



TU/e

Technische Universiteit
Eindhoven
University of Technology



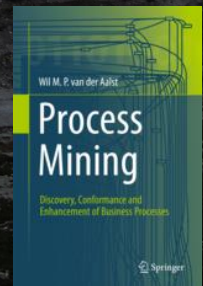
2013
CEE-SEC R

Mine Your Own Business

**Turning (Big) Data into Real Value
using Process Mining**

prof.dr.ir. Wil van der Aalst

***Central and Eastern European Software Engineering
Conference in Russia (CEE-SEC R 2013), Moscow,
October 25th 2013***



SCOOBY DOO, WHERE ARE YOU!
IN:
**MINE YOUR
OWN BUSINESS**

© 1969 HANNA-BARBERA PRODUCTIONS, INC.



Season 1, Episode 4 (1969)

Evidence-based BPM and Auditing

Process Mining:
The missing link

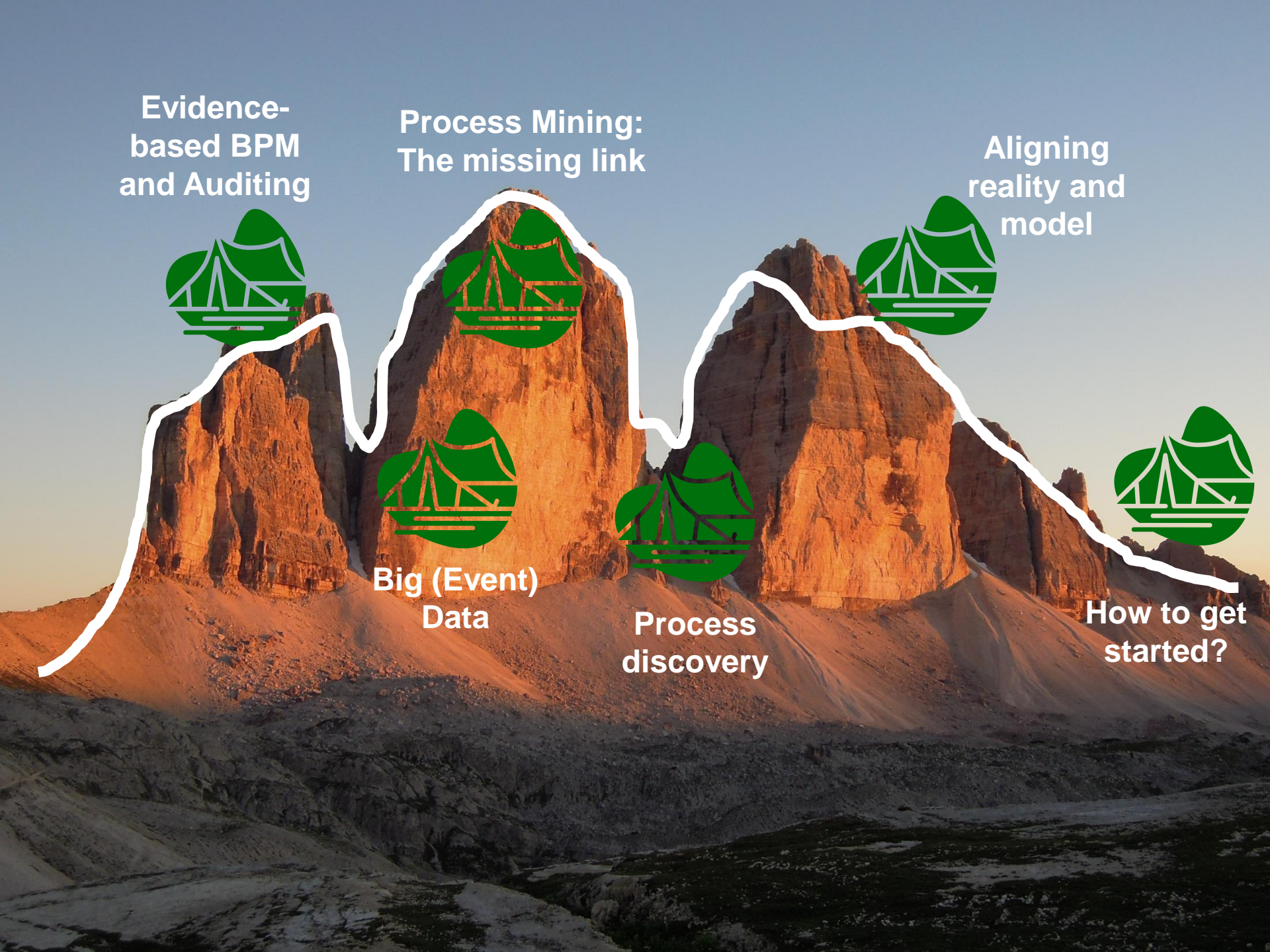
Aligning reality and model



Big (Event)
Data

Process
discovery

How to get
started?



Evidence-based BPM



Process Mining:
The missing link



Aligning
reality and
model



Big (Event)
Data



Process
discovery



How to get
started?

```
switch  
case IDENT:  
    i = internal_func(  
        if(i != -1) { /* rich  
            *value = (*intern_func[i].p)(  
        }  
    else  
        if(find_func(token)){ /* rich  
            call();  
            *value = ret_value;  
        }  
    else *value = find_va  
        get_token();  
    return;  
    ABER;
```

**Why do (larger)
organizations
have software?**



**It is always about
supporting or improving
business processes**



**Business process
problem or IT problem?**



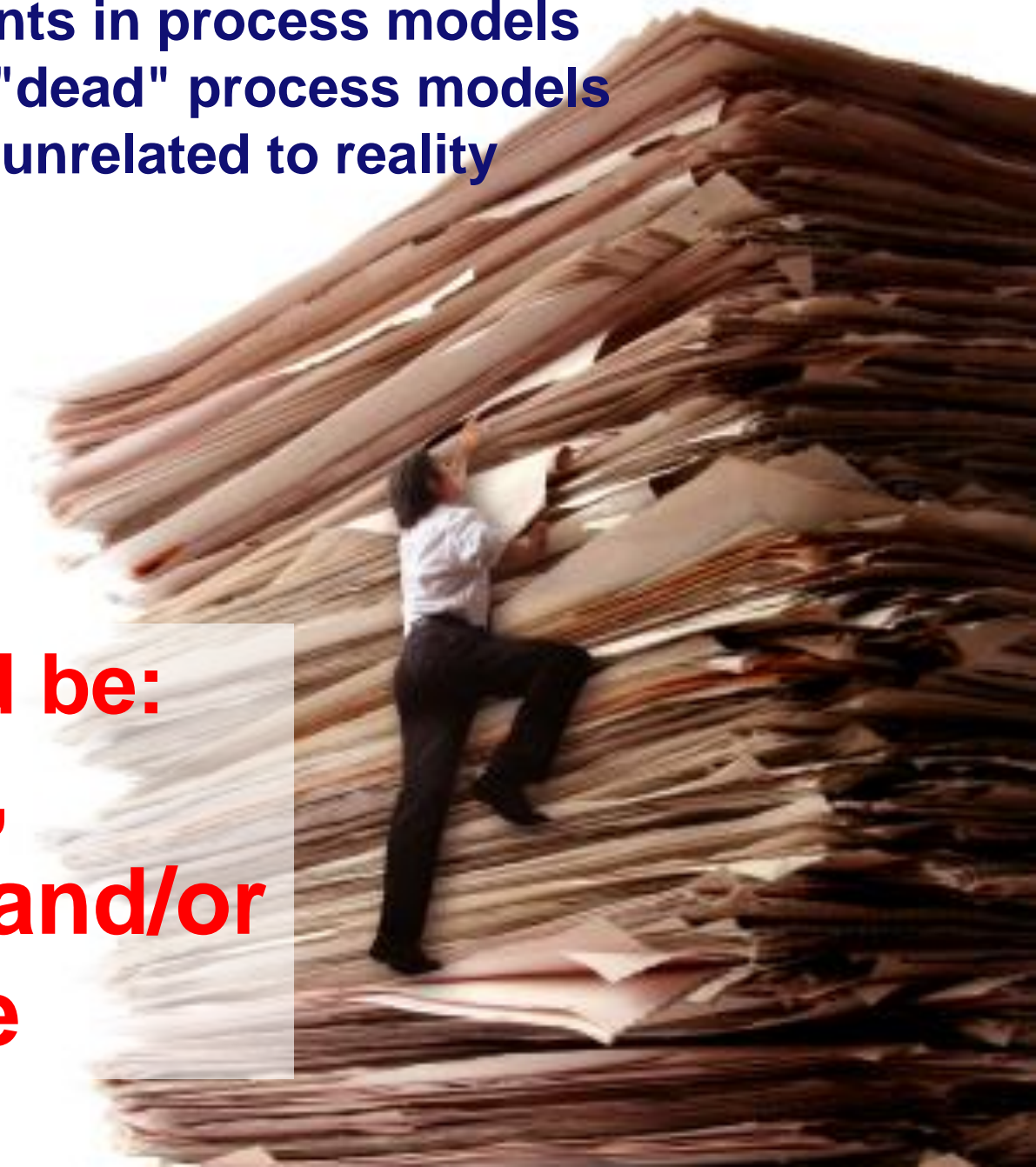


BPM efforts focus on models

- enormous investments in process models
- large collections of "dead" process models
- not taken seriously, unrelated to reality

Models should be:

- descriptive,
- predictive, and/or
- prescriptive





REALITY

Worst game ever.

Evidence-based BPM and Auditing

Process Mining:
The missing link

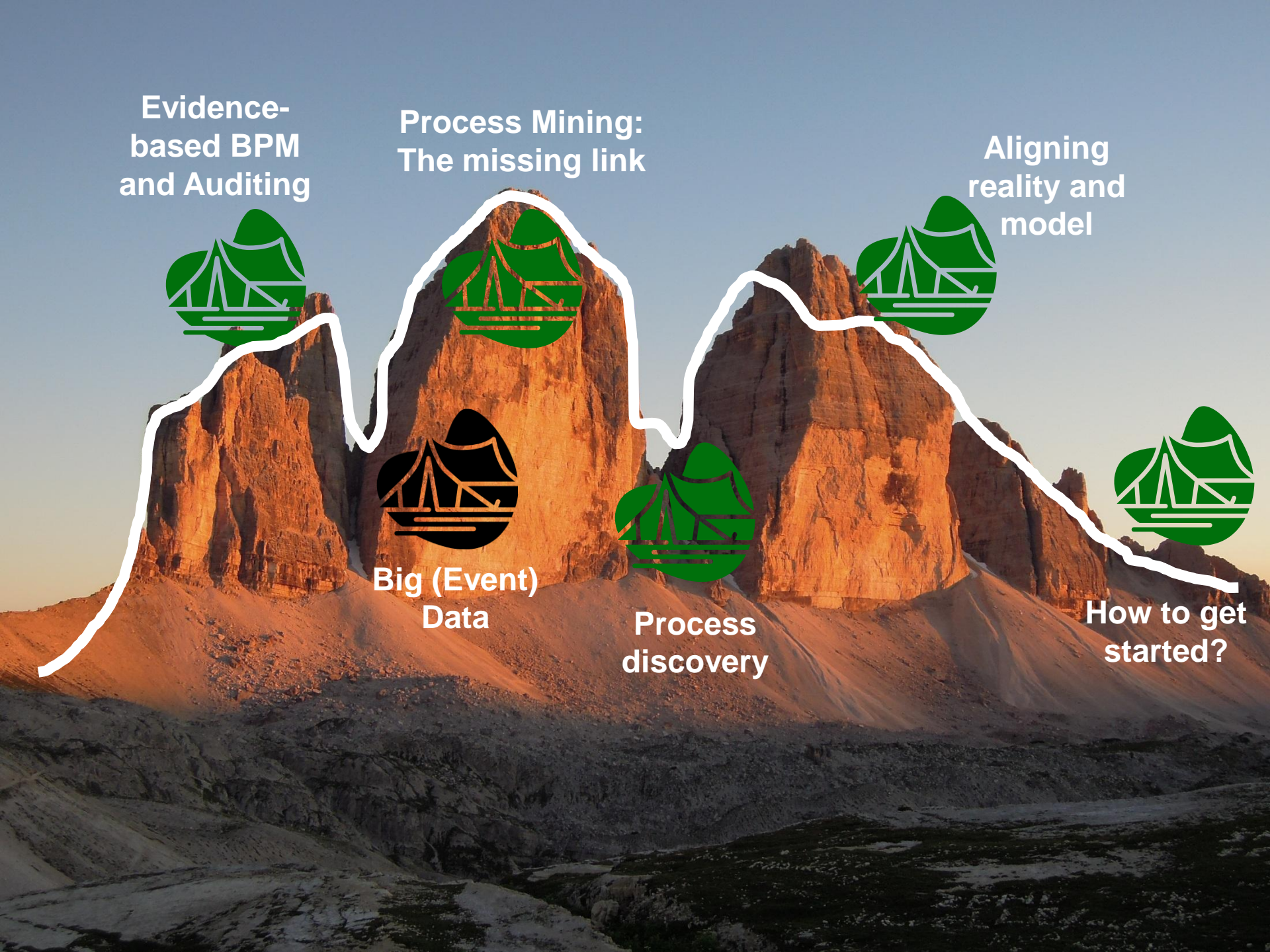
Aligning reality and model



Big (Event)
Data

Process
discovery

How to get
started?



Motivation: Increasing awareness of the value of (Big) Data

- **"In God we trust. All others must bring data"** (William Edwards Deming, statistician),
- **"Data is a precious thing and will last longer than the systems themselves"** (Tim Berners-Lee),
- **"Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital"** (Aaron Levenstein, statistician),
- **"Every 2 days we create as much information as we did up to 2003"** (Eric Schmidt, Google CEO, August 4, 2010).

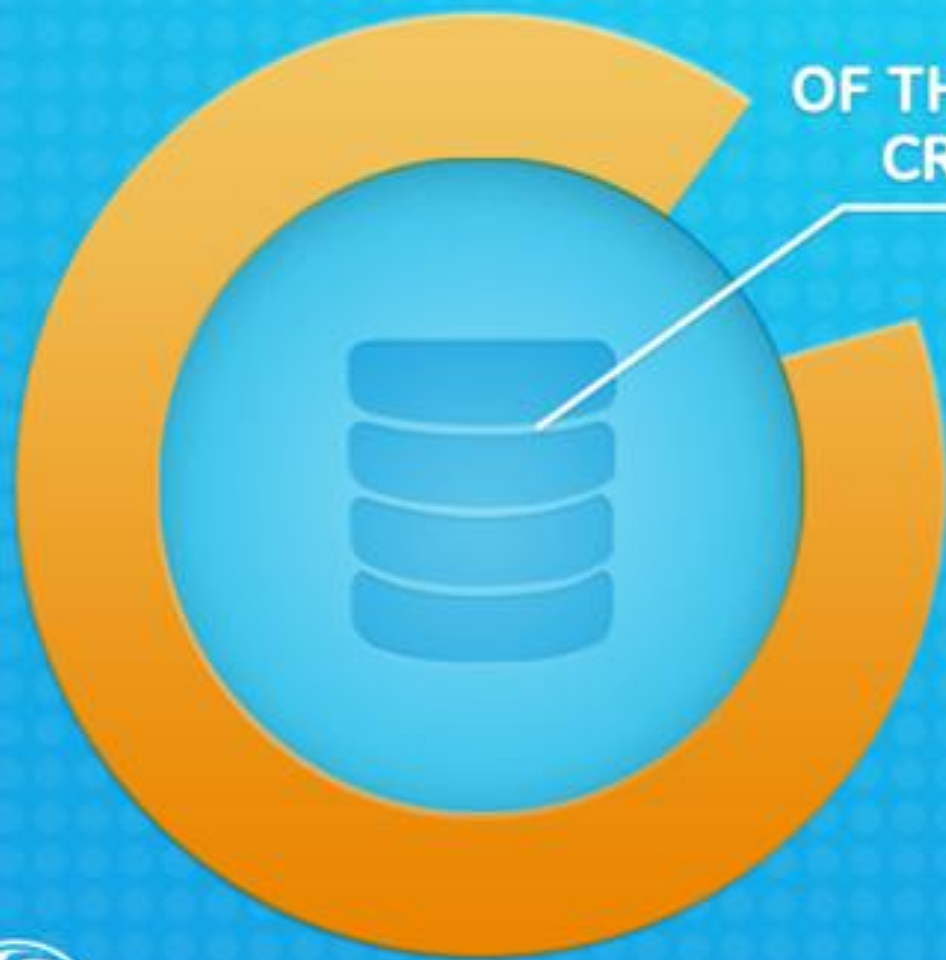
The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

In 10 years we will have 50 times as much data! (IDC)



90%

OF THE WORLD'S DATA WAS
CREATED IN THE LAST
TWO YEARS





THE INTERNET OF

Events

The Economist

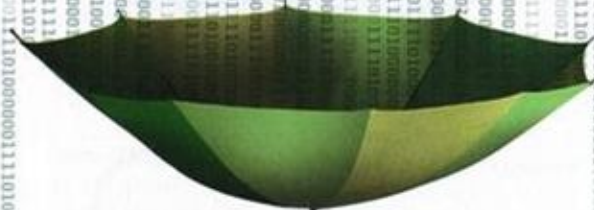
FEBRUARY 27TH - MARCH 5TH 2010

Economist.com

Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT






Big Data ?



Big ... or fast and efficient?



Process-awareness is an essential but often forgotten ingredient when converting big data into real value

Evidence-based BPM and Auditing



Process Mining:
The missing link



Aligning reality and model



Big (Event)
Data



Process
discovery



How to get
started?





process model analysis

(simulation, verification, optimization, gaming, etc.)

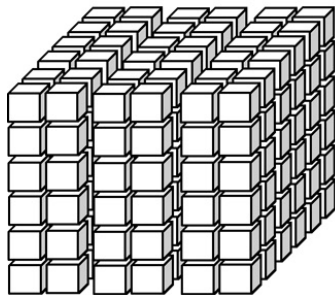


performance-oriented questions, problems and solutions



compliance-oriented questions, problems and solutions

process mining



data-oriented analysis

(data mining, machine learning, business intelligence)



电子工程楼
ELECTRONIC ENGINEERING
BUILDING
↑

0100110011010101010



小草对您微微笑
请您把路绕一绕

KEEP OFF GRASS

绿色大学办公室
修缮中心园林科

0100110011010101010

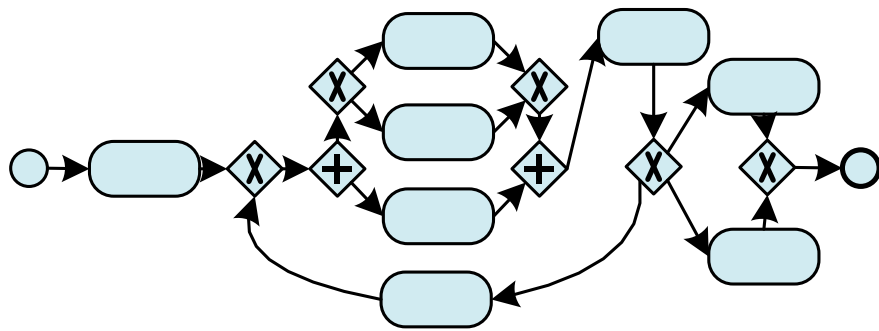
01001001100110101



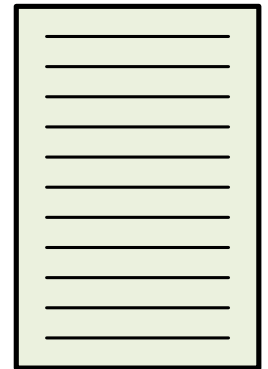


let's play

Play-Out

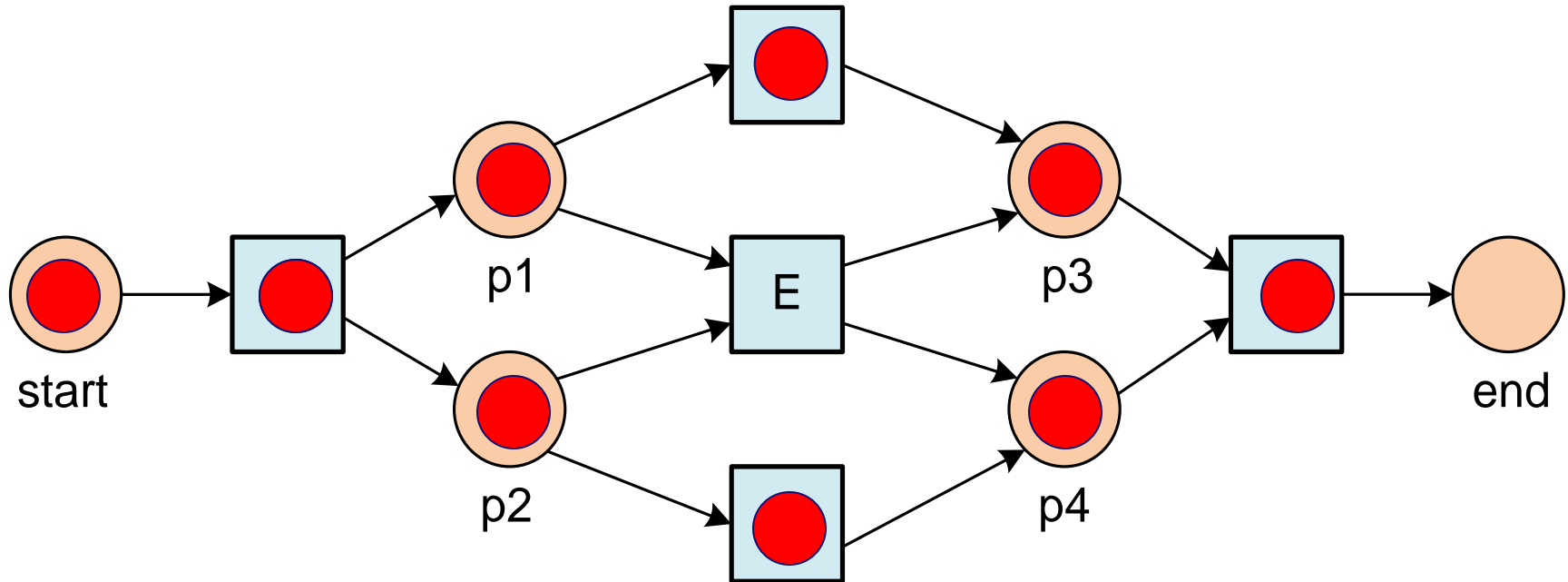


process model



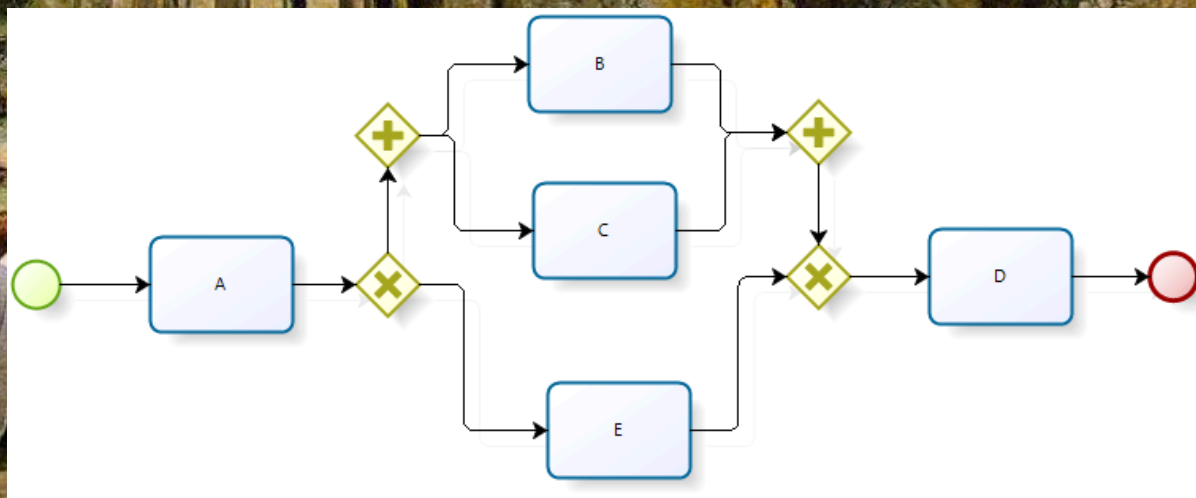
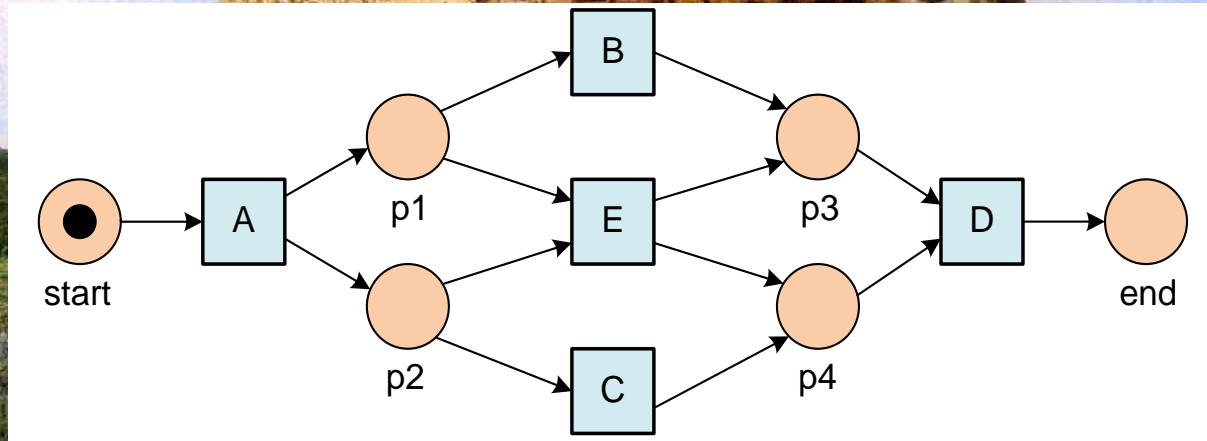
event log

Play-Out (Classical use of models)

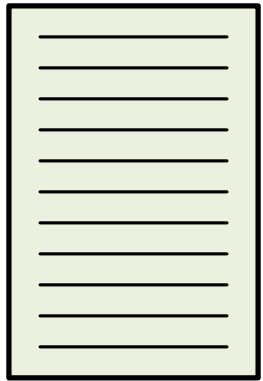


A B C D **A E D** **A E D**
A C B D **A B C D** **A C B D**
A C B D **A E D** **A C B D**

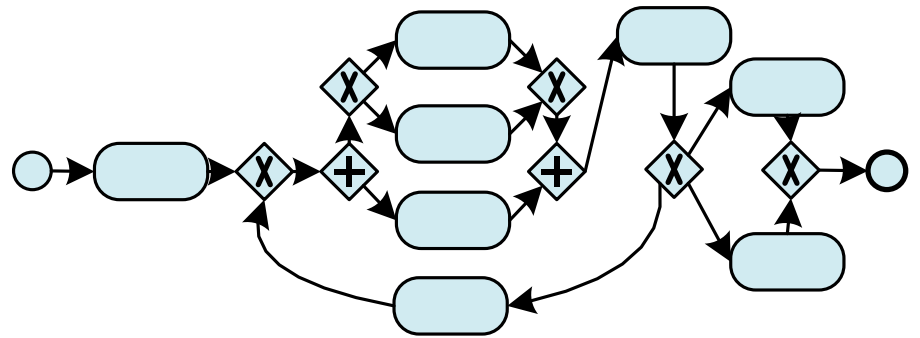
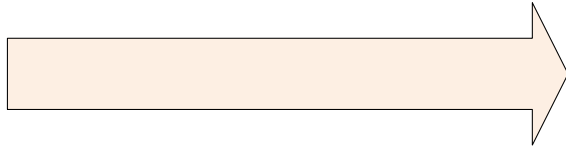
Let's not worry about syntax (there is difference between analysis and presentation)



Play-In



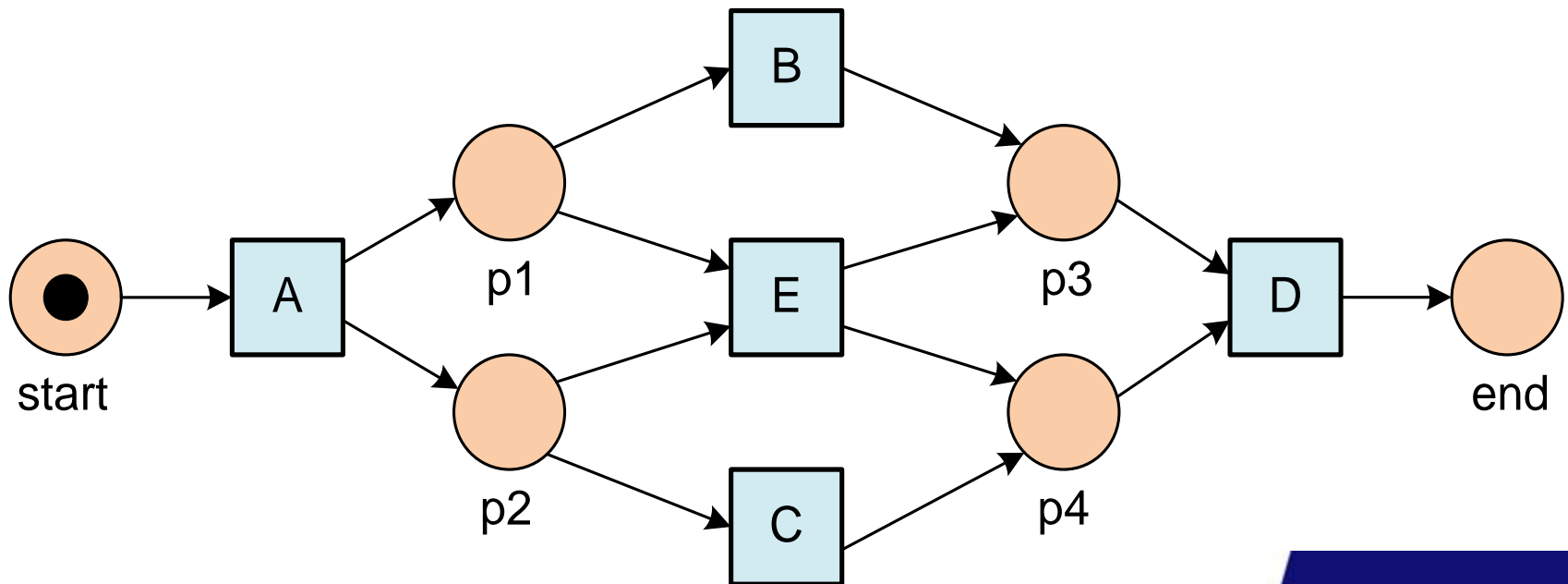
event log



process model

Play-In

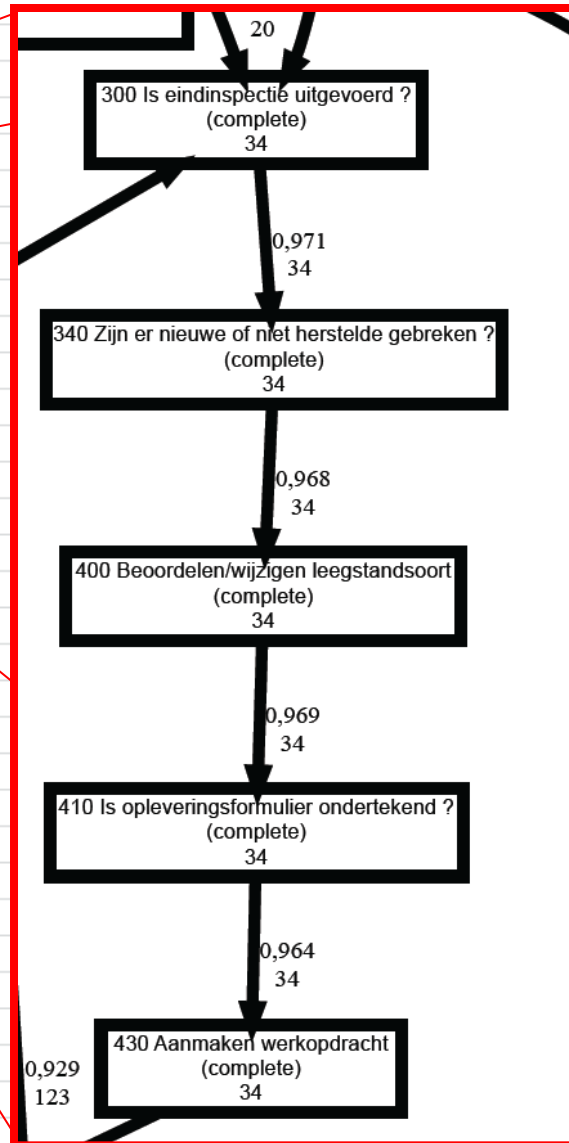
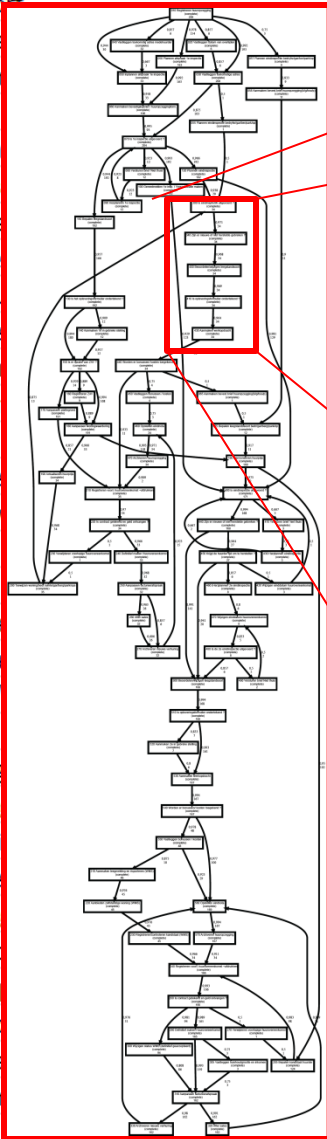
A B C D A E D A E D
A C B D A B C D A C B D
A C B D A E D A C B D



Example Process Discovery

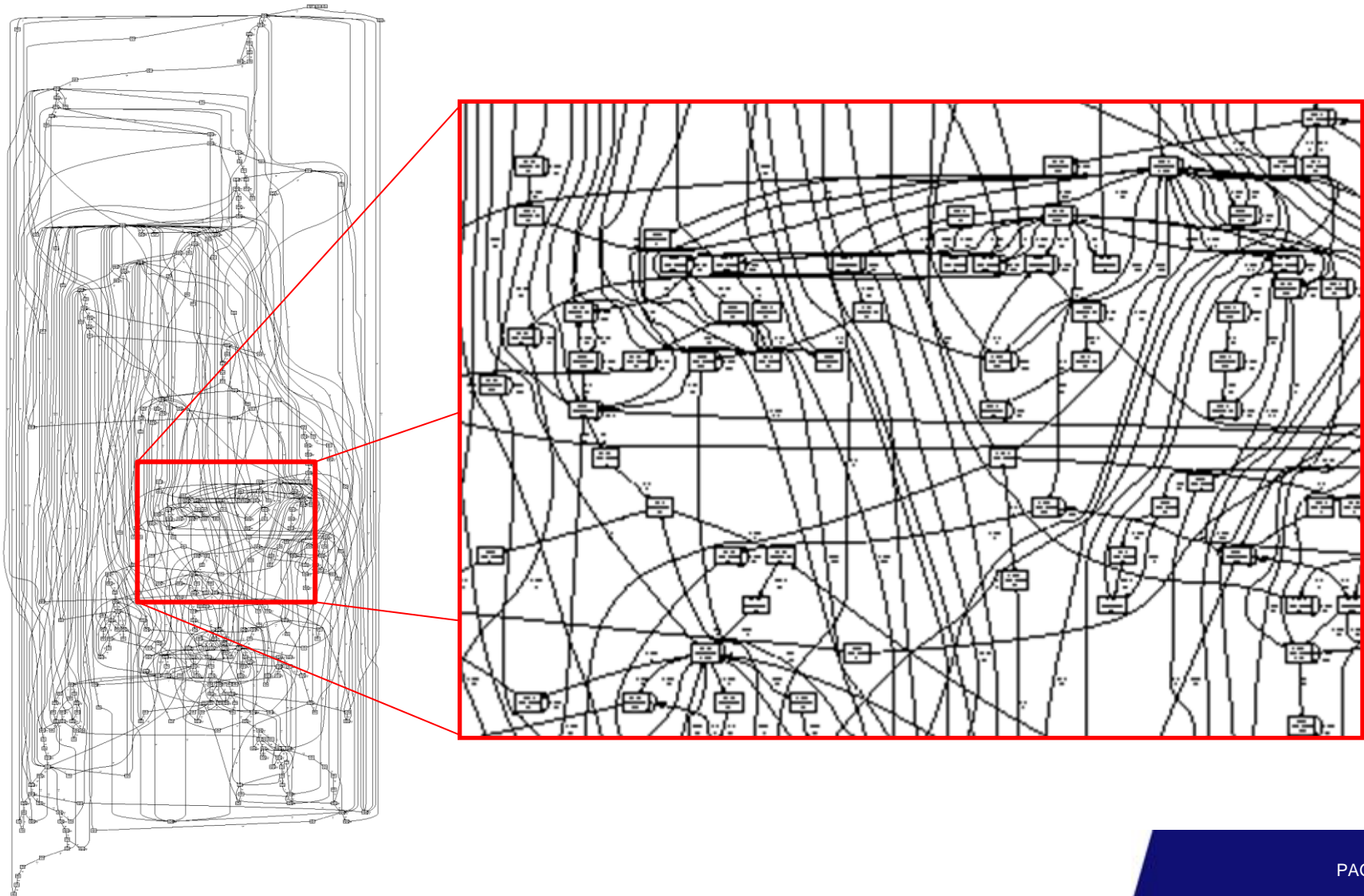
(Vestia, Dutch housing agency, 208 cases, 5987 events)

117315	110 Bepalen leegstandsoort	16.05.2007 14:06:23
117315	120 Plannen eindinspectie	16.05.2007 14:36:01
117315	130 Is het opleveringsform	23.05.2007 09:41:40
117315	150 Is er sprake van ZAV ?	23.05.2007 09:41:51
117315	170 Aanpassen plattegron	23.05.2007 11:57:18
117315	180 Aanpassen woningwa	23.05.2007 09:42:37
117315	190 Actualiseren huurprijs	23.05.2007 09:48:23
117315	200 Toewijzen woning/be	23.05.2007 09:48:29
117315	210 Registreren voorl. hu	10.09.2007 16:24:36
117315	220 Is contract getekend e	11.09.2007 14:56:18
117315	240 Definitief maken Huu	31.03.2008 16:17:12
117315	250 Aanpassen factuureera	09.09.2008 15:39:59
117315	260 After sales	09.09.2008 16:51:24
117315	270 Archiveren nieuwe ve	10.09.2008 07:52:08
117315	300 Is eindinspectie uitge	07.06.2007 14:47:04
117315	340 Zijn er nieuwe of niet	07.06.2007 14:47:06
117315	400 Beoordelen/wijzigen	07.06.2007 14:51:16
117315	410 Is opleveringsformulie	07.06.2007 14:51:26
117315	430 Aanmaken werkopdra	11.06.2007 09:21:39
117315	440 Worden er bonussen/	11.06.2007 09:21:49
117315	460 Opstellen eindnota	08.08.2007 16:18:26
117315	470 Archiveren huuropzeg	09.08.2007 14:42:23
119763	010 Registreren huuropze	09.05.2007 11:19:14
119763	030 Vastleggen toekomst	09.05.2007 12:25:01
119763	050 Inplannen afspraak 1e	09.05.2007 11:59:52
119763	060 Aanmaken bevestigin	09.05.2007 12:31:57
119763	070 Is 1e inspectie uitgev	16.05.2007 13:04:26
119763	100 Gereedmelden 1e ins	16.05.2007 13:43:39
119763	110 Bepalen leegstandsoort	16.05.2007 13:43:28
119763	120 Plannen eindinspectie	16.05.2007 13:42:58
119763	130 Is het opleveringsform	16.05.2007 13:34:49
119763	150 Is er sprake van ZAV ?	16.05.2007 13:34:56



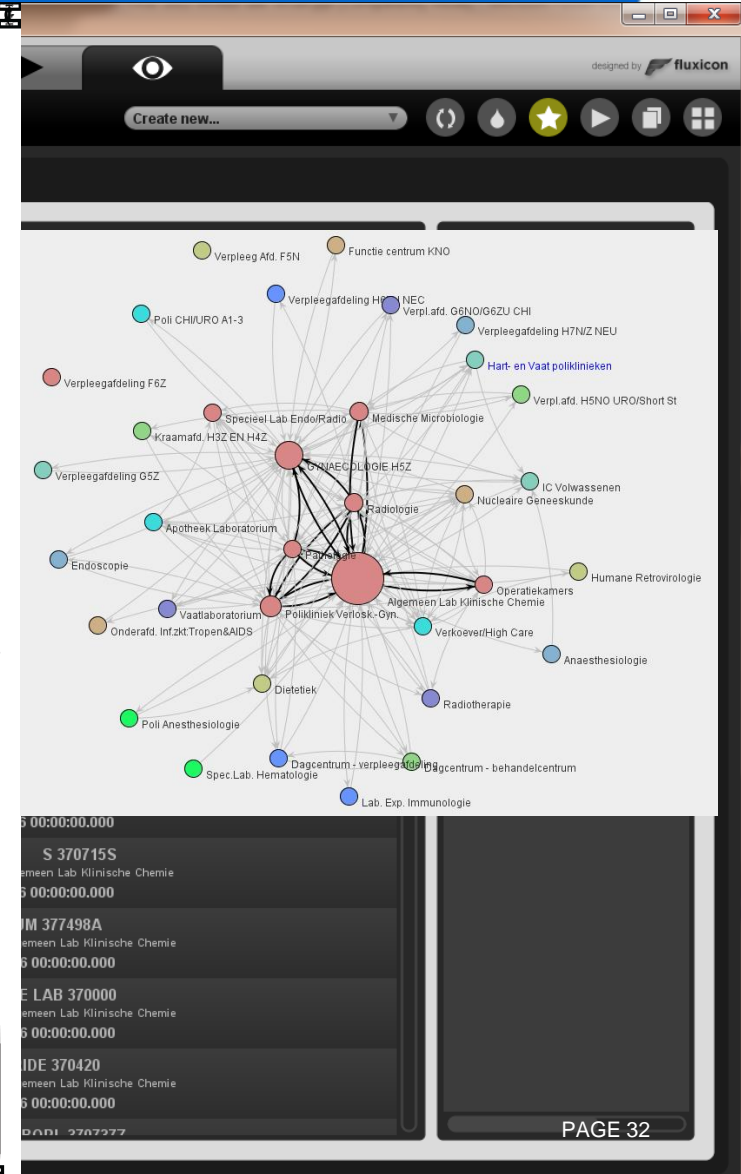
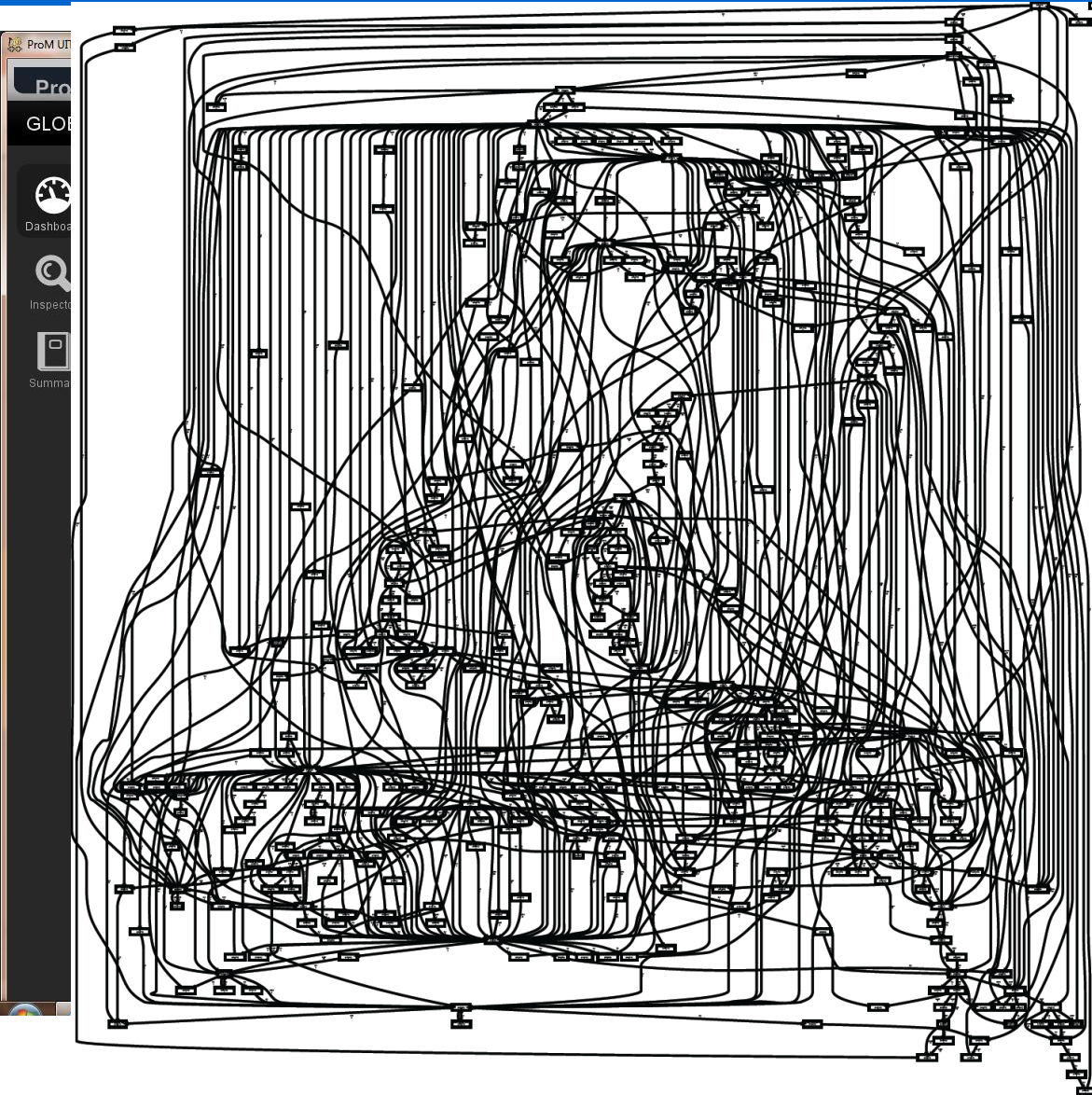
Example Process Discovery

(ASML, test process lithography systems, 154966 events)

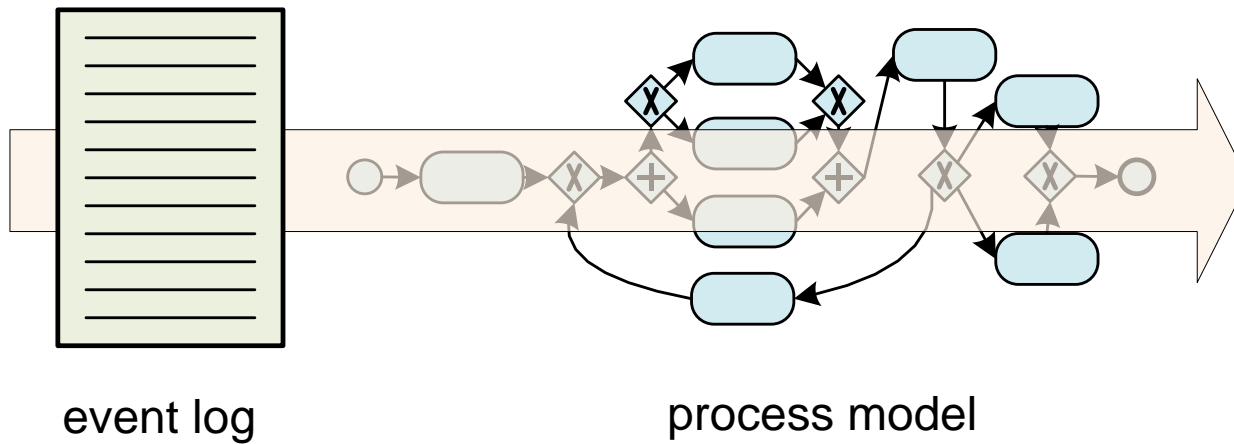


Example Process Discovery

(AMC, 627 gynecological oncology patients, 24331 events)



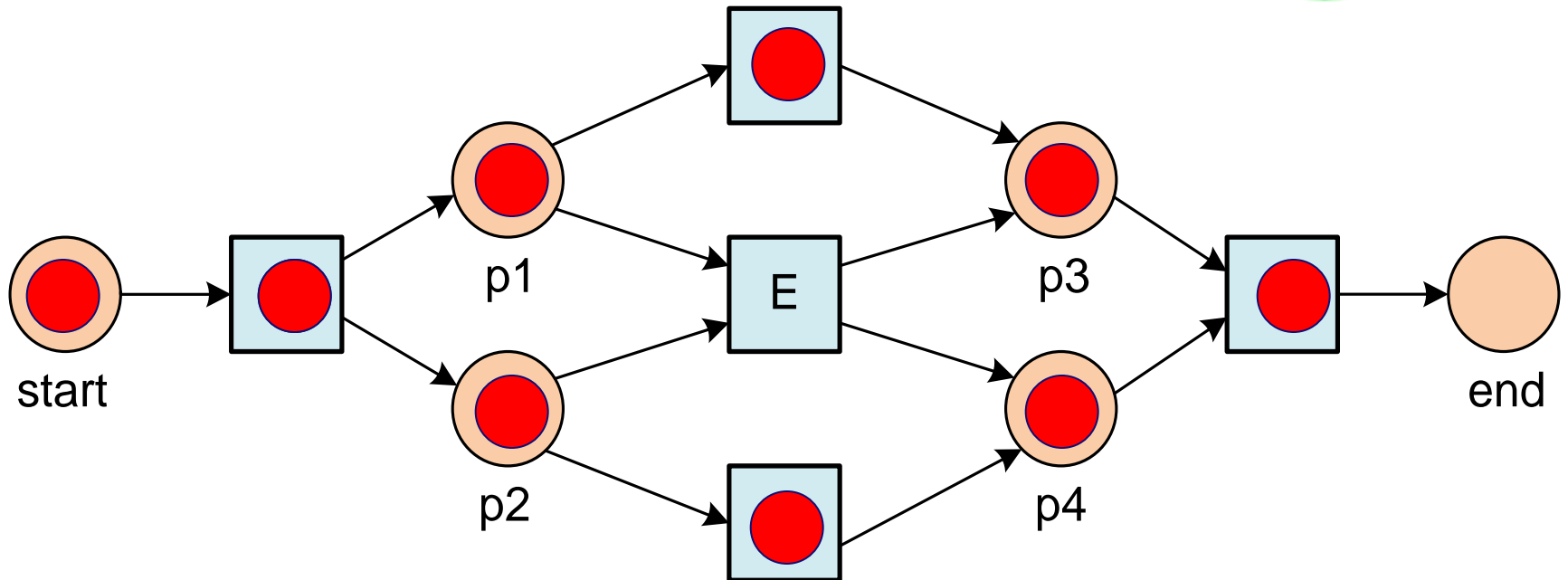
Replay



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

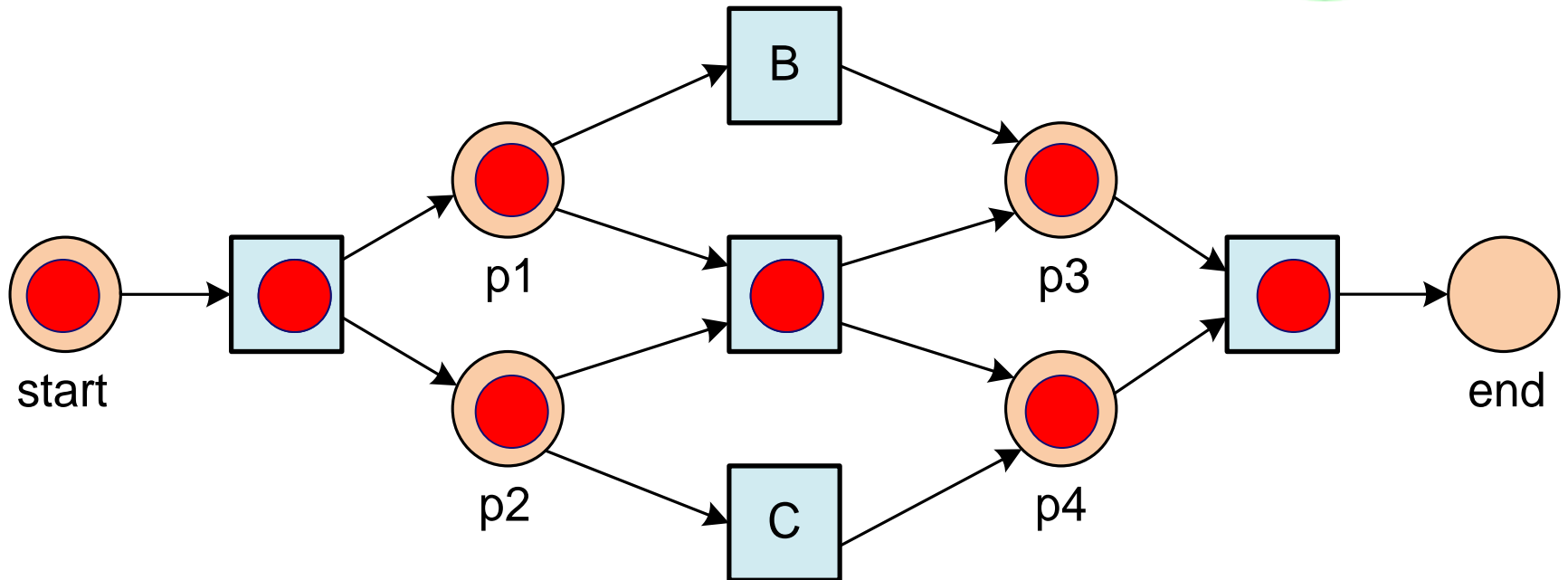
Replay

A B C D



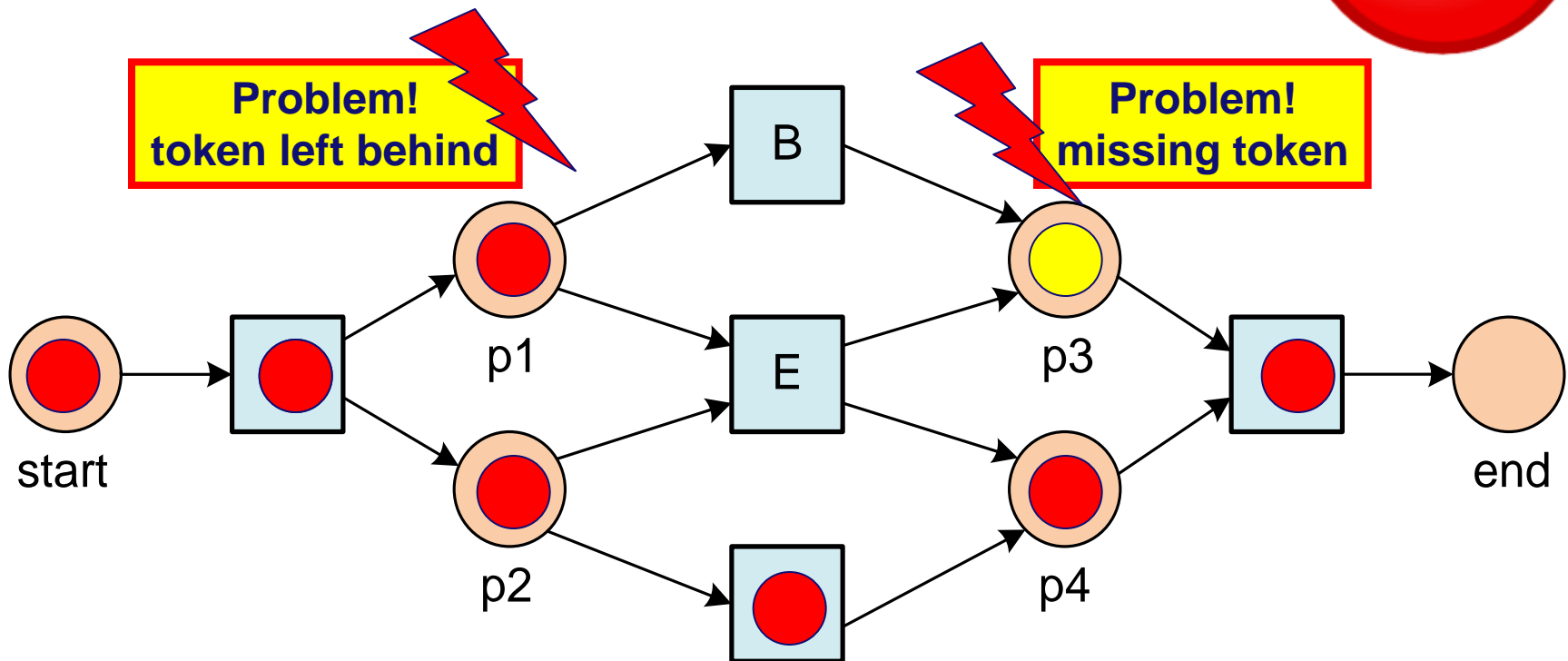
Replay

A E D

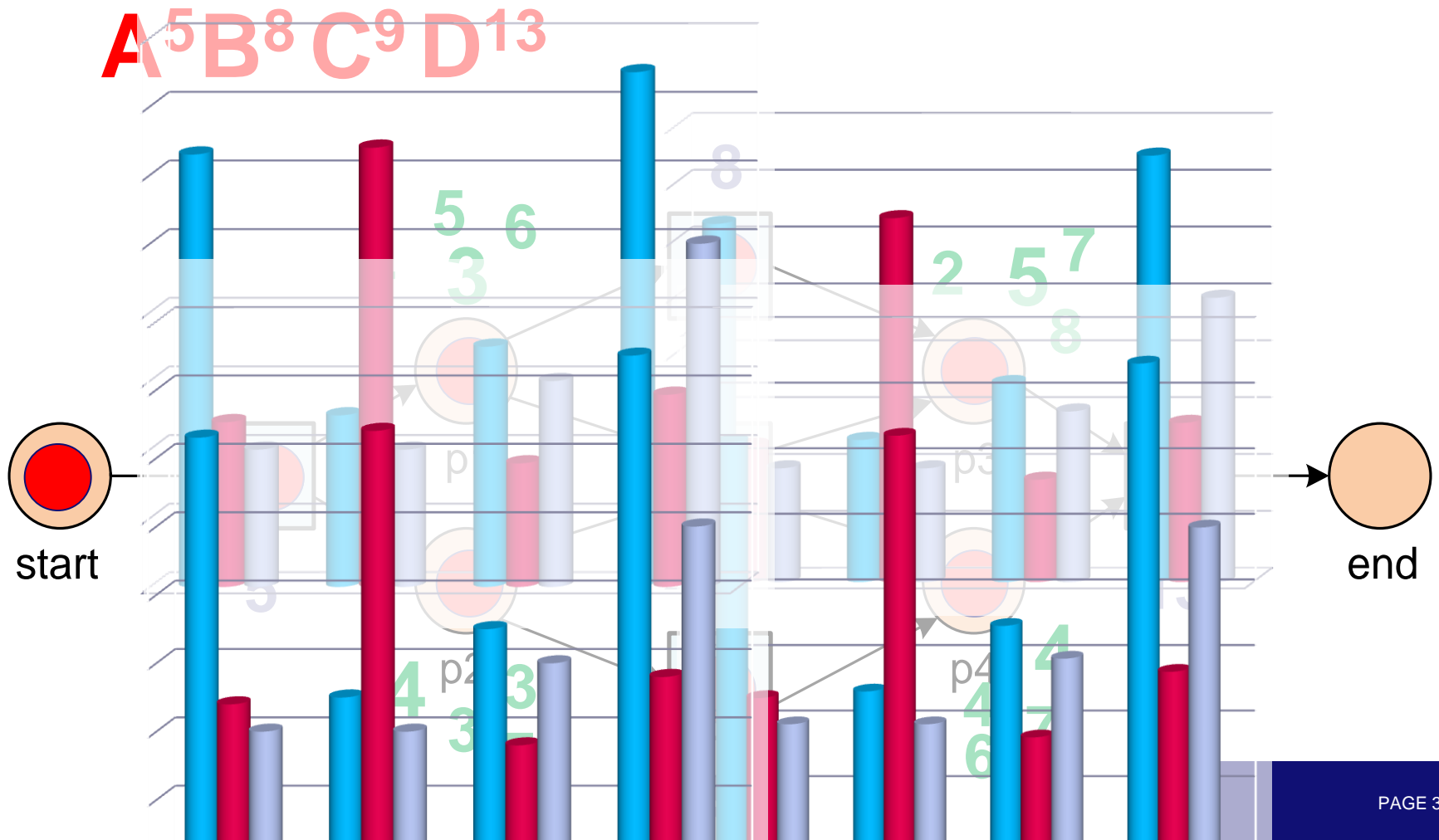


Replay can detect problems

ACD

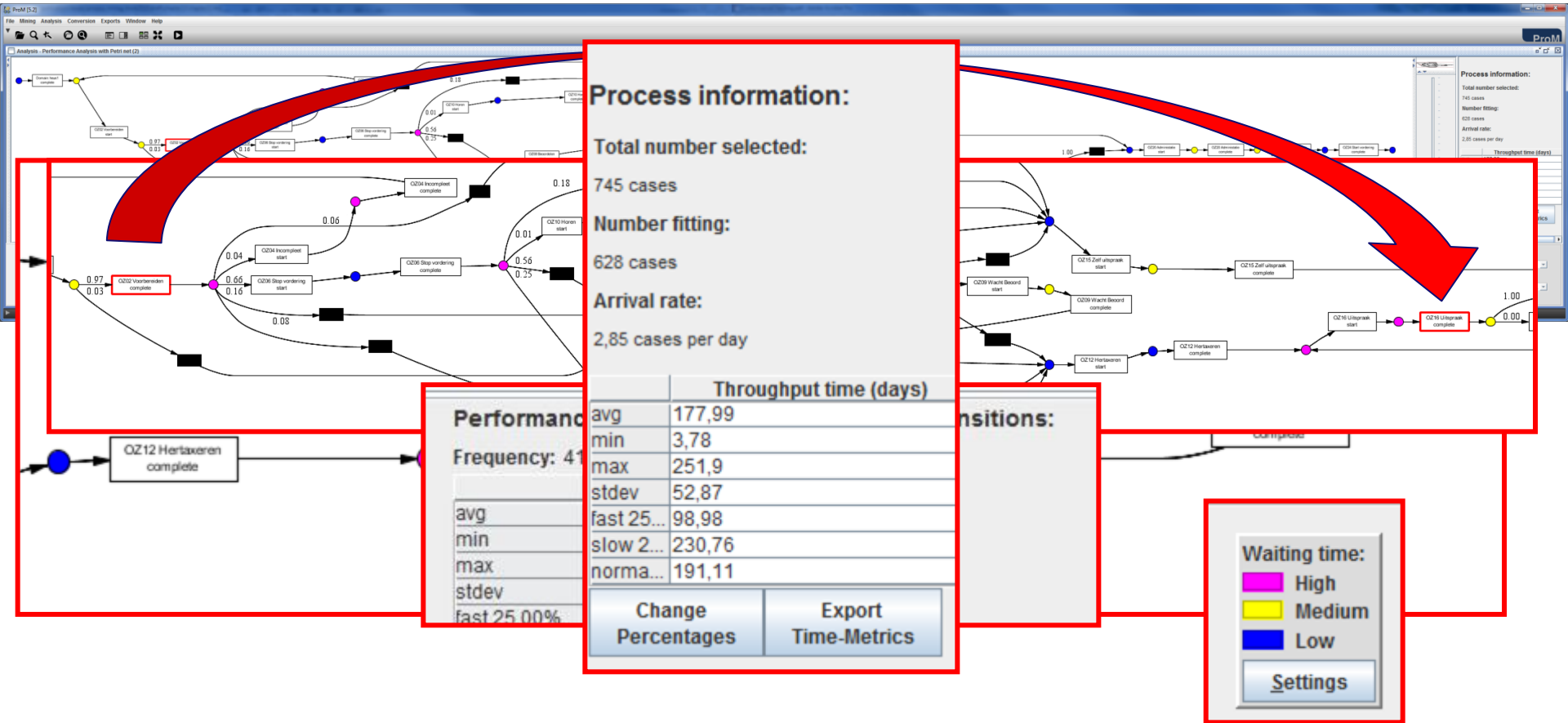


Replay can extract timing information

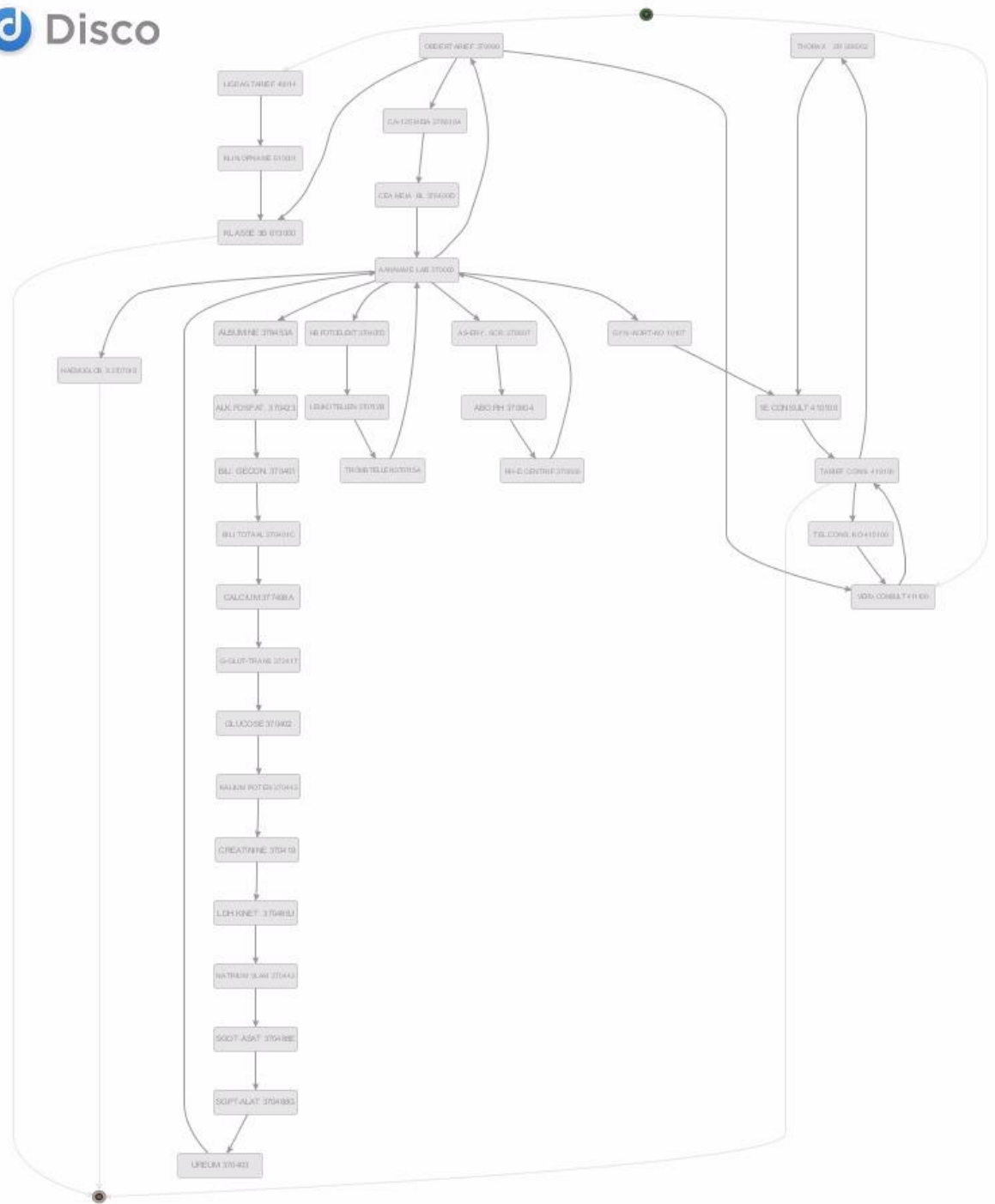
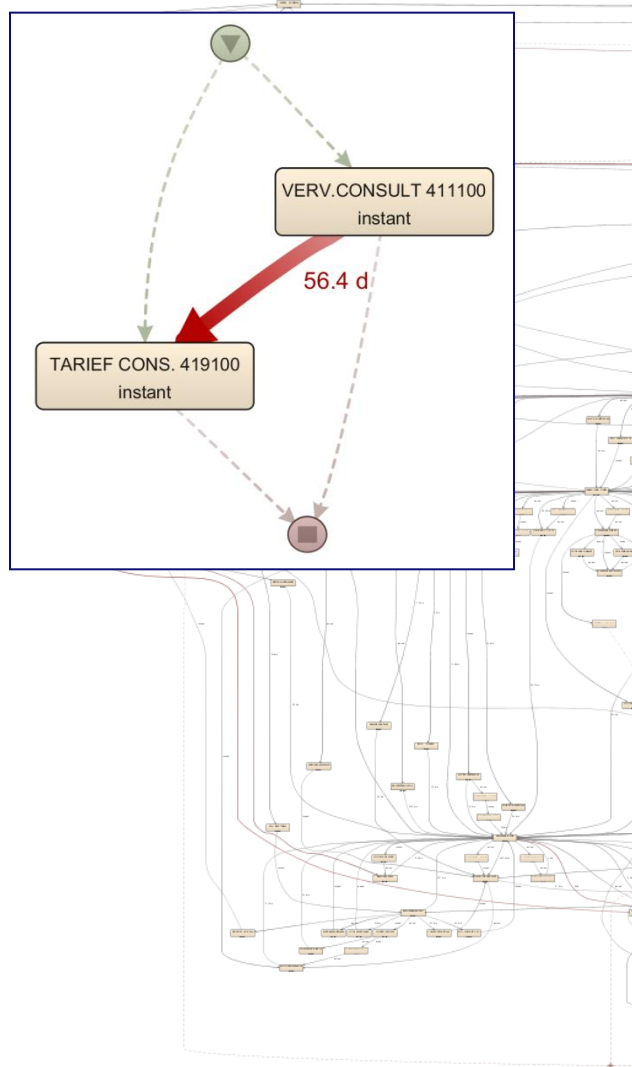


Performance Analysis Using Replay

(WOZ objections Dutch municipality, 745 objections, 9583 event, f= 0.988)



Disco Demo



Models are like the glasses required to see and understand event data!



Evidence-based BPM and Auditing

Process Mining:
The missing link

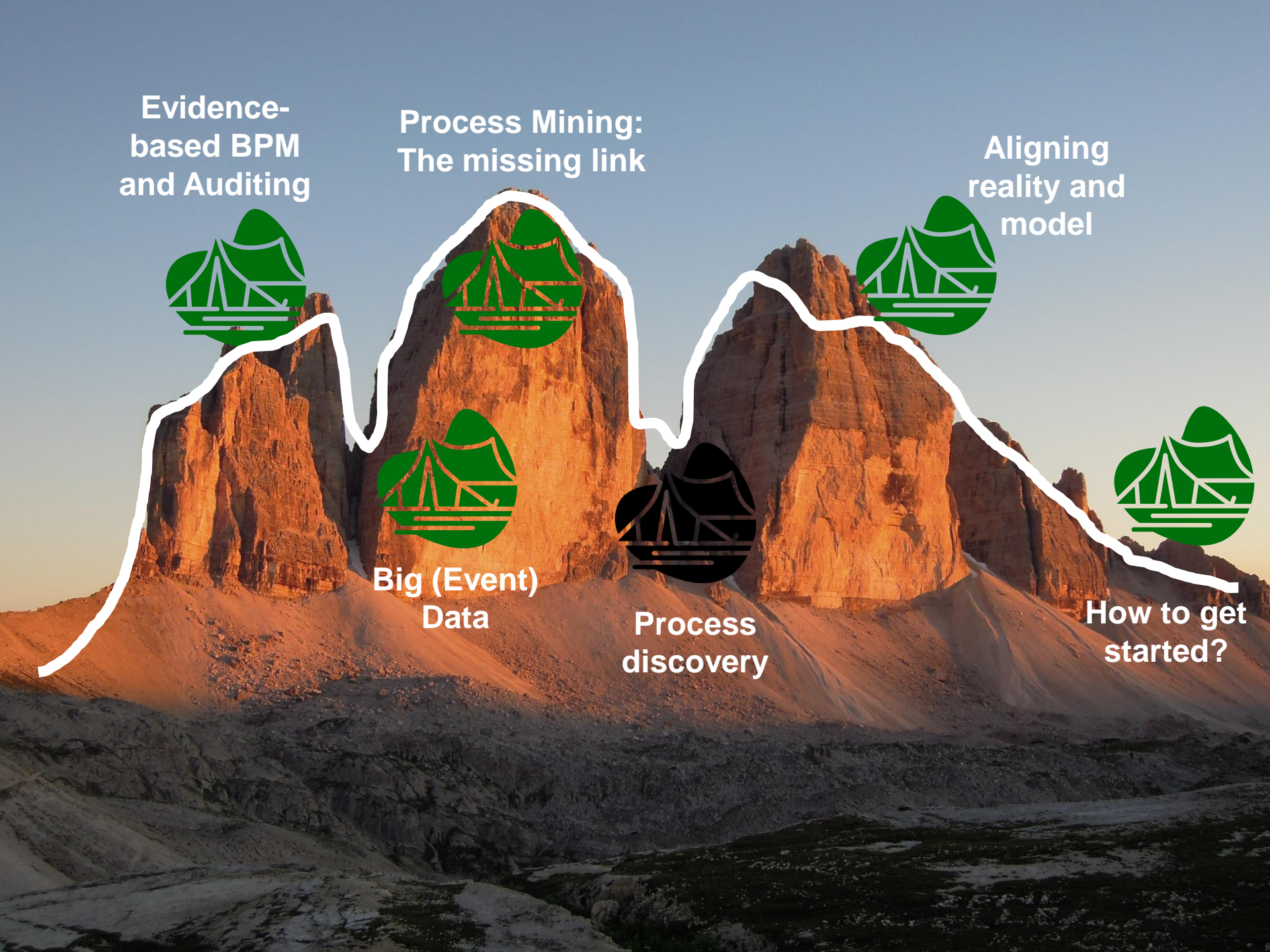
Aligning reality and model



Big (Event)
Data

Process
discovery

How to get
started?



Language identification in the limit (Mark Gold 1967)



A language is **learnable in the limit** if there exists a perfect child that generates only finitely many hypotheses.

Learning is not easy ...



- Even simple languages are not regular languages
- Many languages are not learnable
- Most languages are not learnable from examples, with or without negative examples, etc.

Classical approaches (before 1998) did not consider concurrency and definitely not end-to-end business process models.

reference \cong trace in event log
language \cong process model

Process discovery algorithms (small selection)

automata-based learning

distributed genetic mining

heuristic mining

language-based regions

genetic mining

state-based regions

stochastic task graphs

LTL mining

ETM genetic algorithm

Inductive Miner (infrequent)

fuzzy mining

neural networks

mining block structures

hidden Markov models

α algorithm

multi-phase mining

conformal process graph

$\alpha\#$ algorithm

partial-order based mining

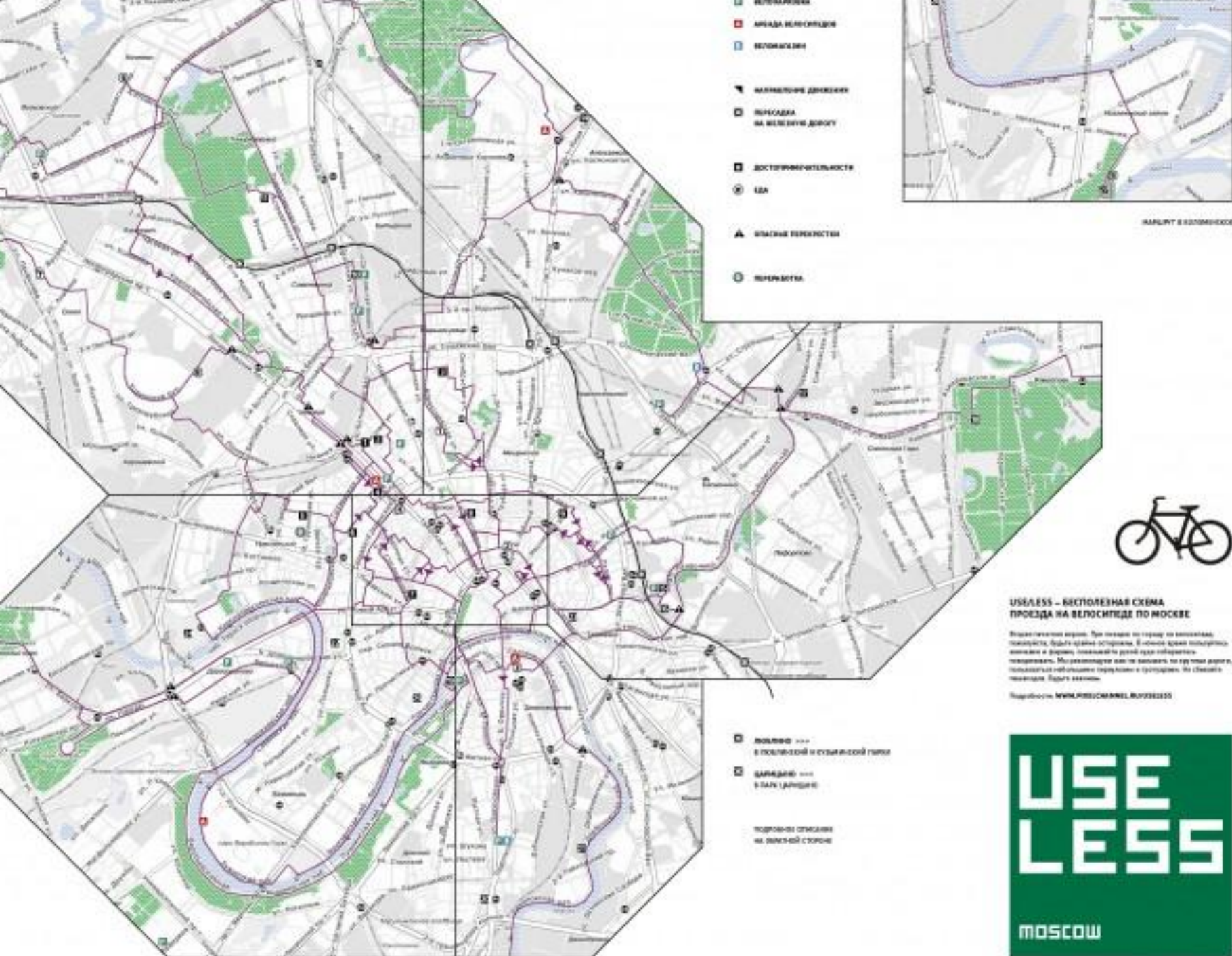
ILP mining

$\alpha++$ algorithm





MINE your MAP



- ВЕЛОПЕШЕД
- АРКАДА ВЕЛОПЕШЕД
- ВЕЛОМОСКИ
- ▼ МАЛЫЕ ПЕШЕХОДНЫЕ ДВИЖЕНИЯ
- ПЕРЕКРЕСТКИ НА ВЕЛОСИДНОЙ ДОРОЖКЕ
- ДОСТОЙНИМЫЕ
- ЦА
- ▲ ПЛОСКИЕ ПЕРЕКРЕСТКИ
- ПЕРЕКРЕСТКИ



КАРТА МОСКОВСКОЙ ОБЛАСТИ



USELESS – ВЕЛОСИДНАЯ СХЕМА ПЕШЕХОДОВ НА ВЕЛОСИДНОЙ ДОРОЖКЕ ПО МОСКВЕ

Визуальное решение. При проектировании велосипедных дорожек, важно учитывать не только их функциональность, но и форму, которая должна быть удобной для пешеходов. Это решение было разработано специально для Москвы. Оно учитывает все требования к велосипедным дорожкам и пешеходным тротуарам. На сайте проекта можно найти все детали. Проект реализован.

Сайт проекта: WWW.USELESS.MOSCOW

- ПОЛНОСТЬЮ – в пешеходной и велосипедной дорожке
- ЧАСТИЧНО – в зоне (длина)

подробные описания на обратной стороне



Evidence-based BPM and Auditing



Process Mining:
The missing link



Aligning reality and model



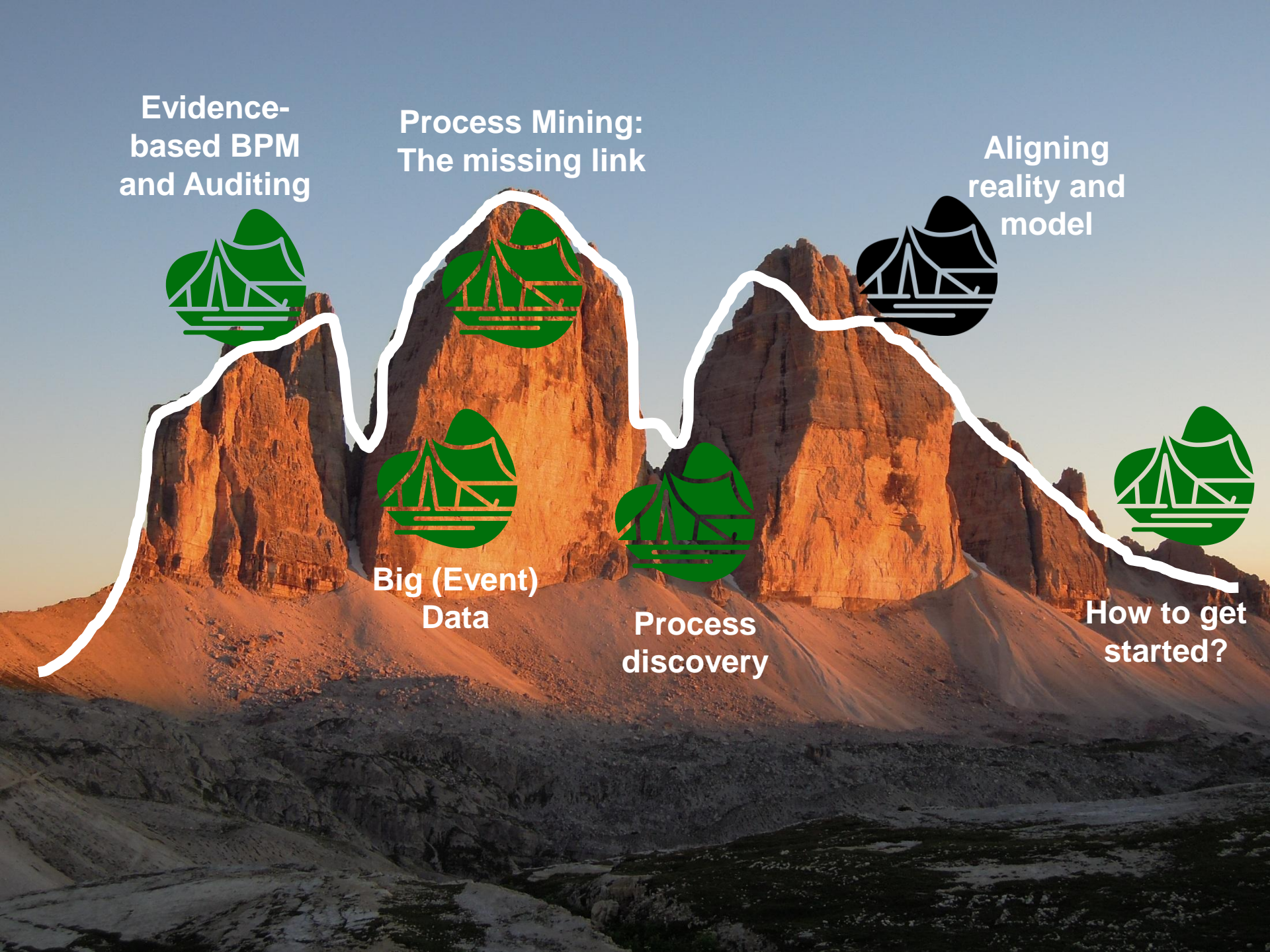
Big (Event) Data



Process discovery



How to get started?



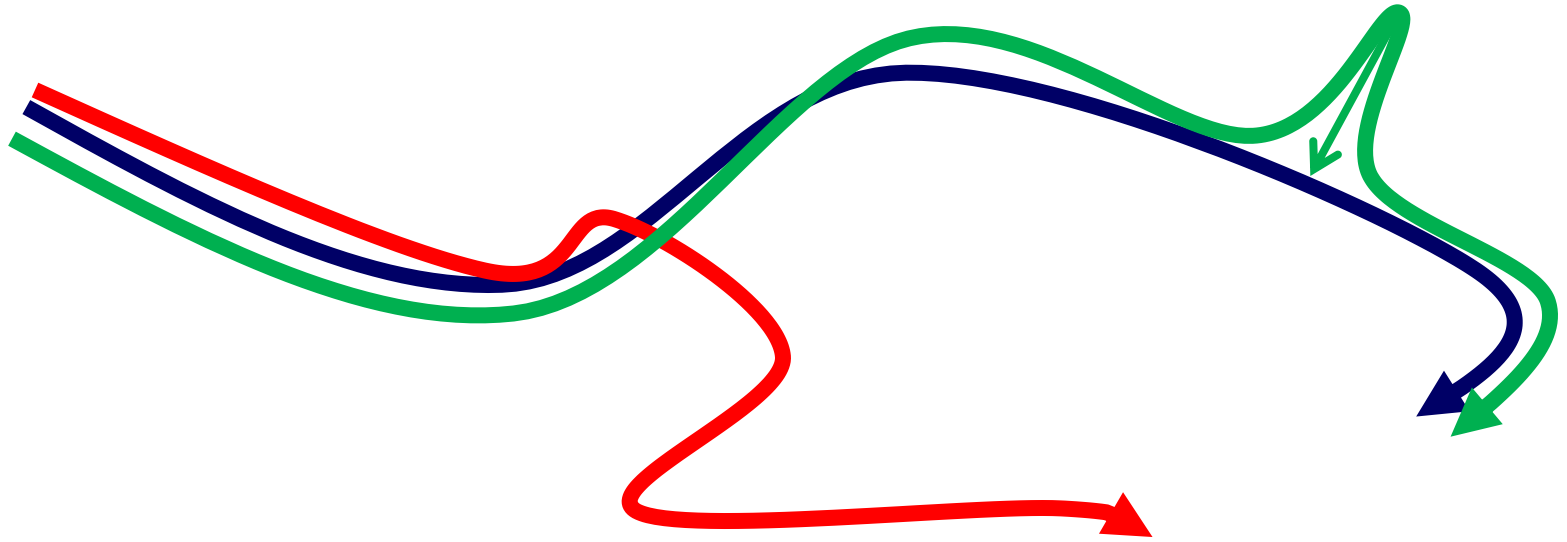
Conformance Checking

The image shows a screenshot of a Microsoft Word document titled "hello world.docx". The document text is "Recent breakthroughs in process mining research make it possible to discover, analyze, and improve business processes based on event data. Activities such as people, machines, and software leave trails. Events such as entering a customer order into SAP, checking in for a flight, and rejecting a building for a patient, and rejecting a building for a patient, are common. Over the years, there has been a spectacular growth of data. Moreover, the digital universe and the physical universe has becoming more and more aligned." The text is annotated with five yellow callout boxes:

- Top left: "an activity that should not happen happened" (points to "analyze")
- Top right: "an activity was executed by the wrong person" (points to "people")
- Middle: "an activity was executed too late" (points to "rejecting a building")
- Bottom left: "an activity that should happen did not happen" (points to "rejecting a building")
- Bottom center: "two activities were swapped" (points to "rejecting a building" and "for a patient")

Page: 1 of 1 | Words: 95 | English (U.S.) | 103%

Alignments are essential!



- conformance checking to diagnose deviations
- squeezing reality into the model to do model-based analysis

<i>a</i>	<i>c</i>	\gg	<i>d</i>	\gg	<i>f</i>	\gg
<i>a</i>	<i>c</i>	<i>b</i>	<i>d</i>	τ	\gg	<i>h</i>
<i>t1</i>	<i>t4</i>	<i>t3</i>	<i>t5</i>	<i>t7</i>		<i>t10</i>

process
model

event log

synchronous
move

<i>a</i>	<i>c</i>	\gg	<i>d</i>	\gg	<i>f</i>	\gg
<i>a</i>	<i>c</i>	<i>b</i>	<i>d</i>	τ	\gg	<i>h</i>
<i>t1</i>	<i>t4</i>	<i>t3</i>	<i>t5</i>	<i>t7</i>		<i>t10</i>

move on
model only

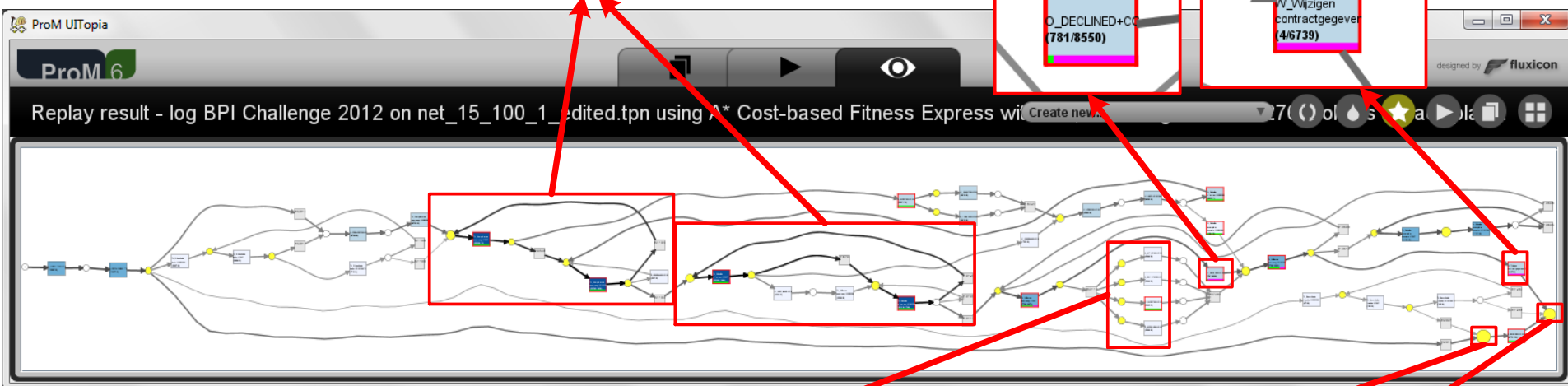
move on log
only

Example: BPI Challenge 2012

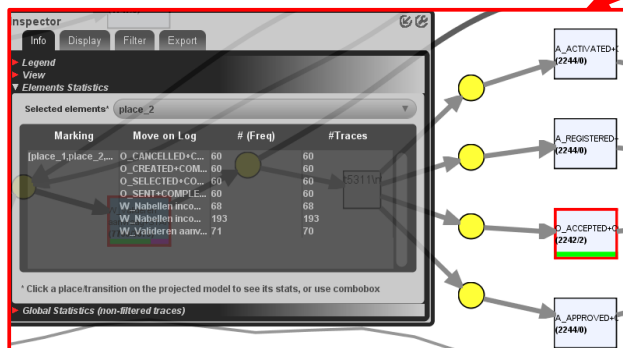
(Dutch financial institute, doi:10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f)

Loops of “W_Completeren aanvraag” and “W_Nabellen offertes” are often performed

“O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped



Many moves on log of “O_CANCELLED”, “O_CREATED”, “O_SELECTED”, “O_SENT” occurred with the same frequency value (i.e. 60) before parallel branch



Marking	Move on Log	# (Freq)	#Traces
[place_11]	A_ACCEPTED+COMPLETE	19	19
	A_PREACCEPTED+COMPLETE	481	481
	W_Afhandelen leads+COMPLETE	2431	2431
	W_Afhandelen leads+SCHEDULE	2431	2431
	W_Completeren aanvraag+COMPLETE	67	67
	W_Completeren aanvraag+SCHEDULE	481	481
	W_Completeren aanvraag+START	578	481

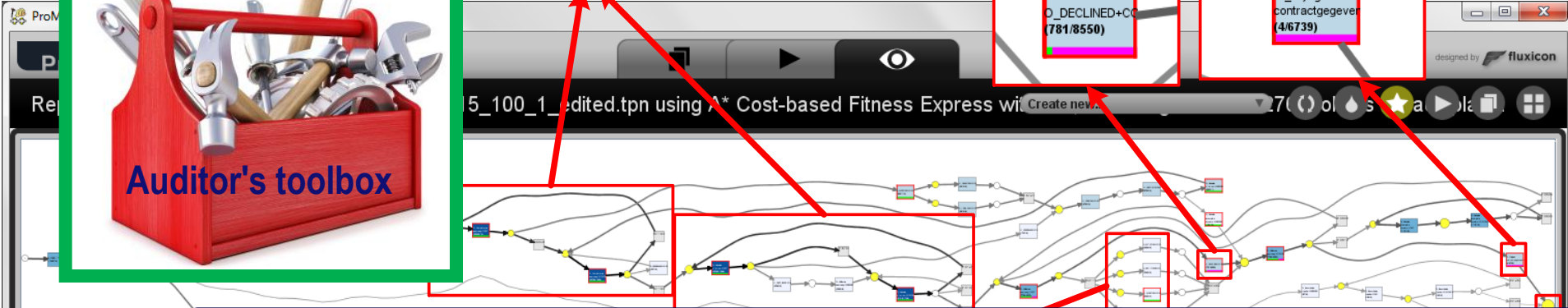
Marking	Move on Log	# (Freq)	#Traces
[place_42]	A_ACCEPTED+COMPLETE	16	16
	A_CANCELLED+COMPLETE	1087	1087
	A_DECLINED+COMPLETE	89	89
	A_PREACCEPTED+COMPLETE	156	156
	O_CANCELLED+COMPLETE	524	524
	O_DECLINED+COMPLETE	24	24
	W_Afhandelen leads+COMPLETE	2233	2225

Many moves on log of “W_Afhandelen leads” (> 2200 times) occurred in the end of traces



Loops of “W_Completeren aanvraag” and “W_Aan offertes” are often performed

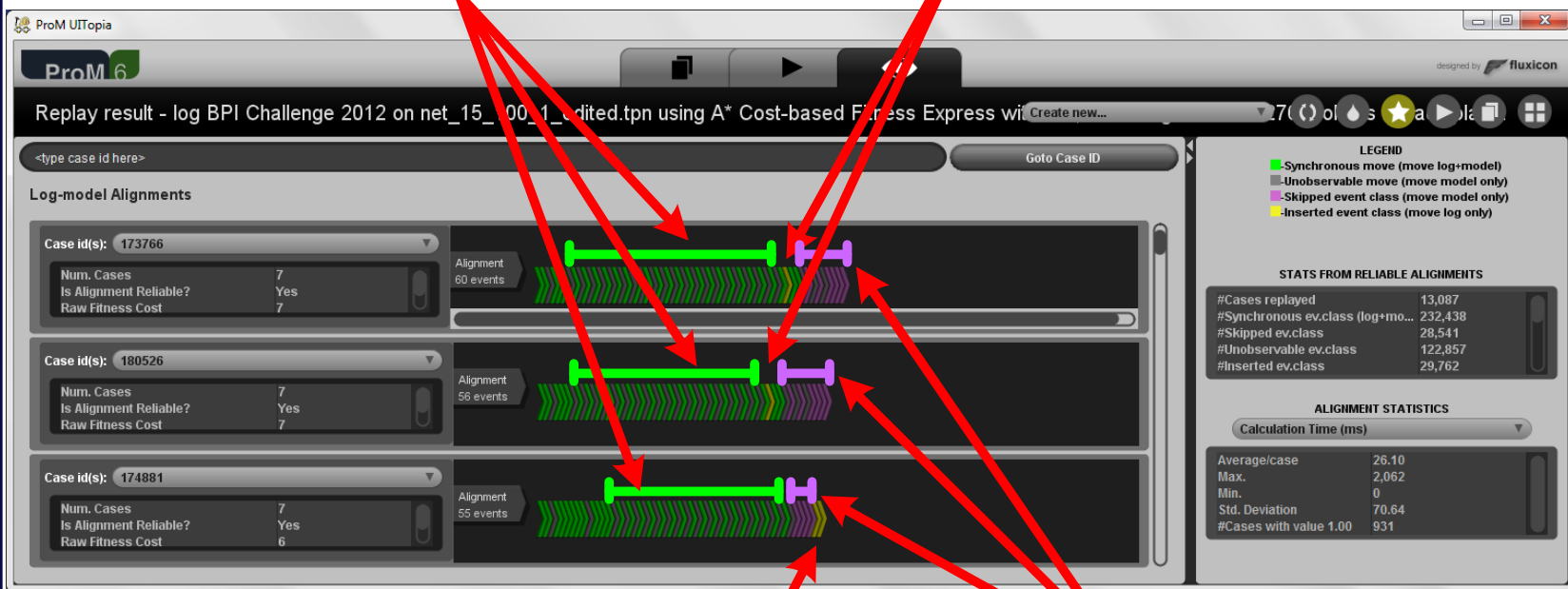
“O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped



Synchronous moves of “Completeren aanvraag”

Move on log of “Completeren aanvraag”

Many moves of “O_CANCELLED”, “O_CREATED”, “O_SELECTED”, “O_SENT” or “O_COMPLETED” with the frequency variable (value 60) before

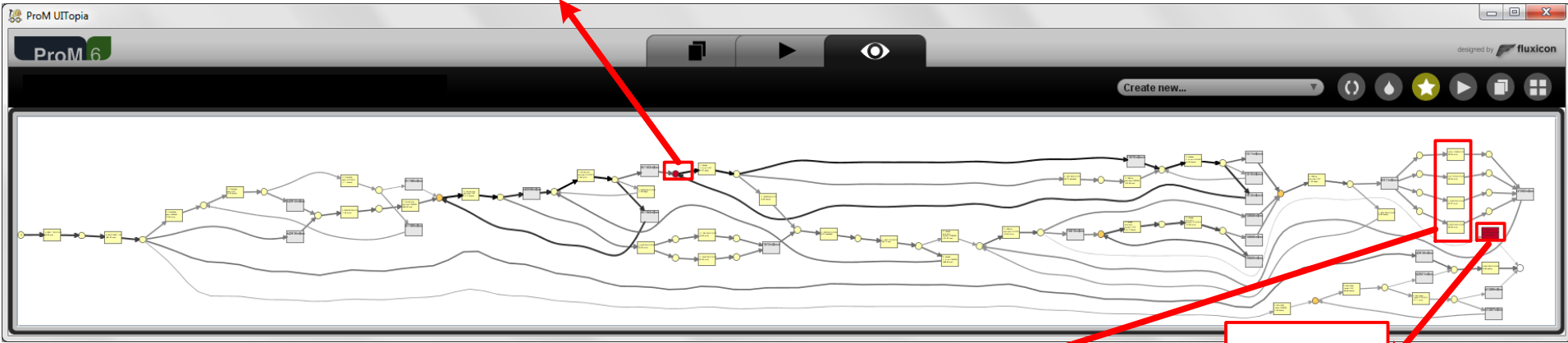


Move on log of “O_CANCELLED” and “A_CANCELLED”

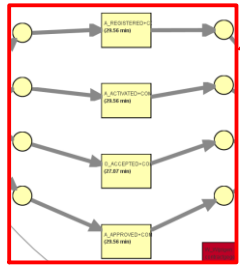
Moves on model towards end of traces

Property	Min.	Max.	Avg.	Std. Dev	Freq.
Waiting time	0.00 ms	29.78 days	2.83 days	3.30 days	24,229
Synchronization time	0.00 ms	0.00 ms	0.00 ms	0.00 ms	24,229
Sojourn time	0.00 ms	29.78 days	2.83 days	3.30 days	24,229

The average waiting time for the input place of "W_Nabellen offertes+START" is very long (2.83 days) compares to the average waiting time of other places



"O_ACCEPTED" has average sojourn time of 27.07 minutes, while "A_REGISTERED", "A_ACTIVATED", and "A_APPROVED" have average sojourn time of 29.56 minutes

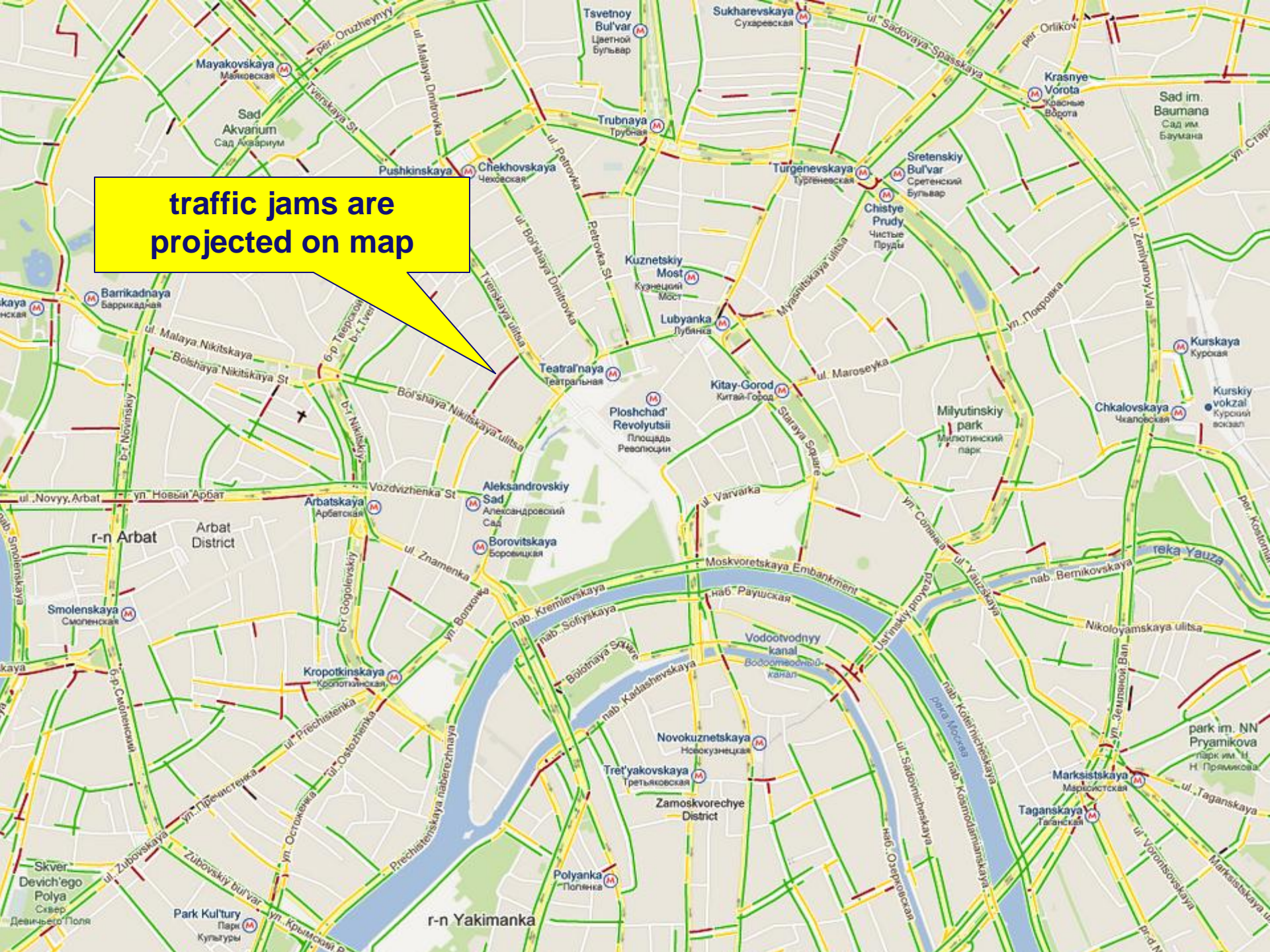


Property	Min.	Max.	Avg.	Std. Dev	Freq.
Throughput time	0.00 ms	0.00 ms	0.00 ms	0.00 ms	4
Waiting time	1.55 hours	3.43 months	1.14 months	1.55 months	4
Sojourn time	1.55 hours	3.43 months	1.14 months	1.55 months	4
#Unique cases ...	4				

Activity "W_Wijzigen contractgegevens" is the bottleneck, but it occurred rarely (only 4 times)



traffic jams are projected on map



Demand TomTom!
Do not settle for restrictive
information systems and
static process models

recommend:
turn right

predict: when
will I be home

adapt: use real-
time traffic
information



Evidence-based BPM and Auditing



Process Mining:
The missing link



Aligning
reality and
model



Big (Event)
Data



Process
discovery



How to get
started?



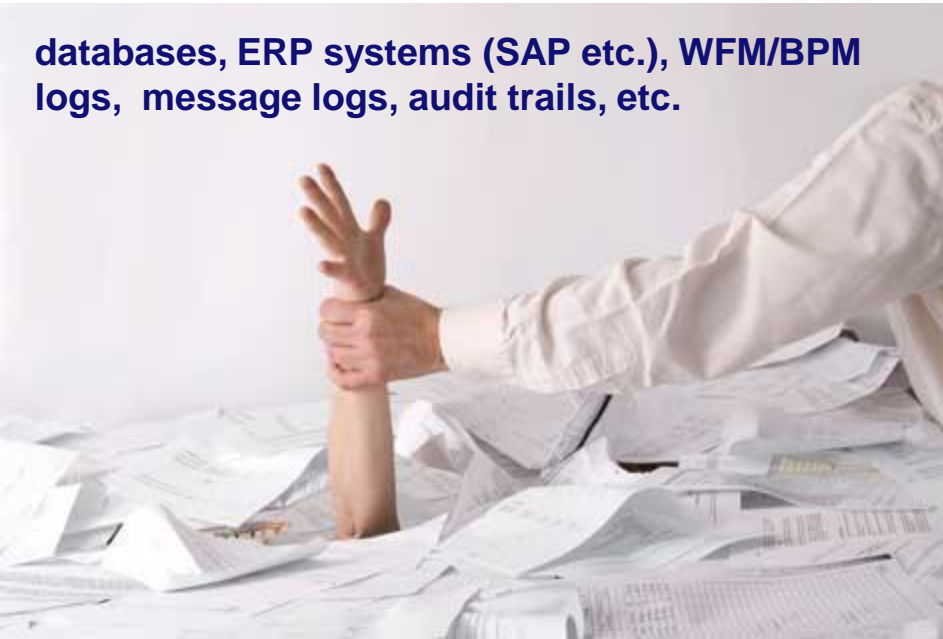
How to get started?



Collect data: Events are everywhere!



databases, ERP systems (SAP etc.), WFM/BPM logs, message logs, audit trails, etc.

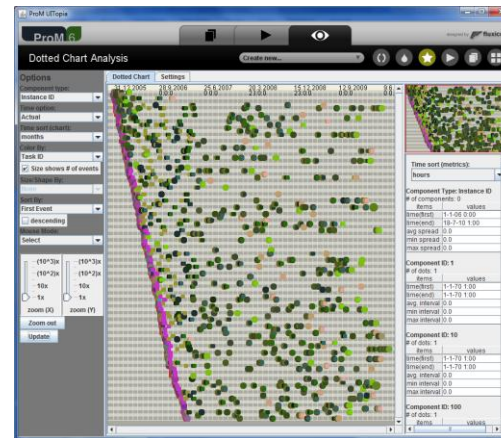
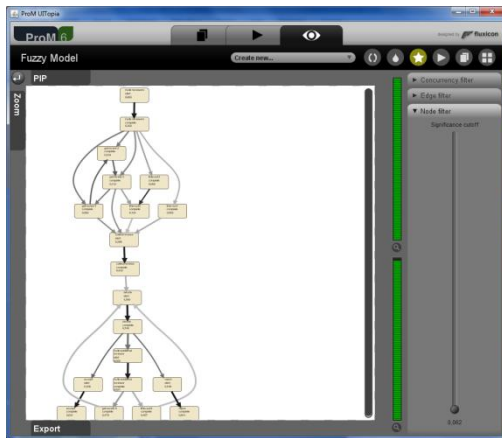


- **Minimal requirement:** events referring to an activity name and a process instance.
- **Good to have:** timestamps, resource information, additional data elements.
- **Challenges:** scoping and sometimes correlation.

Get at a process mining tool, e.g., ProM



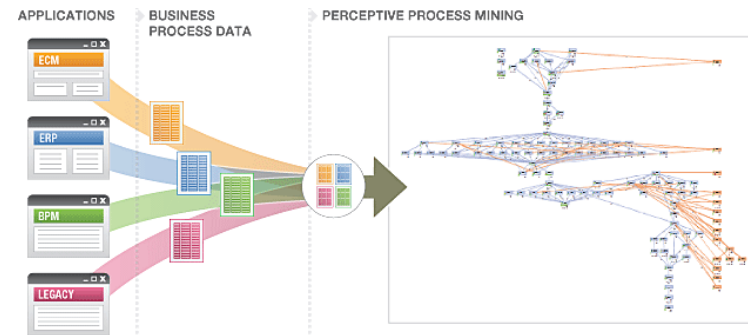
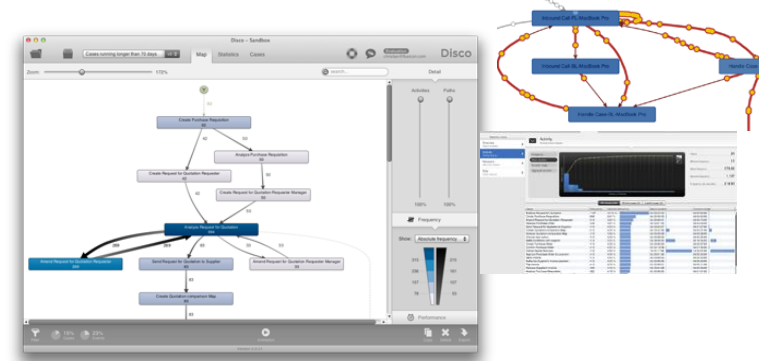
600+ plug-ins available covering the whole process mining spectrum



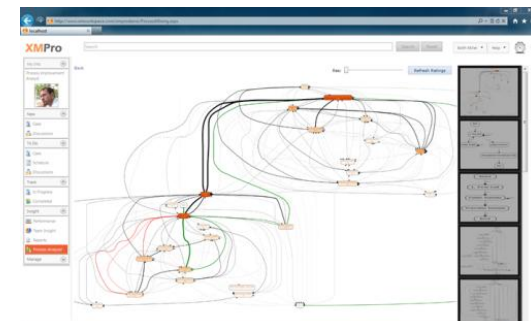
Download from: www.processmining.org

Commercial process mining tools

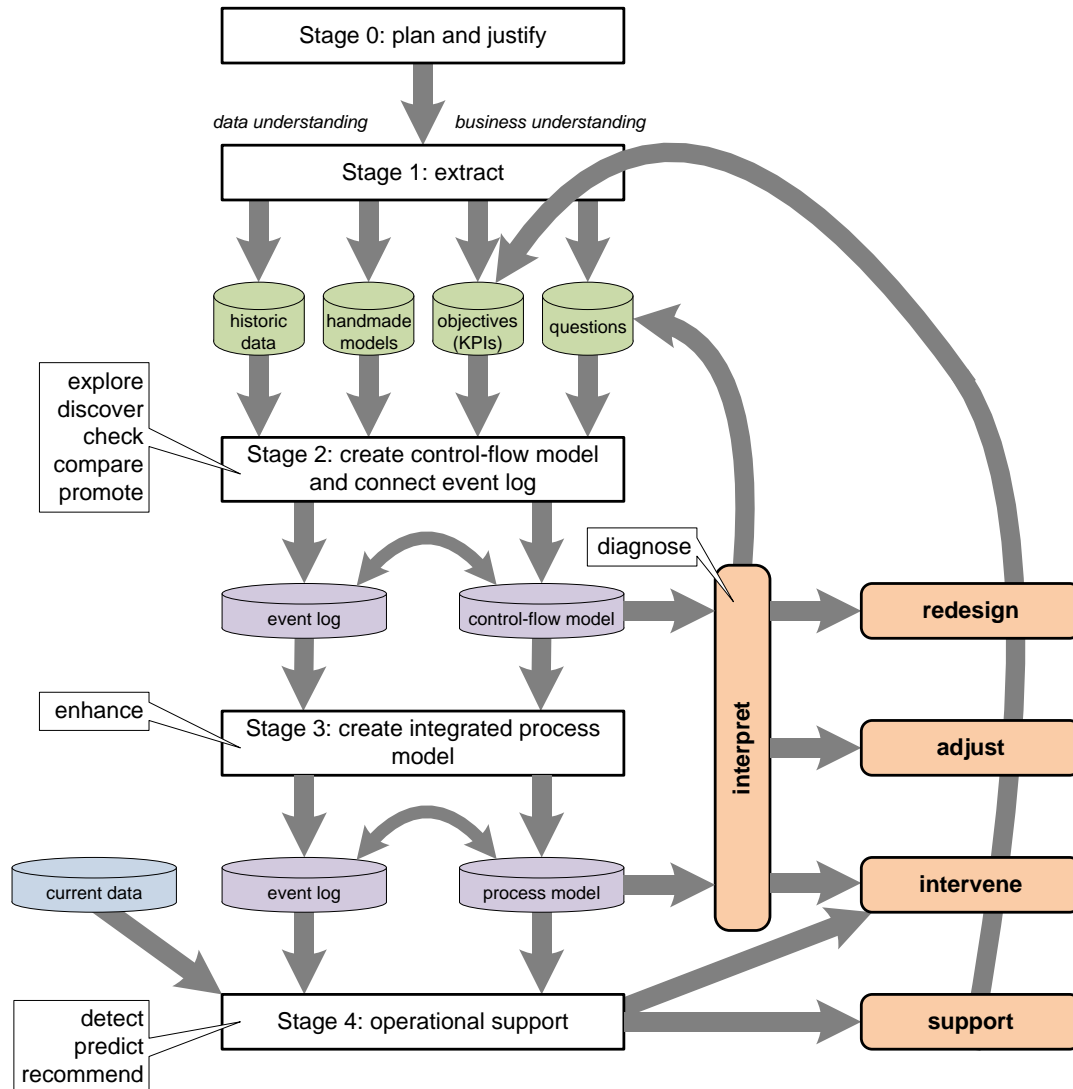
- **Disco** (Fluxicon)
- **Perceptive Process Mining** (before Futura Reflect and BPM|one)
- ARIS Process Performance Manager
- QPR ProcessAnalyzer
- Celonis Discovery
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAalyzer (XMPro)
- ...



Example of a dedicated process mining consulting firm: **ProcessGold AG**.



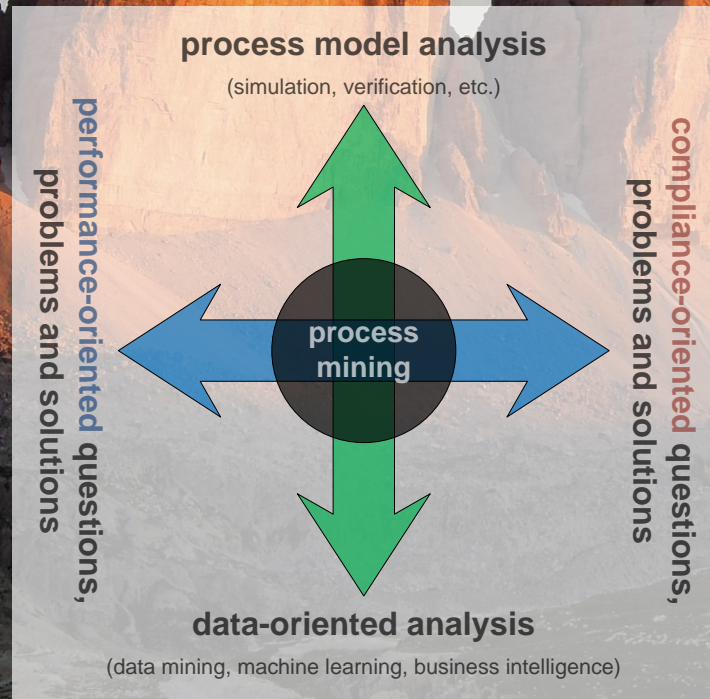
Approach: Start simple



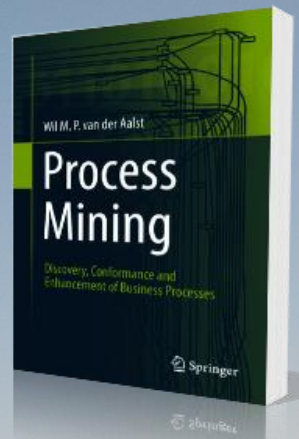
Questions:

- What kind problems would you like to address (cost, time, risk, compliance, service, etc.)?
- Related to discovery, conformance, enhancement?
- Iterative process: can be “curiosity driven” initially.

Join our expedition: Mine your processes!



Learn more?



Introduction to Process Mining: Turning (Big) D...

90% OF THE APPLICATIONS TAKE MORE THAN 9 DAYS

Process Mining Movie

PAIS lab International Laboratory of Process-Aware Information Systems (PAIS Lab)

Printout

- About
- Staff
- Research
- Links
- Tools
- Science seminar
- Events
- Press about us
- Publications

Search

[printer-friendly version](#)

Processes are everywhere and without information systems supporting these processes society would come to a grinding halt. When you rent a car, book a flight, buy a book, file a tax declaration, or transfer money there are process-aware information systems making this possible. A Process-Aware Information System (PAIS) is a software system that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models. Example PAISs are Business Process Management (BPM) systems, Workflow Management (WFM) systems, Enterprise Resource Planning (ERP) systems, and case handling systems. Given the importance of such systems and related analysis techniques the National Research University Higher School of Economics (HSE) created the *International Laboratory of Process-Aware Information Systems*.

The laboratory was established in January 2013 and is supervised by professor [Wim van der Aalst](#), one of the leading computer scientists in the world and the most influential researcher in areas such as business process management and process mining.

The notion of a process model is foundational for PAISs. A process model aims to capture the different ways in which a case (i.e., process instance) can be handled. A plethora of notations exists to model operational business processes (e.g., Petri nets, BPMN, UML, and EPCs). These notations have in common that processes are described in terms of activities (and possibly subprocesses). The ordering of these activities is modeled by describing causal dependencies. Moreover, the process model may also describe temporal properties, specify the creation and use of data, e.g., to model decisions, and stipulate the way that resources interact with the process (e.g., roles, allocation rules, and priorities).

Process mining techniques play a central role in the lab because of the incredible growth of event data. Process mining techniques can be used to extract knowledge from event data, discover models, align logs and models, measure conformance, diagnose bottlenecks, and predict future events. Today's processes leave many trails in data bases, audit trails, message logs, transaction logs, etc. Therefore, it makes sense to relate these event data to process models independent of their particular notation. Process models discovered based on the actual behavior tend to be very different from the process models made by humans. Moreover, conformance checking techniques often reveal important deviations between models and reality.

Traditionally, process models and system specifications tend to be static and disconnected from the real processes and system. Process mining techniques provide a means to establish a direct connection between processes, models, and systems. Moreover, event data can be used to breathe life into process models and unite domain experts, IT experts, managers and users of PAISs.

HSE's International Laboratory of Process-Aware Information Systems aims to address urgent challenges related to business process management, process mining, and information systems development. The laboratory uses a mixture of formal methods (e.g., Petri nets and other models for concurrency), data-driven analysis (data/process mining), and systems engineering.

http://www.youtube.com/watch?v=7oat7MatU_U

<http://pais.hse.ru/>