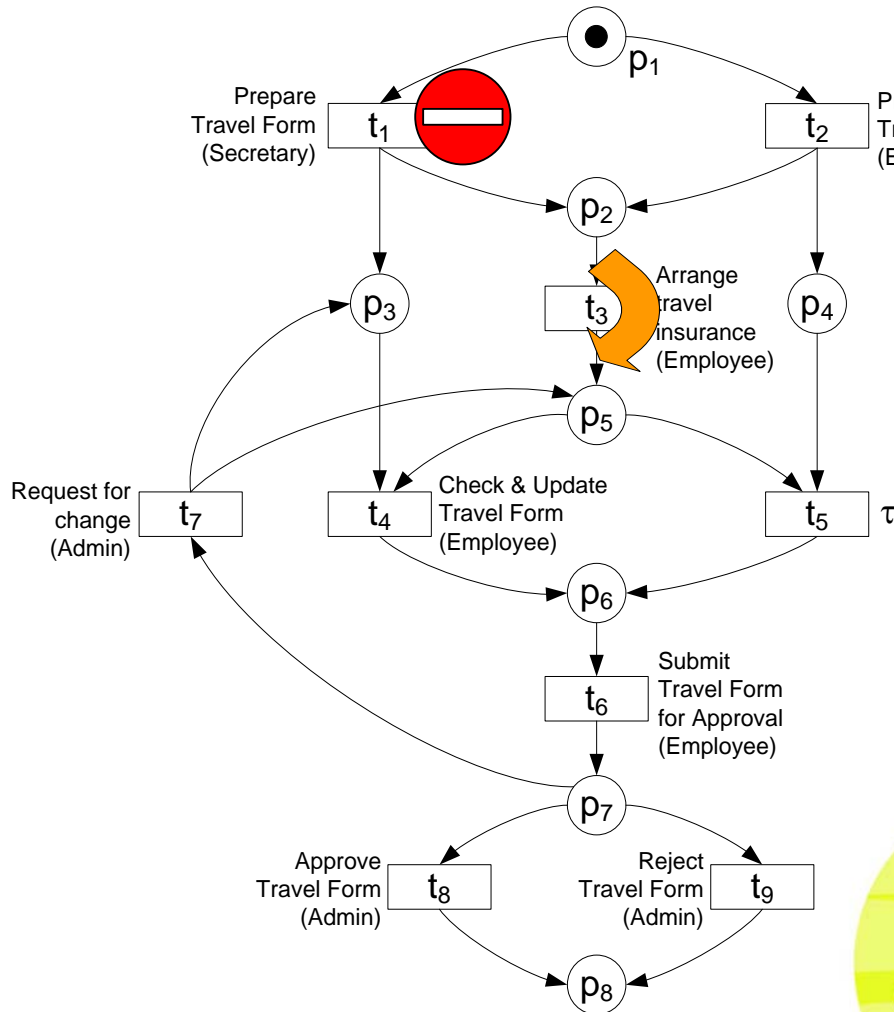
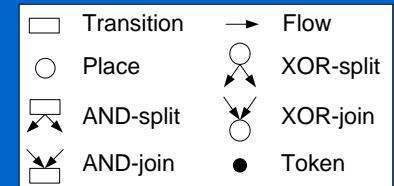
Configurable Model + Configuration = Configured model

- Question 1:
Is a particular configuration correct?
- Question 2:
Is there a correct configuration?
- Question 3:
How to characterize the set of all correct configurations?
- Question 4:
How to auto-complete a configuration?

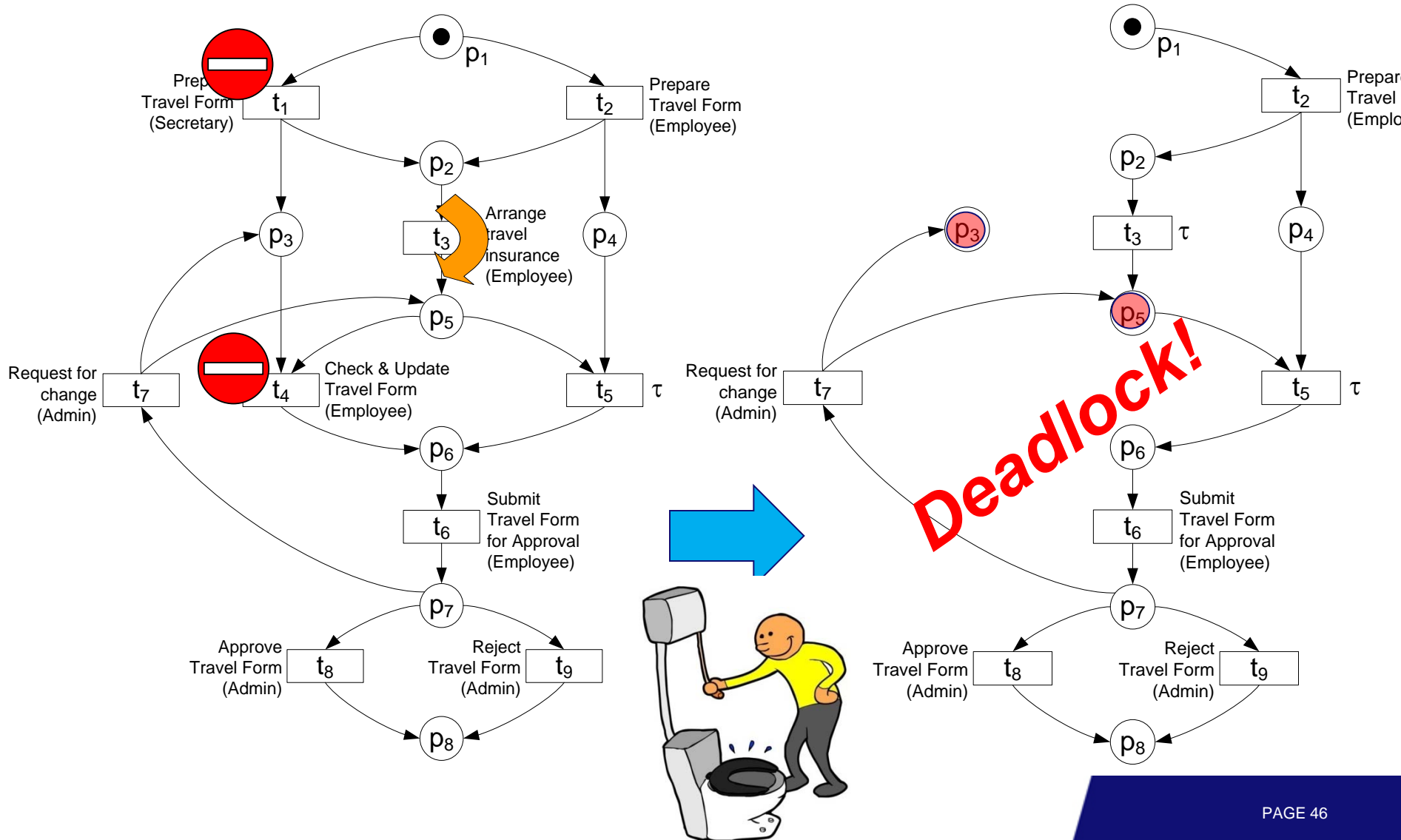
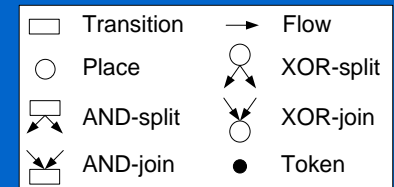
Can t_3 be blocked?



Block t_1 and hide t_3 ?



Block t_4 also?

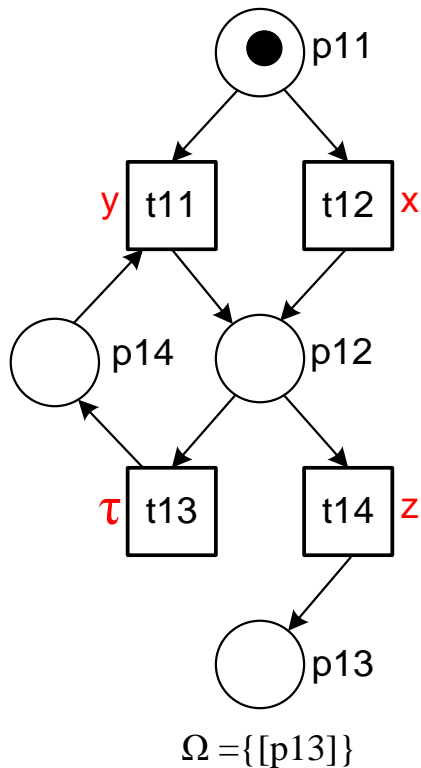


Existing approaches

- Most approaches only consider the syntactical issues or simply create the configured model and analyze it (i.e., trail and error).
- Naïve approach: enumerate all possibilities and check, or trail-and-error at configuration time.
- Better approach based on **partner synthesis**: construct a **configuration guideline** at design time!



Partner synthesis



partner??

=

**desirable
property
e.g. weakly
terminating**

Given a service:

- **Is this a partner?**
- **Is there a partner?**
- **How to describe all partners?**

Partner synthesis

- Existing approaches can do the following:
 - Check whether there exists a partner.
 - Construct partners having desirable properties.
 - Characterize all such partners.

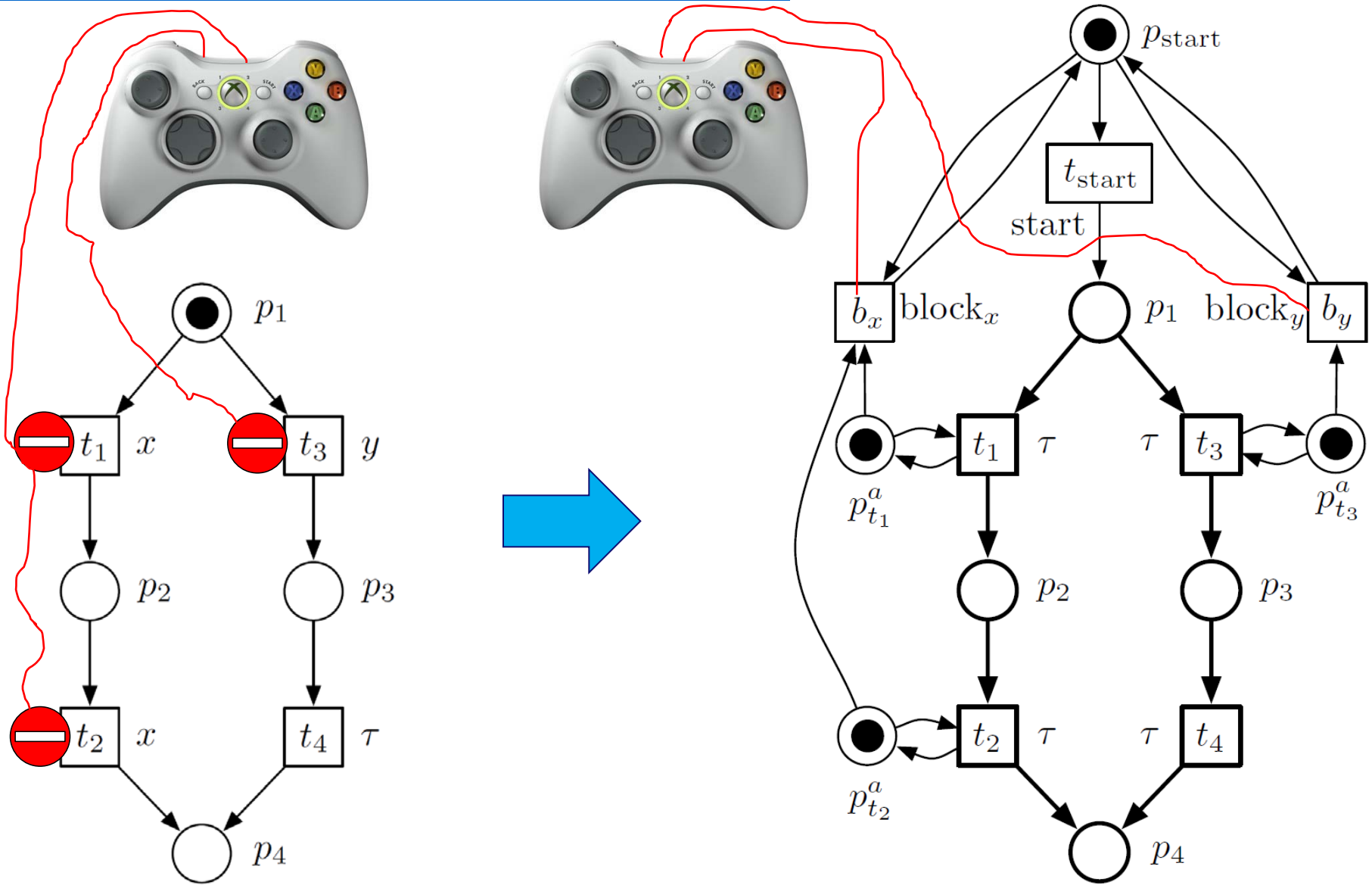
cf. [Karsten Wolf. Does My Service Have Partners?. T. Petri Nets and Other Models of Concurrency 2: 152-171 (2009)]

- These are implemented in **Wendy**.

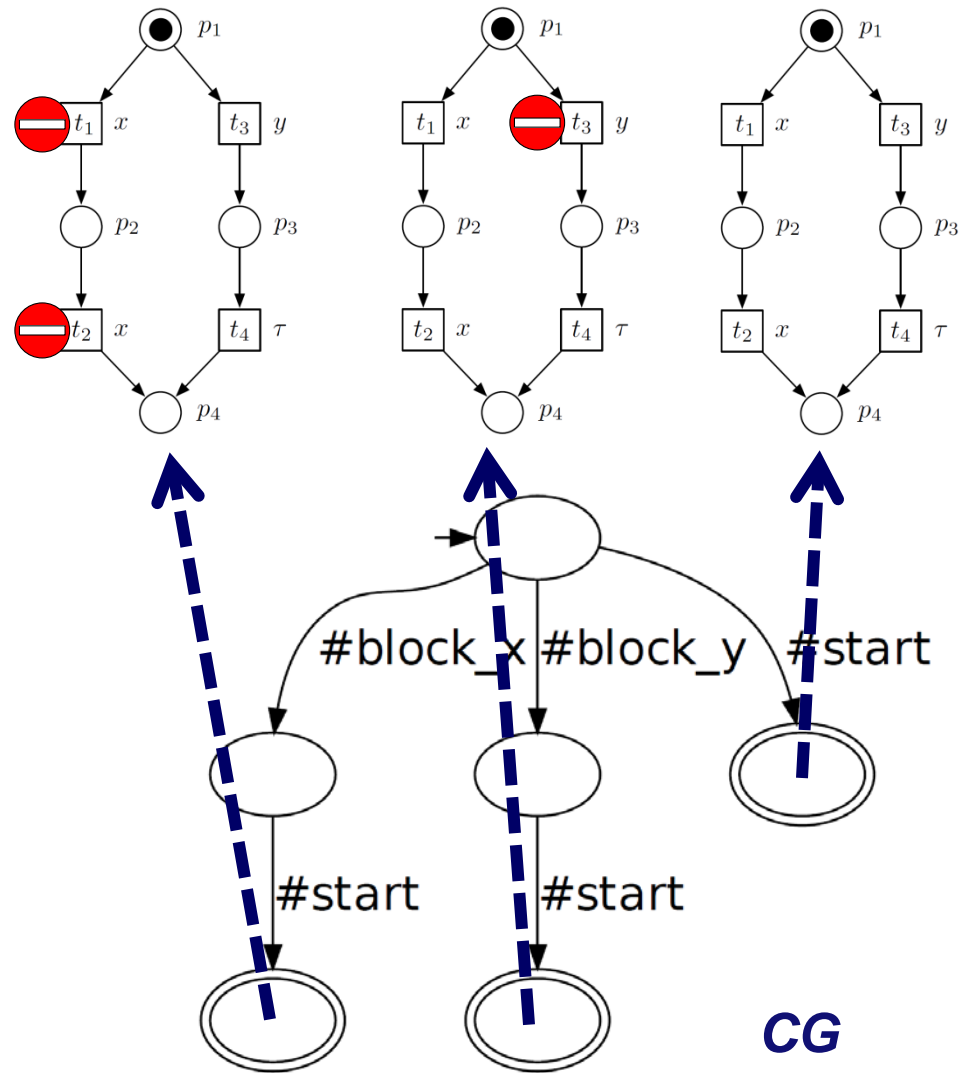
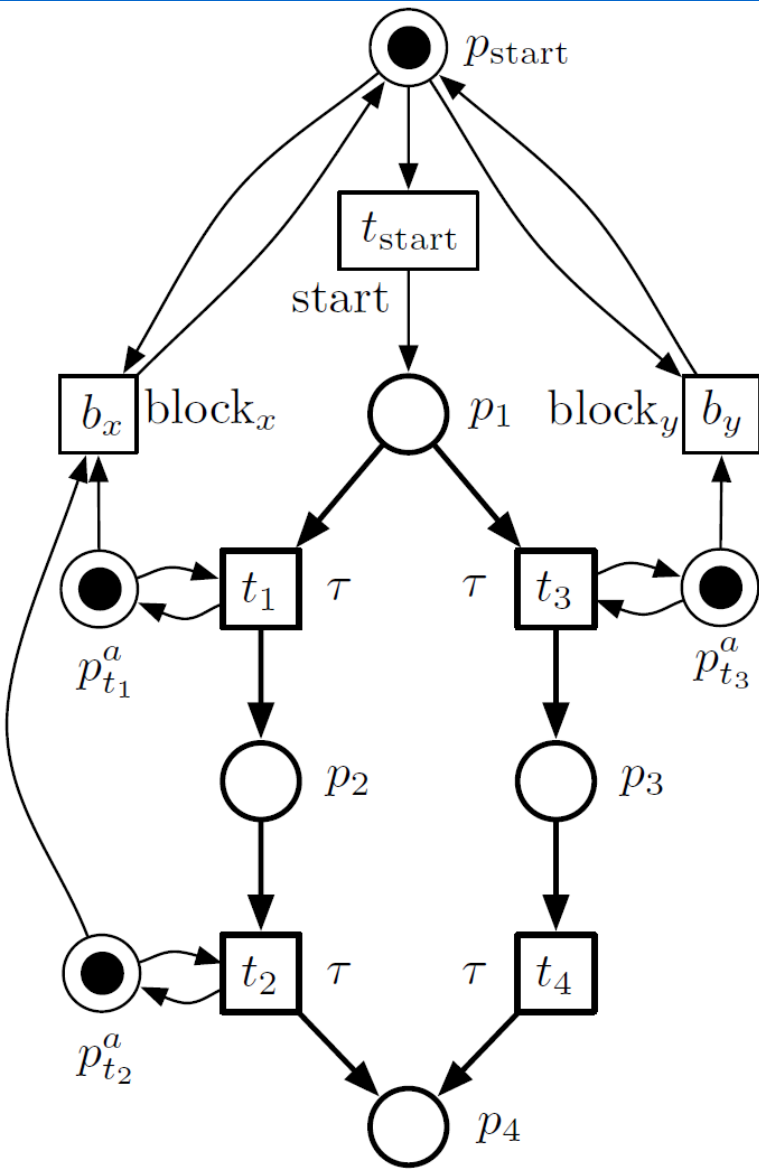
cf. [Niels Lohmann, Daniela Weinberg. Wendy: A Tool to Synthesize Partners for Services. Petri Nets 2010: 297-307.]

- How can this be used for ensuring the correctness of a configurable model and its configurations?

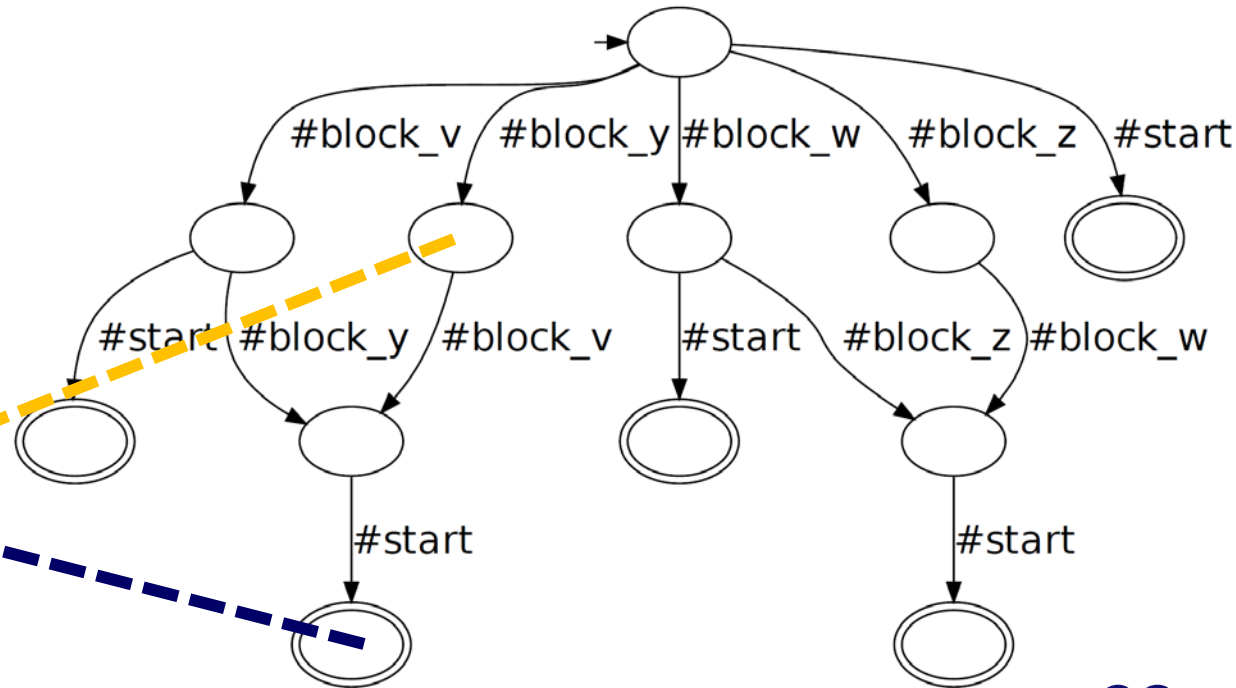
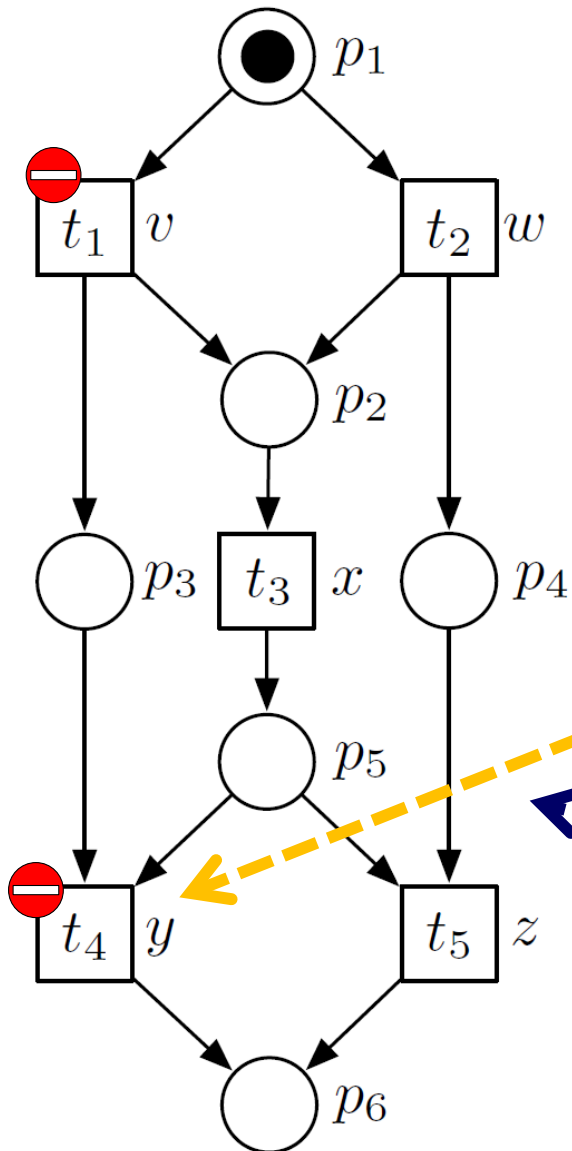
Configurable interface (allow by default)



Configuration guideline (allow by default)

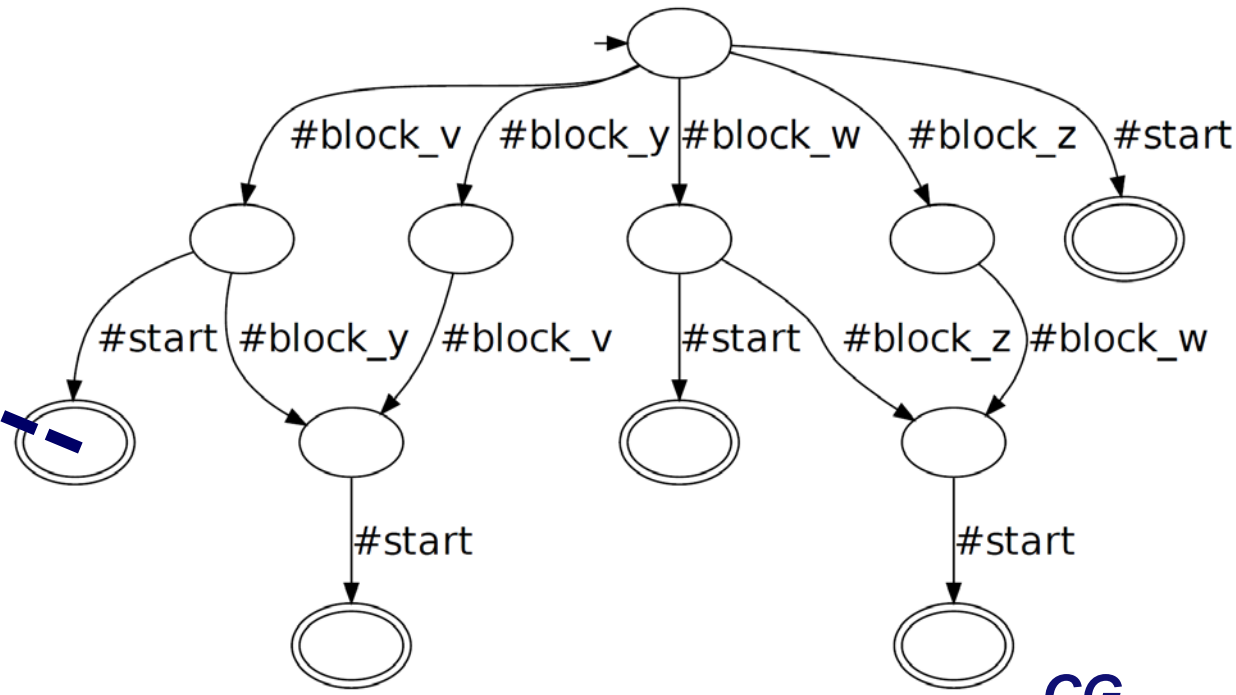
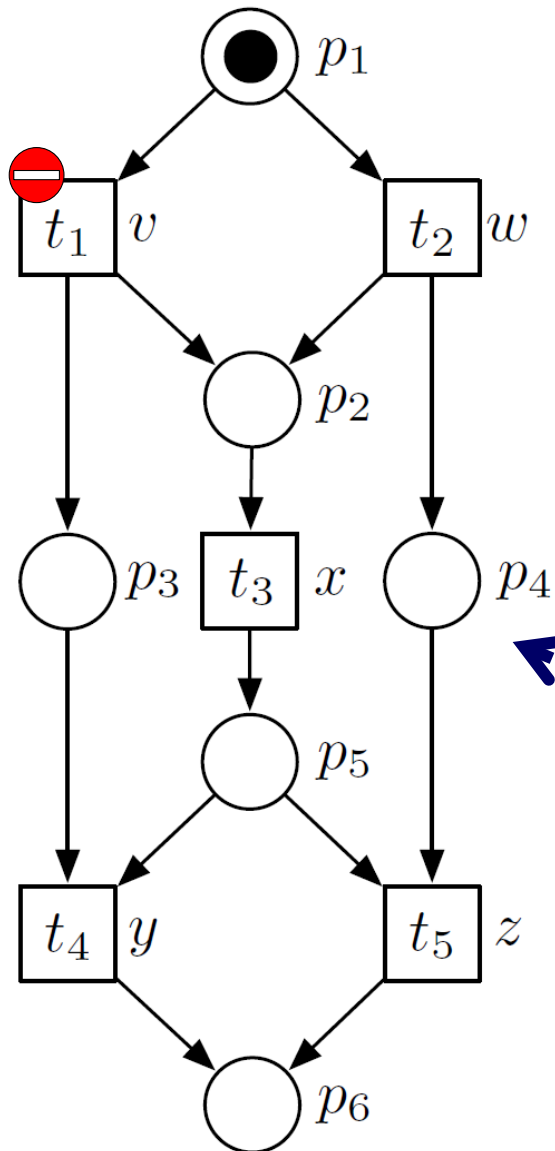


Configuration guideline (allow by default)



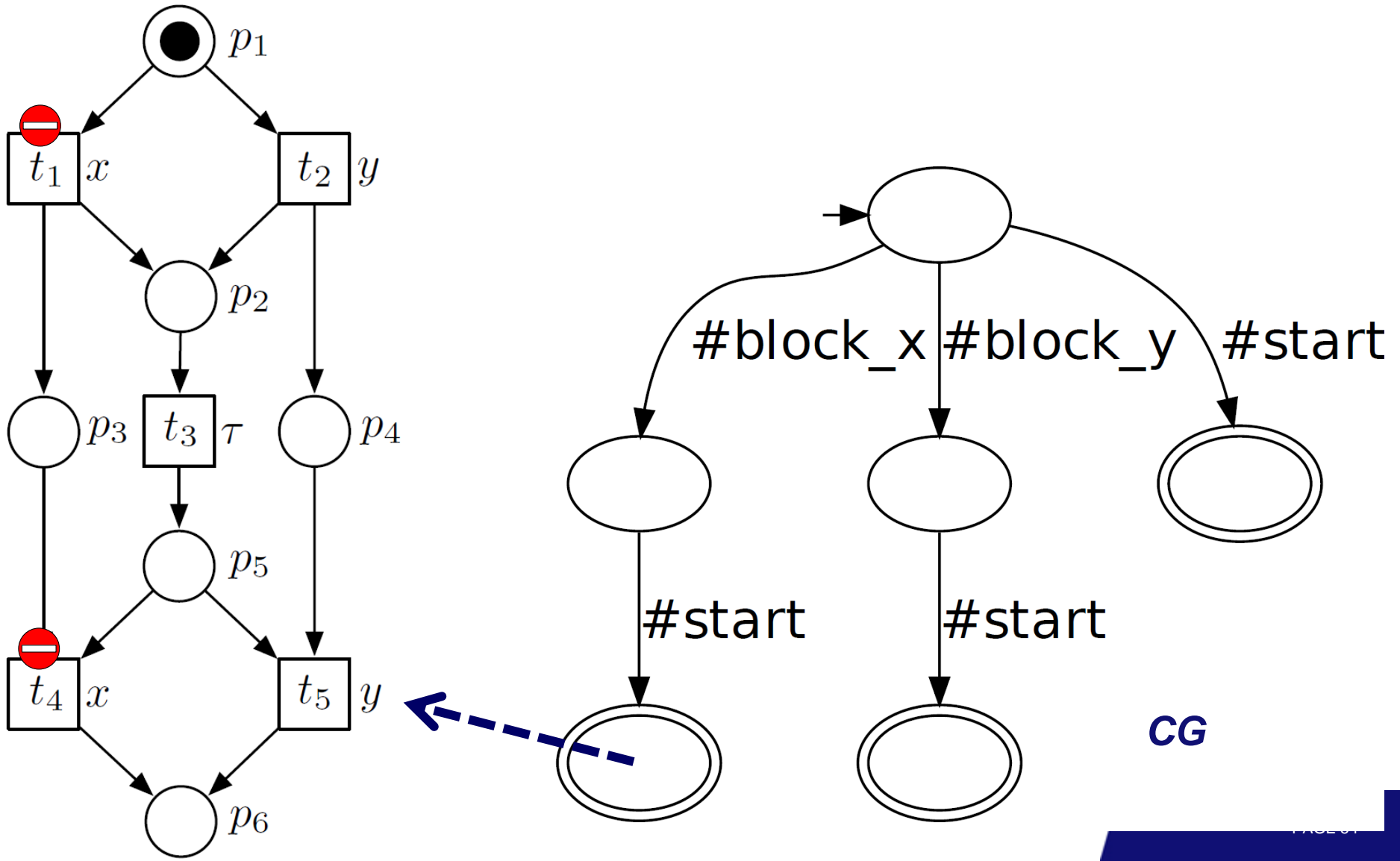
CG

Configuration guideline (allow by default)

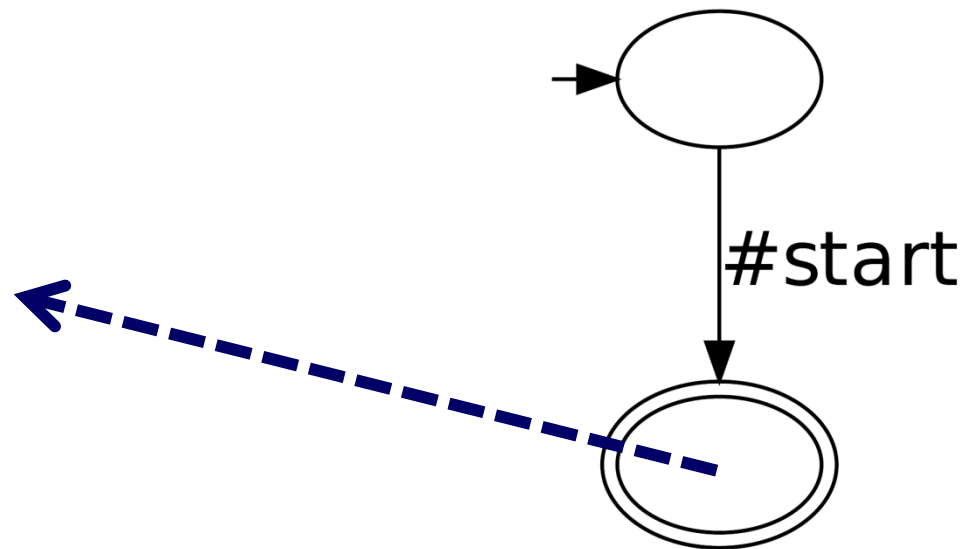
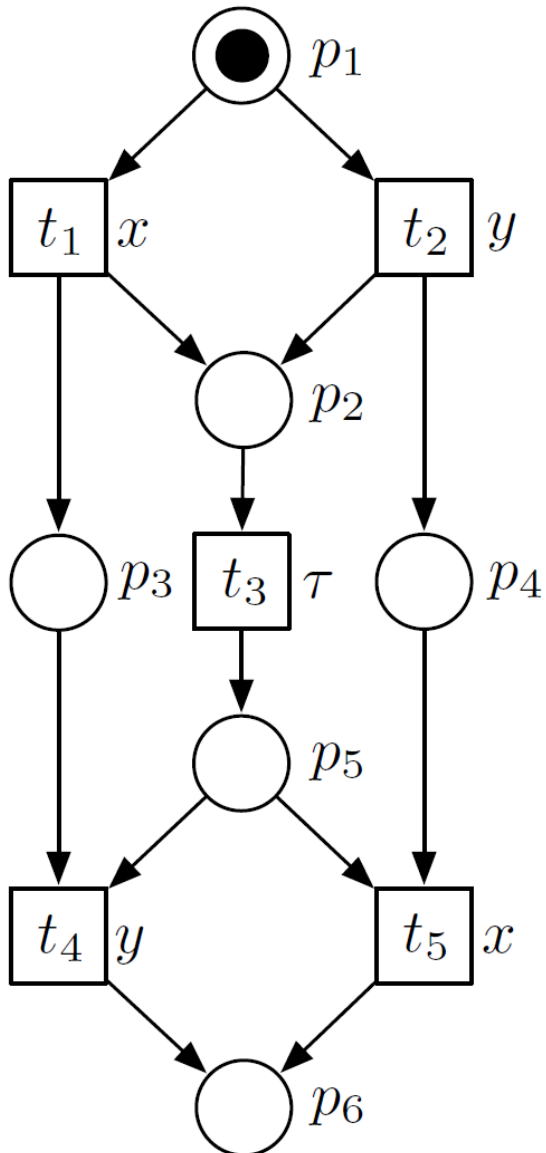


CG

Configuration guideline (allow by default)

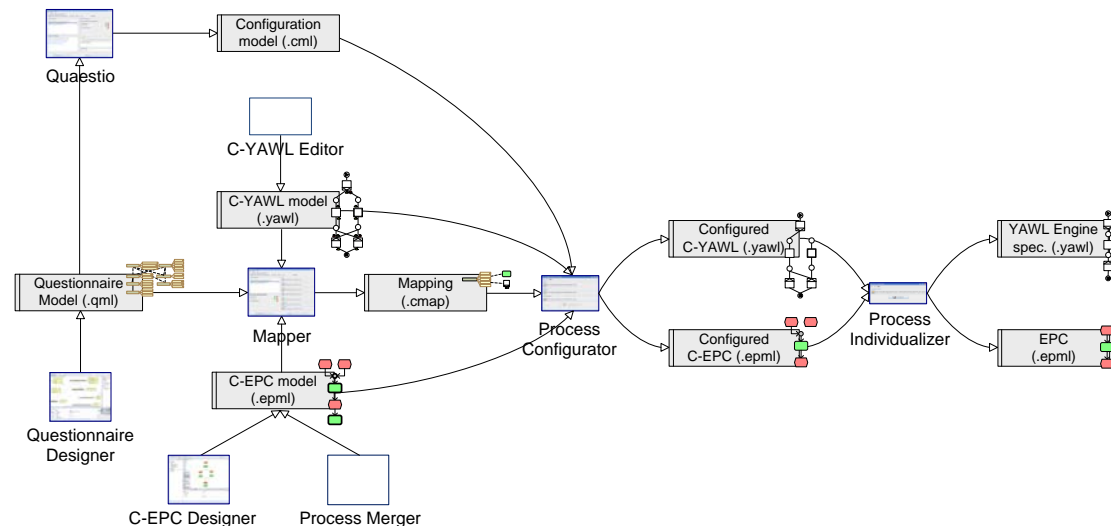


Configuration guideline (allow by default)



Tool support

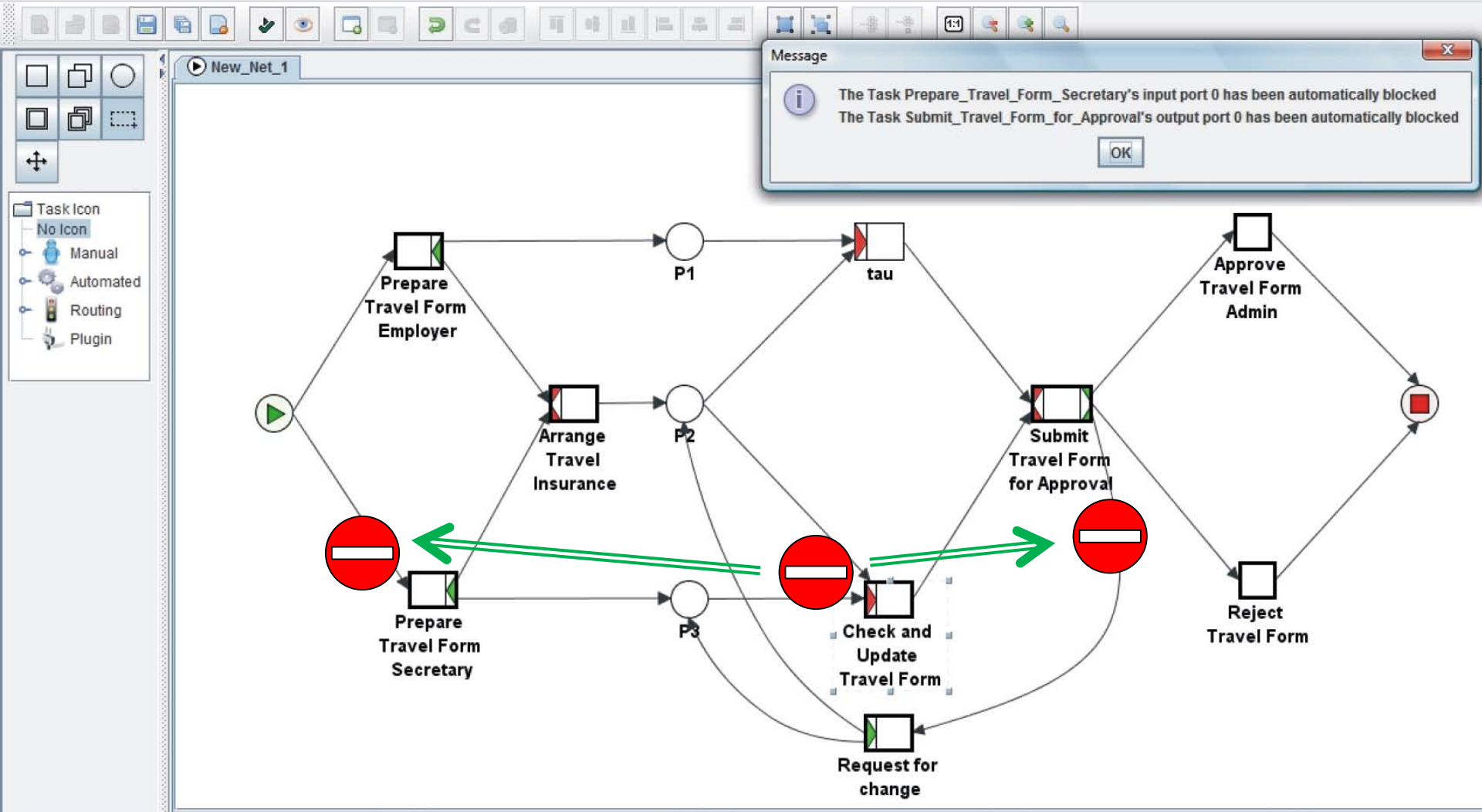
- **YAWL editor**
 - Creating C-YAWL models
 - Configuring C-YAWL models
 - Verification and auto-completion (using SAT solver and/or Wendy)
- **YAWL engine**
- **YAWL services**



C-YAWL

YAWLEditor - D:\Users\marcello\Desktop\tool\paper example\paper_example.yawl

Specification Net Edit Elements Tools View Help



Message



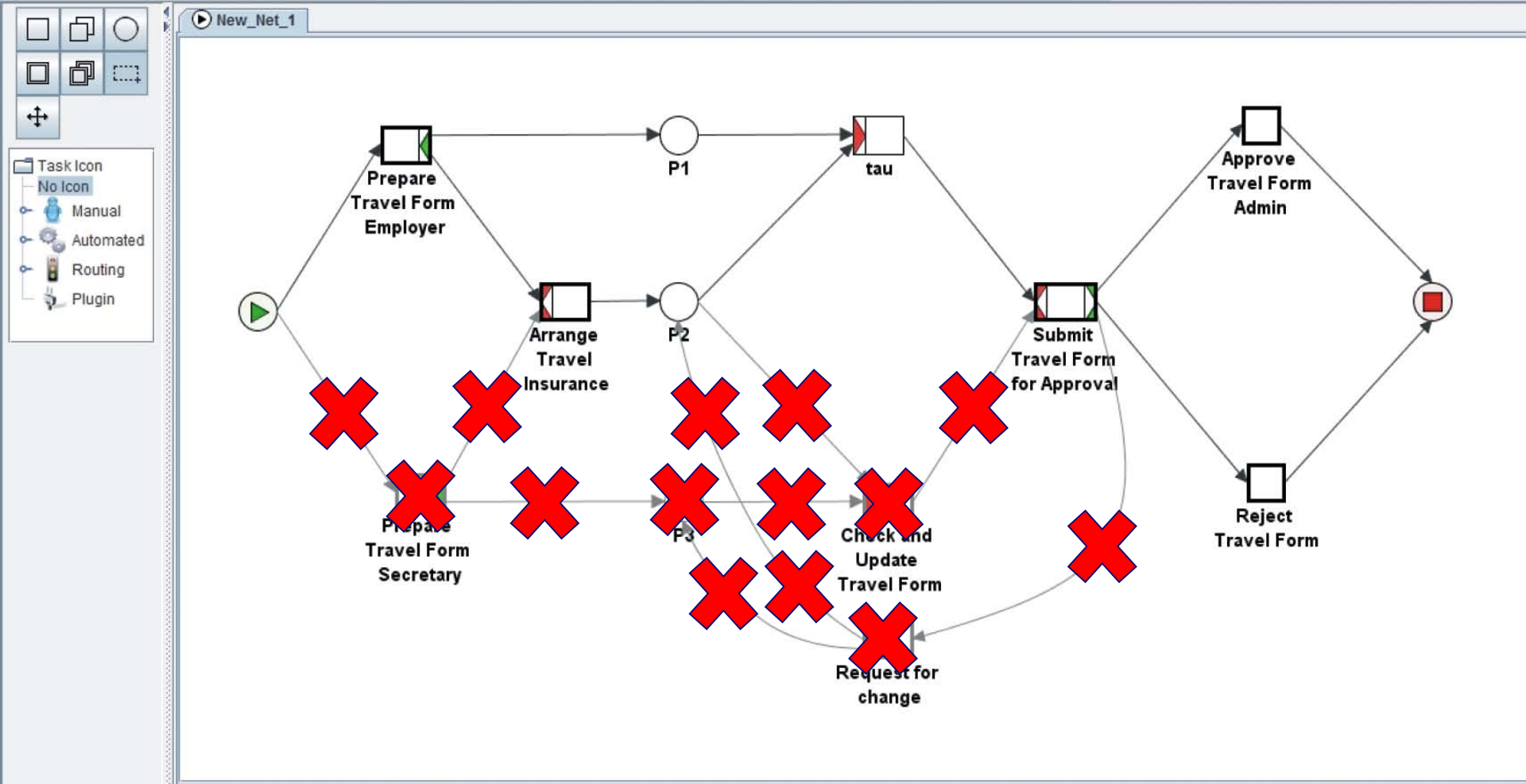
The Task Prepare_Travel_Form_Secretary's input port 0 has been automatically blocked
The Task Submit_Travel_Form_for_Approval's output port 0 has been automatically blocked

OK

C-YAWL

YAWLEditor - D:\Users\marcello\Desktop\tool\paper example\paper_example.yawl

Specification Net Edit Elements Tools View Help



Use the palette toolbar to edit the selected net.

Conclusion

- **BPM in the cloud triggers the need for configurable process models.**
- **Configuration is important, however, existing reference models are crap!**
- **Challenges:**
 - **Design of configurable models (language+approach)**
 - **Analysis of configurable models, e.g., ensuring correctness**
 - **Discovering configurable models**
 - **Cross-organizational process mining**

More information

- **www.processconfiguration.com**
(various references to configuration literature and a comprehensive toolset)
- **www.win.tue.nl/coselog**
(webpage of the CoSeLoG project)
- **www.processmining.org**
(webpage for process mining)
- **www.yawlfoundation.org**
(C-YAWL)
- **service-technology.org**
(analysis of services)



References

(Configurable Process Models)

- M. Rosemann and W.M.P. van der Aalst. A Configurable Reference Modelling Language. *Information Systems*, 32(1):1-23, 2007.
- F. Gottschalk, W.M.P. van der Aalst, and M.H. Jansen-Vullers. SAP WebFlow Made Configurable: Unifying Workflow Templates into a Configurable Model. In *BPM 2007*, volume 4714 of *Lecture Notes in Computer Science*, pages 262-270. Springer-Verlag, Berlin, 2007.
- F. Gottschalk, W.M.P. van der Aalst, M.H Jansen-Vullers, and M. La Rosa. Configurable Workflow Models. *International Journal of Cooperative Information Systems*, 17(2):177-221, 2008.
- M. La Rosa, W.M.P. van der Aalst, M. Dumas, and A.H.M. ter Hofstede. Questionnaire-based Variability Modeling for System Configuration. *Software and Systems Modeling*, 8(2):251-274, 2009.
- F. Gottschalk, W.M.P. van der Aalst, and H.M. Jansen-Vullers. Configurable Process Models: A Foundational Approach. In F. Lehner, H. Nosekabel, and P. Kleinschmidt, editors, *Proceedings of the Multikonferenz Wirtschaftsinformatik 2006 (MKWI '06)*. GITO-Verlag, Berlin, 2006.
- ...

References

(Correctness Issues)

- W.M.P. van der Aalst, M. Dumas, F. Gottschalk, A.H.M. ter Hofstede, M. La Rosa, and J. Mendling. Correctness-Preserving Configuration of Business Process Models. In *FASE 2008*, volume 4961 of *Lecture Notes in Computer Science*, pages 46-61. Springer-Verlag, Berlin, 2008.
- W.M.P. van der Aalst, M. Dumas, F. Gottschalk, A.H.M. ter Hofstede, M. La Rosa, and J. Mendling. Preserving Correctness During Business Process Model Configuration. *Formal Aspects of Computing*, 22(3):459-482, 2010.
- W.M.P. van der Aalst, N. Lohmann, M. La Rosa, and J. Xu. Correctness Ensuring Process Configuration: An Approach Based on Partner Synthesis. In *BPM 2010*, volume 6336 of *Lecture Notes in Computer Science*, pages 95-111. Springer-Verlag, Berlin, 2010.
- M.T. Wynn, H.M.W. Verbeek, W.M.P. van der Aalst, A.H.M. ter Hofstede, and D. Edmond. Business Process Verification: Finally a Reality! *Business Process Management Journal*, 15(1):74-92, 2009.
- H.M.W. Verbeek, T. Basten, and W.M.P. van der Aalst. Diagnosing Workflow Processes using Woflan. *The Computer Journal*, 44(4):246-279, 2001.

References

(Process Mining)

- M.H. Jansen-Vullers, W.M.P. van der Aalst, and M. Rosemann. Mining Configurable Enterprise Information Systems. *Data and Knowledge Engineering*, 56(3):195-244, 2006.
- F. Gottschalk, T. Wagemakers, M.H. Jansen-Vullers, W.M.P. van der Aalst, and M. La Rosa. Configurable Process Models: Experiences From a Municipality Case Study. In *CAiSE'09*, volume 5565 of *Lecture Notes in Computer Science*, pages 486-500. Springer-Verlag, Berlin, 2009.
- W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128-1142, 2004.
- W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, and H.M.W. Verbeek. Business Process Mining: An Industrial Application. *Information Systems*, 32(5):713-732, 2007.
- ...