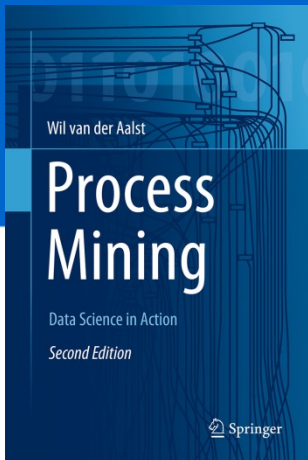


*Big Software on the Run: Where Software Meets Data*  
*Winter School, October 2016, Ede, NL*

# Making Sense From Software Using Process Mining

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



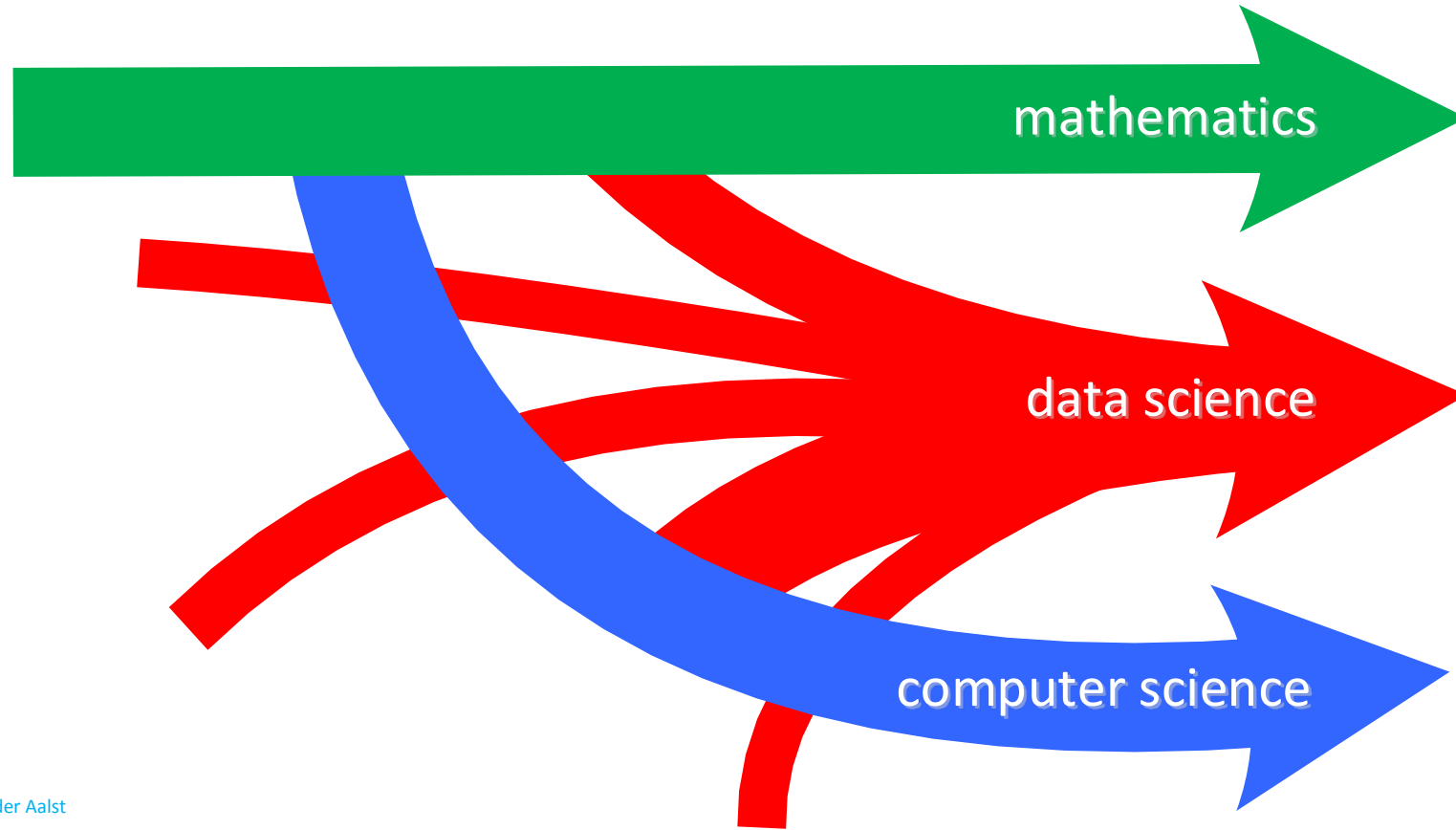
**TU/e**

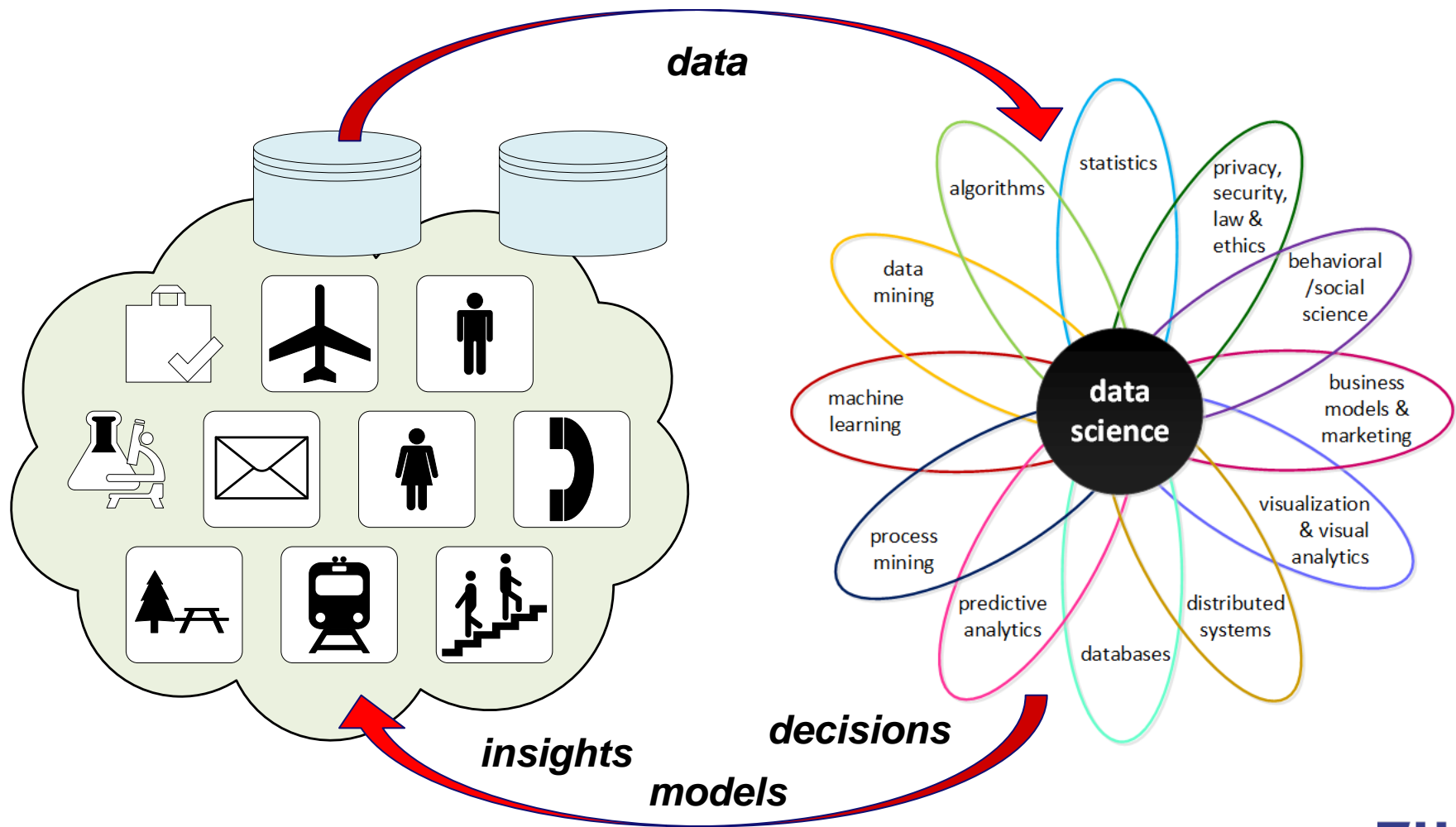
Technische Universiteit  
**Eindhoven**  
University of Technology

**Where innovation starts**

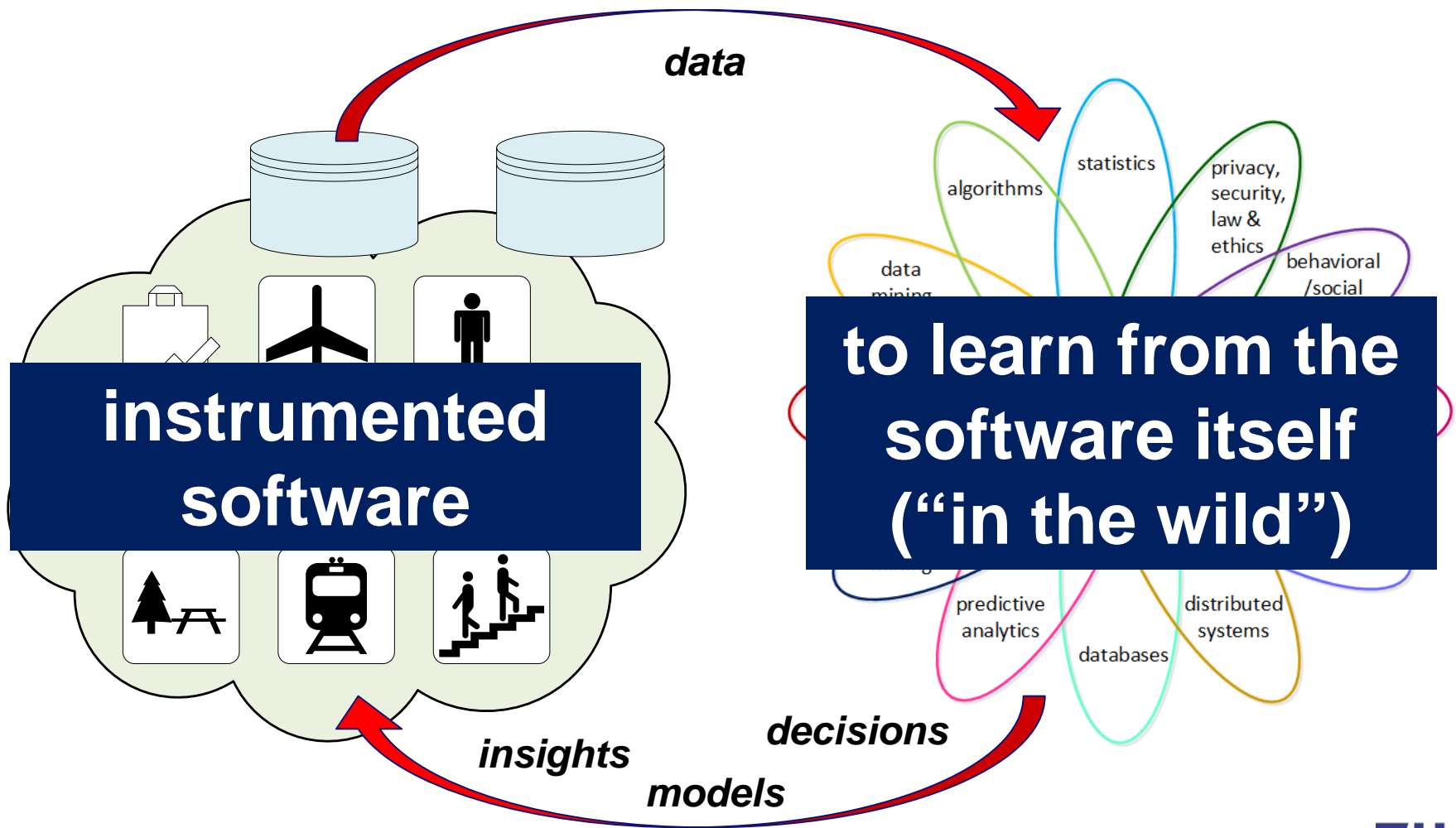
*The bigger picture*

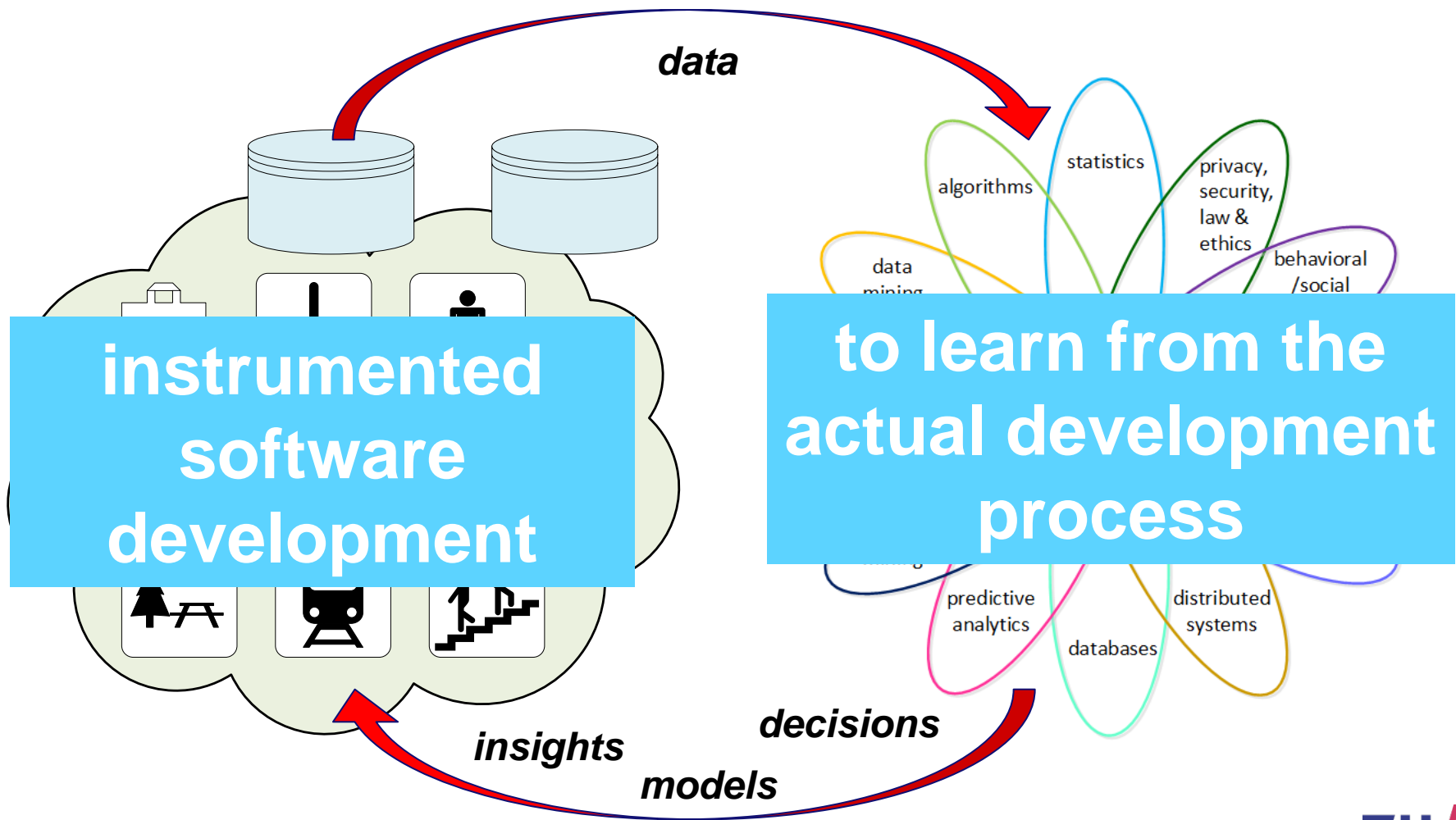
# A new discipline ...



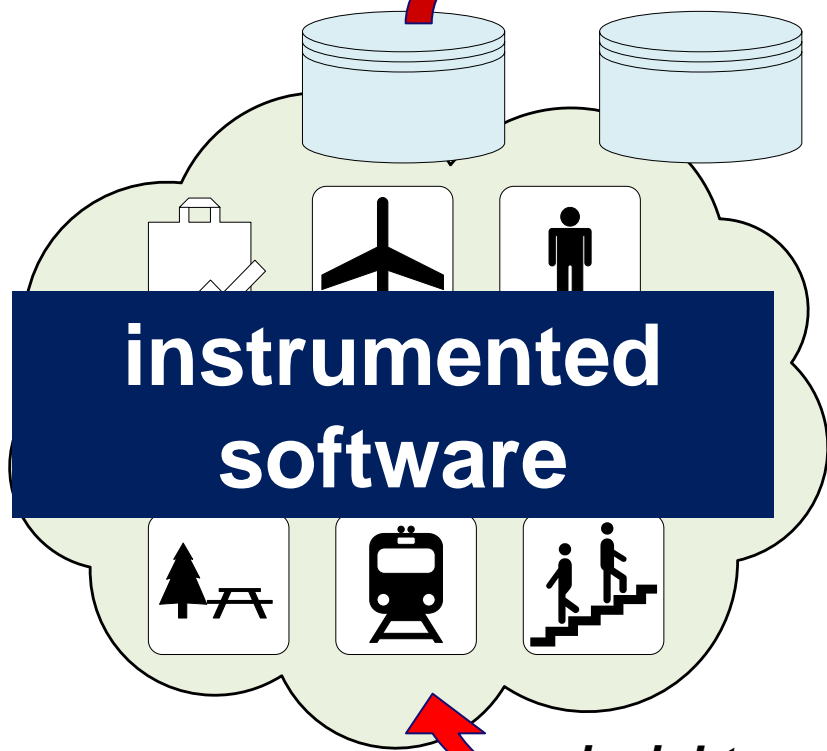






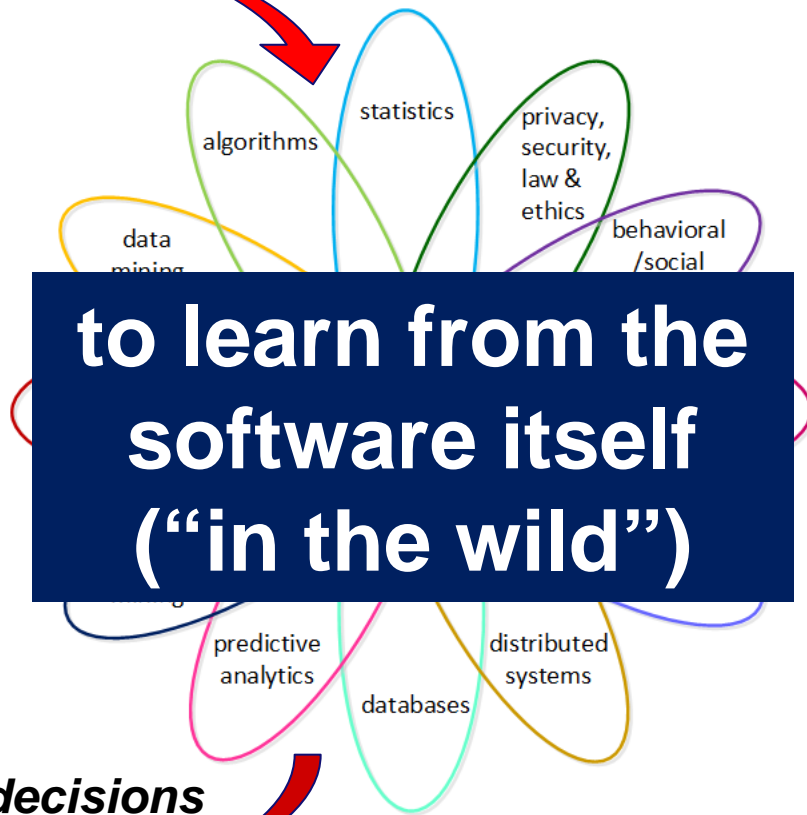


# Focus BSR



**instrumented  
software**

*data*



**to learn from the  
software itself  
("in the wild")**

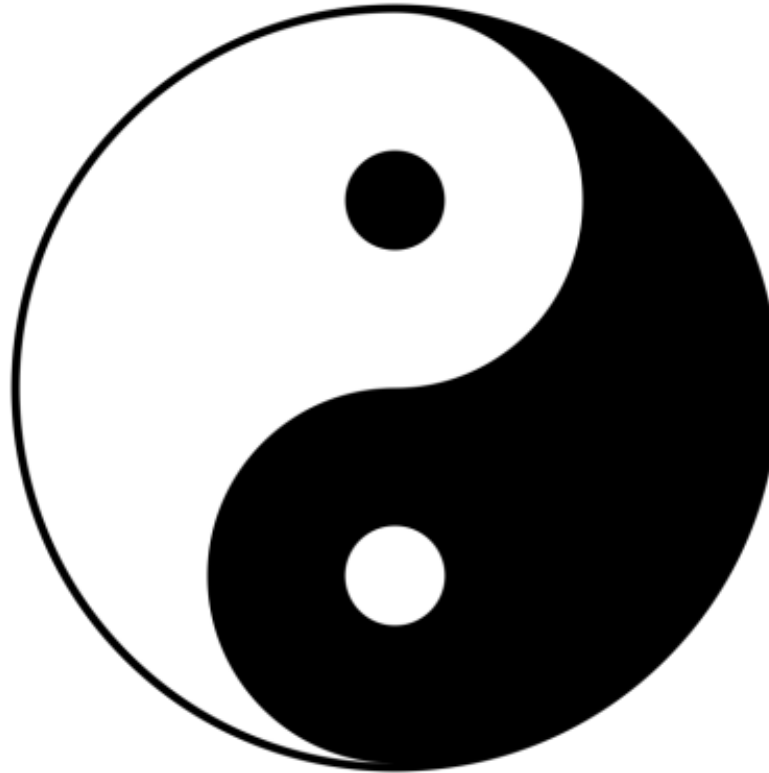
*insights*

*decisions*

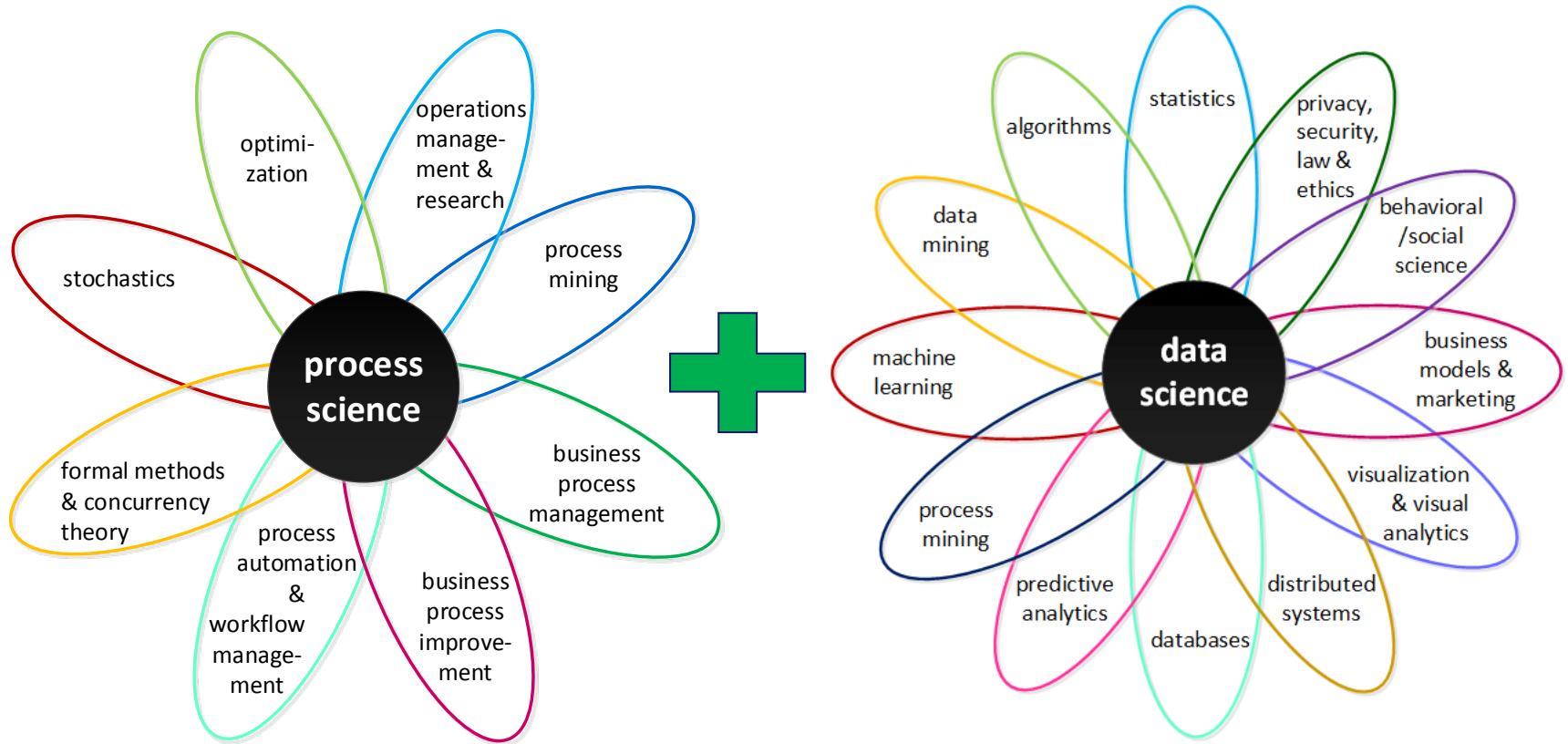
*models*

# Learning from the software itself

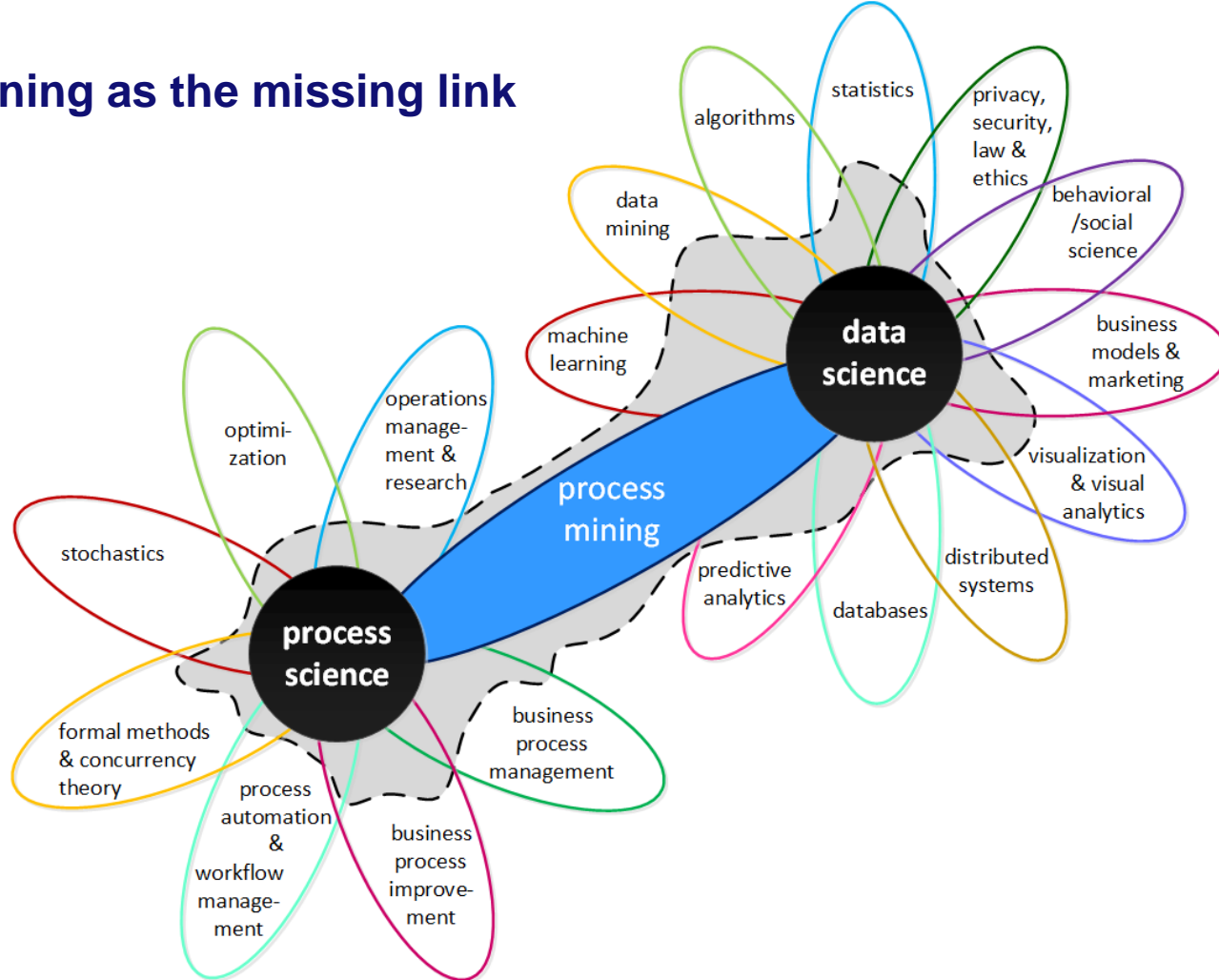
**code /  
structure**



**behavior**



# process mining as the missing link



*Spreadsheets for events  
(also coming from software)*

# Spreadsheet: Killer App for early computers

C11 (L) TOTAL C1 25

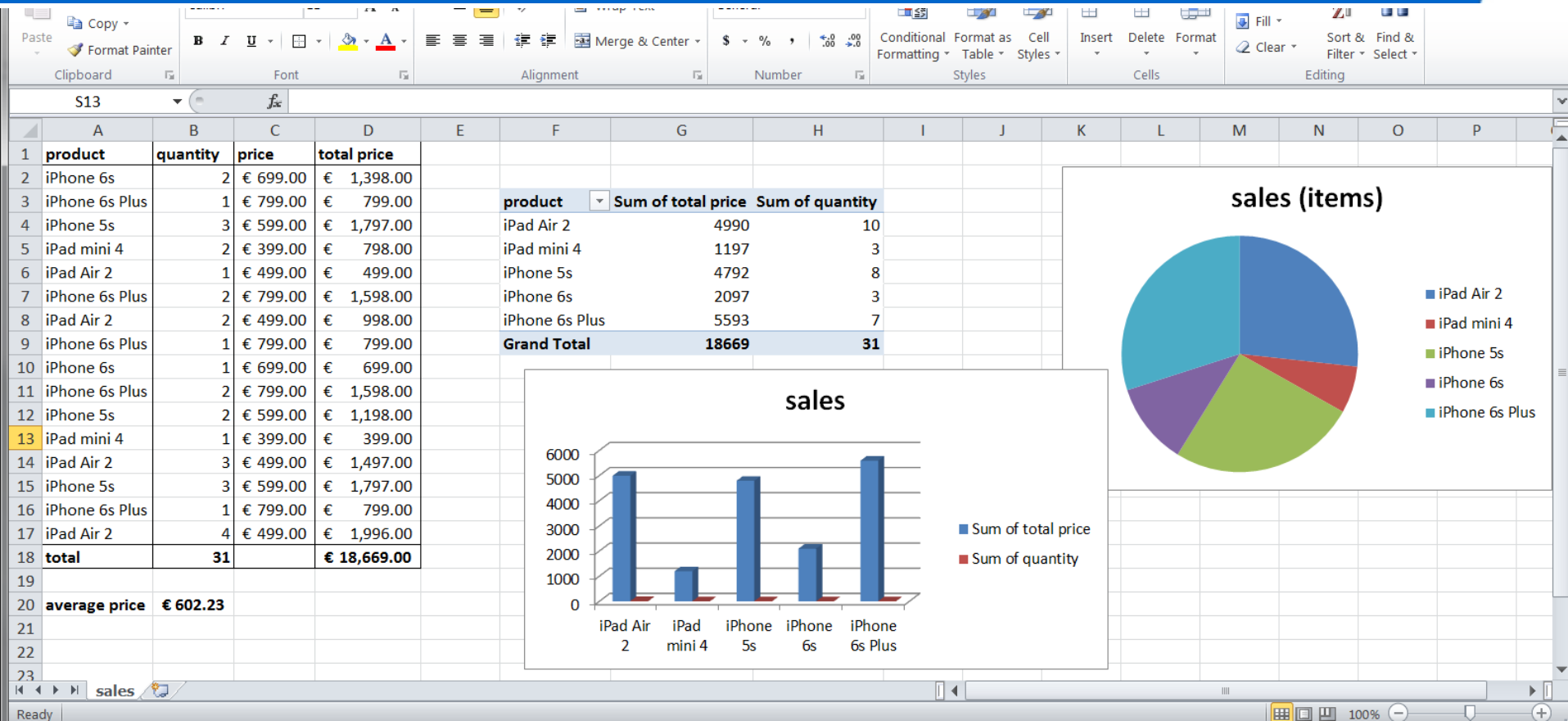
	A	B	C	D
1	ITEM	NO.	UNIT	COST
2	MUCK RAKE	43	12.95	556.85
3	BUZZ CUT	15	6.75	101.25
4	TOE TONER	250	49.95	12487.50
5	EYE SNUFF	2	4.95	9.90
			SUBTOTAL	13155.50
			9.75% TAX	1282.66
			TOTAL	14438.16

- **VisiCalc** (killer app for Apple II, Oct. 1979)
- **Lotus 1-2-3** (killer app for IBM PC 1983)
- **Microsoft Excel** (1985)





# Spreadsheet: Static data

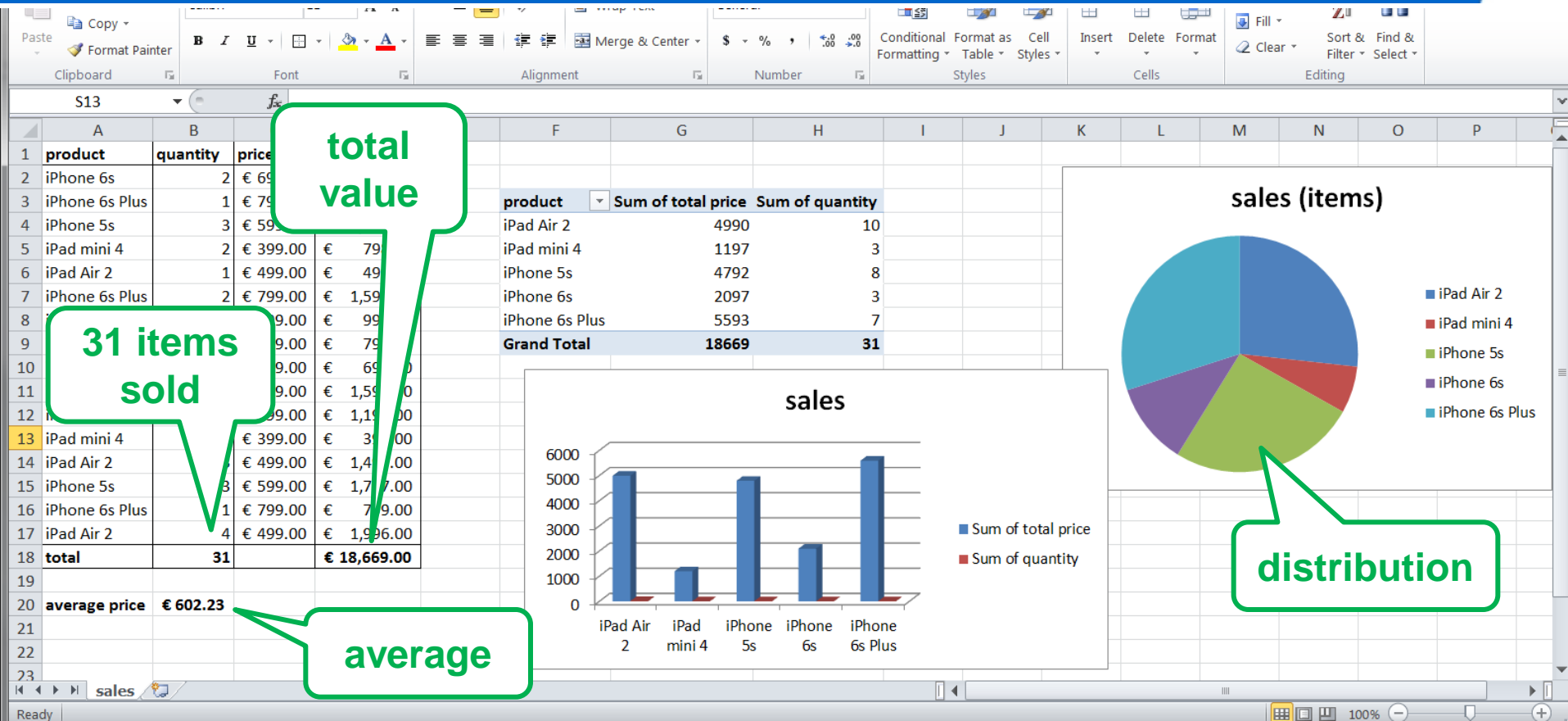


# Spreadsheet: Static data

The screenshot shows a spreadsheet with a table of product sales data. A red box highlights the table, which has columns for product, quantity, price, and total price. A blue callout labeled 'fact' points to the 'iPhone 5s' row. A green callout labeled 'derived' points to the 'total price' column.

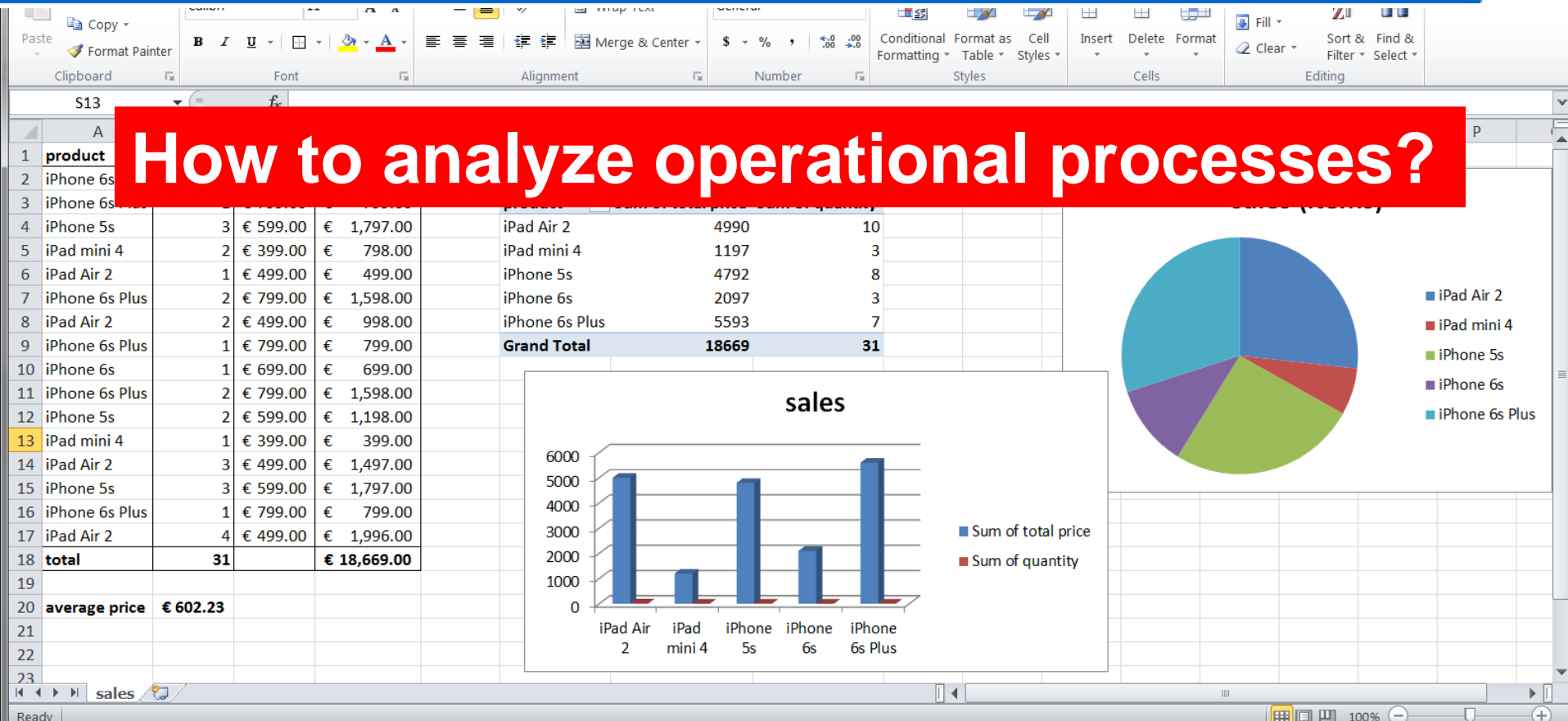
product	quantity	price	total price
iPhone 6s	2	€ 699.00	€ 1,398.00
iPhone 6s Plus	1	€ 799.00	€ 799.00
iPhone 5s	3	€ 599.00	€ 1,797.00
iPad mini 4	2	€ 399.00	€ 798.00
iPad Air 2	1	€ 499.00	€ 499.00
iPhone 6s Plus	2	€ 799.00	€ 1,598.00
iPad Air 2	2	€ 499.00	€ 998.00
iPhone 6s Plus	1	€ 799.00	€ 799.00
iPhone 6s	1	€ 699.00	€ 699.00
iPhone 6s Plus	2	€ 799.00	€ 1,598.00
iPhone 5s	2	€ 599.00	€ 1,198.00
iPad mini 4	1	€ 399.00	€ 399.00
iPad Air 2	3	€ 499.00	€ 1,497.00
iPhone 5s	3	€ 599.00	€ 1,797.00
iPhone 6s Plus	1	€ 799.00	€ 799.00
iPad Air 2	4	€ 499.00	€ 1,996.00
total	31		€ 18,669.00
average price	€ 602.23		

# Spreadsheet: Static data



# Spreadsheet: Static data

## How to analyze operational processes?



# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/11 13:26:29.000
4	185783	050 Plannen van toekomstige activiteiten	somebody	2007/09/11 13:26:29.000
5	185783	060 Aanvragen van formulier	somebody	2007/09/11 13:26:29.000
6	185783	070 Is 1e in gebreke staande?	somebody	2007/09/11 13:26:29.000
7	185783	100 Geraden maken	somebody	2007/09/11 13:26:29.000
8	185783	120 Plannen van toekomstige activiteiten	somebody	2007/09/24 10:55:56.000
9	185783	400 Inspectie uitgevoerd?	somebody	2007/09/24 10:56:06.000
10	185783	440 Zijn er nieuwe of niet herstelde gebreken?	somebody	2007/09/24 10:56:10.000
11	185783	450 Krijgt de huurder tijd om te herstellen?	somebody	2007/09/24 10:56:10.000
12	185783	500 Beoordelen/wijzigen leegstandsoort	somebody	2007/09/24 10:57:02.000
13	185783	110 Bepalen leegstandsoort	somebody	2007/09/24 10:57:42.000
14	185783	510 Is opleveringsformulier ondertekend?	somebody	2007/09/24 10:57:48.000
15	185783	130 Is het opleveringsformulier ondertekend?	somebody	2007/09/24 10:57:48.000
16	185783	140 Aanmaken 1e in gebreke staande	somebody	2007/09/24 11:01:00.000
17	185783	150 Is er sprake van ZAV?	somebody	2007/09/24 11:37:00.000
18	185783	160 Aanpassen woningwaardering	somebody	2007/09/24 11:37:00.000
19	185783	170 Harmoniseren huurprijs	somebody	2007/09/24 11:40:00.000
20	185783	180 Bepalen kandidaat huurder	somebody	2007/09/24 11:47:40.000
21	185783	190 Aanpassen plattegrond	somebody	2007/09/24 12:10:58.000
22	185783	200 Stellen van de huurprijs	somebody	2007/10/30 11:00:00.000
23	185783	210 Stellen van de huurprijs	somebody	2007/10/30 11:00:00.000
24	185783	220 Stellen van de huurprijs	somebody	2007/10/30 11:00:00.000
25	185783	230 Stellen van de huurprijs	somebody	2007/10/30 11:00:00.000
26	185783	240 Stellen van de huurprijs	somebody	2007/11/28 12:00:00.000
27	185783	250 Is contract getekend en geld ontvangen?	somebody	2007/12/10 10:44:00.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Opstellen eindnota	somebody	2007/12/12 11:19:41.000

row = event

resource

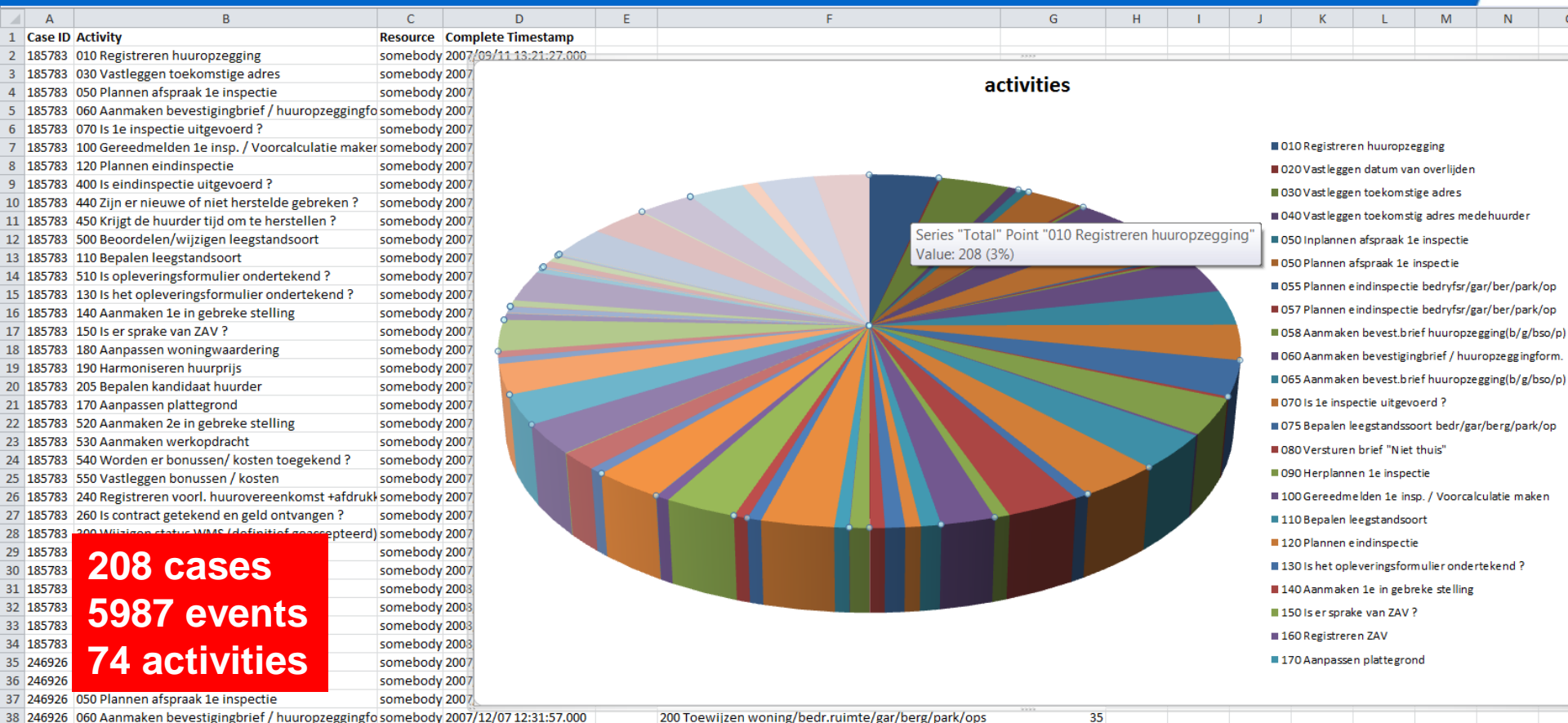
case  
identifier

activity  
name

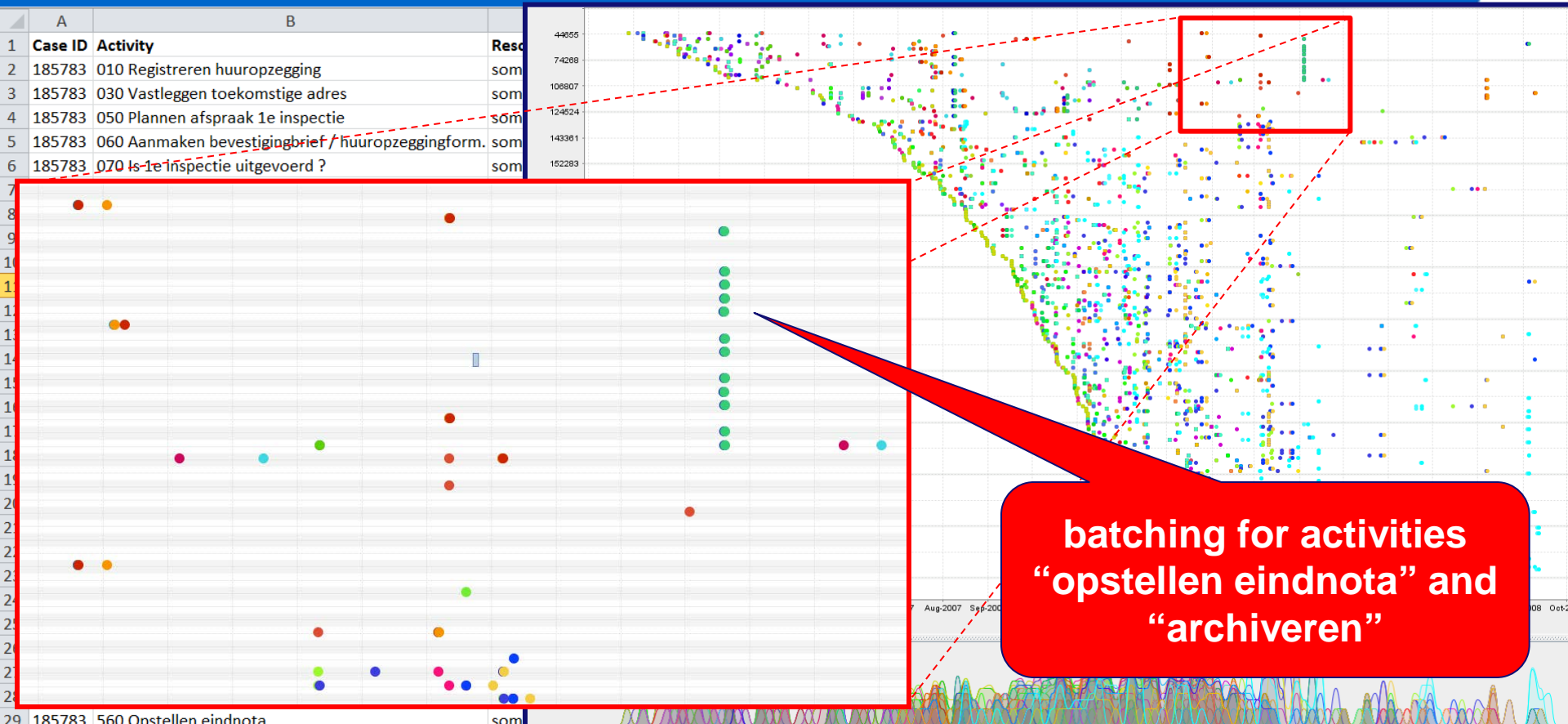
timestamp

- Input: **events** (“things that have happened”)
- Mandatory per event:
  - **case identifier**
  - **activity name**
  - **timestamp/date**
- Optional
  - **resource**
  - **transaction type**
  - **costs**
  - ...

# Process Mining: Spreadsheet for behavior



# Process Mining: Spreadsheet for behavior





A young girl with brown hair, wearing a bright pink puffer jacket and blue pants, is riding a red and white bicycle on a dirt path. The path is bordered by grass on the right and a paved road on the left. In the background, there is a road intersection with traffic lights and some buildings. Two speech bubbles are overlaid on the image: one pointing to the girl and another pointing to a line on the grass.

**Loesje van  
der Aalst**

**desire line**

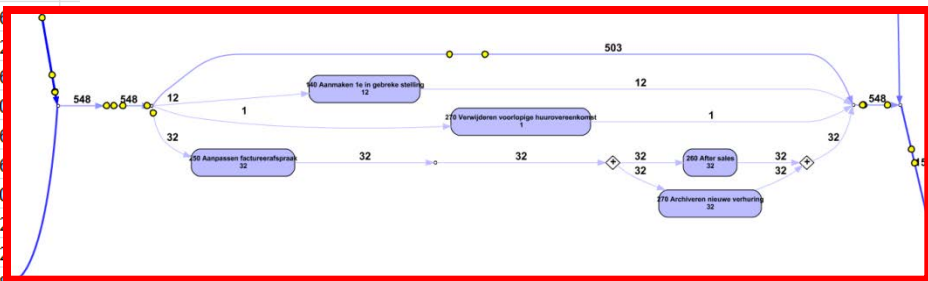
**Process Discovery**



# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/11 13:26:29.000
4	185783	050 Plannen afspraak 1e inspectie	somebody	2007/09/11 13:29:34.000
5	185783	060 Aanmaken bevestigingsbrief / huuropzeggingform.	somebody	2007/09/11 13:41:36.000
6	185783	070 Is 1e inspectie uitgevoerd ?	somebody	2007/09/24 08:39:32.000
7	185783	100 Is 1e inspectie / Voorcalculatie maken	somebody	2007/09/24 08:41:26.000
8	185783	120 Is 1e inspectie gestelde gebreken ?	somebody	2007/09/24 08:51:00.000
9	185783	400 Is 1e inspectie herstelde gebreken ?	somebody	2007/09/24 10:55:56.000
10	185783	400 Is 1e inspectie herstelde gebreken ?	somebody	2007/09/24 10:56:00.000
11	185783	400 Is 1e inspectie herstelde gebreken ?	somebody	2007/09/24 10:56:10.000
12	185783	500 Is 1e inspectie gestandsoort	somebody	2007/09/24 10:57:02.000
13	185783	110 Is 1e inspectie gestandsoort	somebody	2007/09/24 10:57:42.000
14	185783	510 Is opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:06.000
15	185783	130 Is het opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:19.000
16	185783	140 Aanmaken 1e in gebreke stelling	somebody	2007/09/24 11:01:58.000
17	185783	150 Is er sprake van ZAV ?	somebody	2007/09/24 11:37:33.000
18	185783	180 Aanpassen woningwaardering	somebody	2007/09/24 11:37:44.000
19	185783	190 Harmoniseren huurprijs	somebody	2007/09/24 11:40:01.000
20	185783	205 Bepalen kandidaat huurder	somebody	2007/09/24 11:47:42.000

*process discovery*



# Process Mining: Spreadsheet for behavior

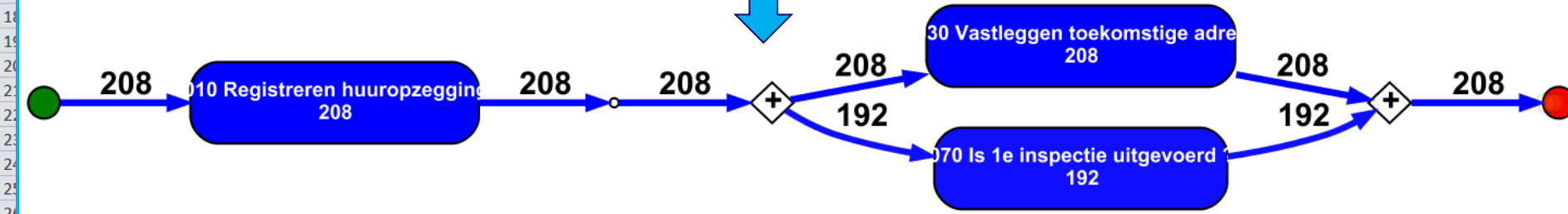
	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzeggng	somebody	2007/09/11 13:21:27.000

*process discovery*



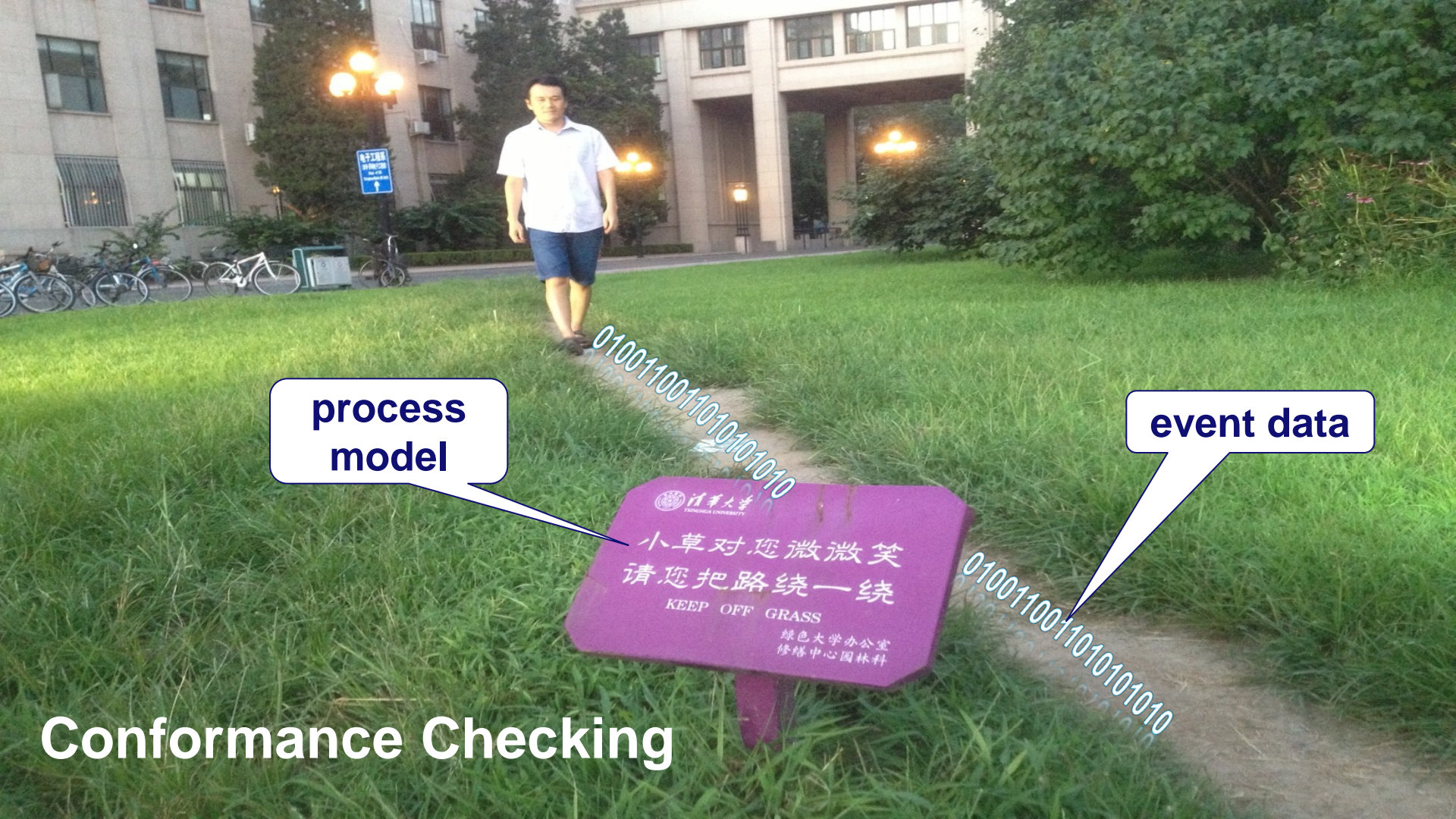
NO  
modeling  
needed!

9	185783	400 Is eindinspectie uitgevoerd ?	somebody	2007/09/24 10:55:56.
16	185783	140 Aanmaken 1e in gebreke stelling	somebody	2007/09/24 11:01:58.
17	185783	150 Is er sprake van ZAV ?	somebody	2007/09/24 11:37:33.



27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000





process  
model

event data

Conformance Checking



A photograph of a golf course. A light-colored gravel or sand path leads from the foreground towards a dark wooden fence in the middle ground. The path is bordered by green grass on both sides. In the background, there are more trees and a green lawn. Two callout boxes are present: one on the left pointing to the path and one on the right pointing to the fence.

**desire line**

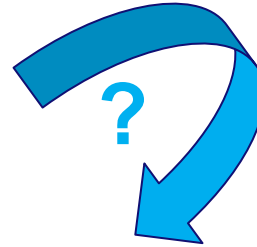
**very safe  
system**

**Conformance Checking**

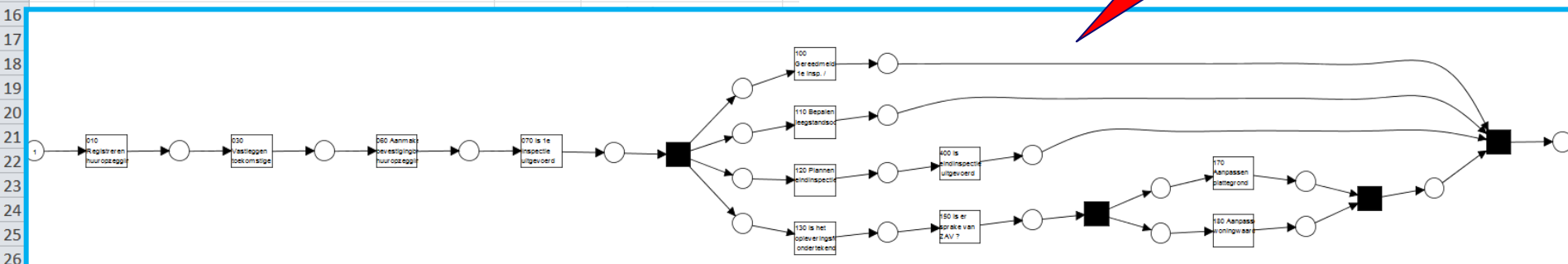
# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/11 13:26:29.000
4	185783	050 Plannen afspraak 1e inspectie	somebody	2007/09/11 13:29:34.000
5	185783	060 Aanmaken bevestigingsbrief / huuropzeggingform.	somebody	2007/09/11 13:41:36.000
6	185783	070 Is 1e inspectie uitgevoerd ?	somebody	2007/09/24 08:39:32.000
7	185783	100 Gereedmelden 1e insp. / Voorcalculatie maken	somebody	2007/09/24 08:41:26.000
8	185783	120 Plannen eindinspectie	somebody	2007/09/24 08:51:00.000
9	185783	400 Is eindinspectie uitgevoerd ?	somebody	2007/09/24 10:55:56.000
10	185783	440 Zijn er nieuwe of niet herstelde gebreken ?	somebody	2007/09/24 10:56:06.000
11	185783	450 Krijgt de huurder tijd om te herstellen ?	somebody	2007/09/24 10:56:10.000
12	185783	500 Beoordelen/wijzigen leegstandsoort	somebody	2007/09/24 10:57:02.000
13	185783	110 Bepalen leegstandsoort	somebody	2007/09/24 10:57:42.000
14	185783	510 Is opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:08.000
15	185783	130 Is het opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:19.000

*conformance checking*



**discovered or  
hand-made**

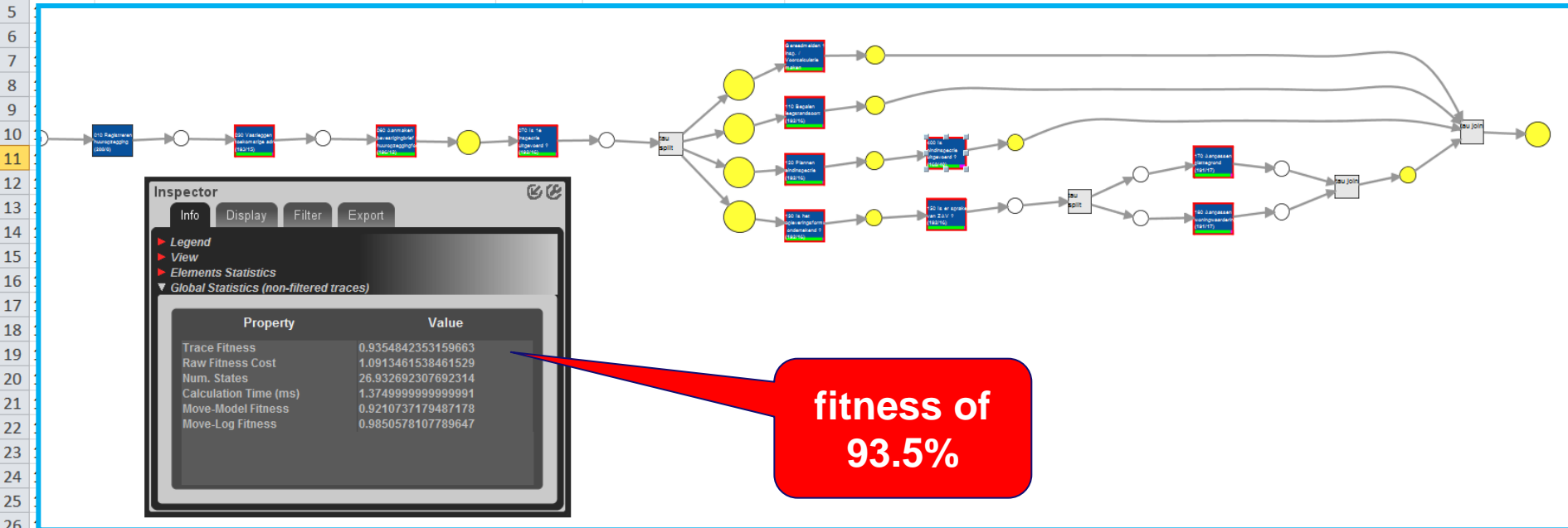


27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/11 13:26:29.000
4	185783	050 Plannen afspraak 1e inspectie	somebody	2007/09/11 13:29:34.000

*conformance checking*

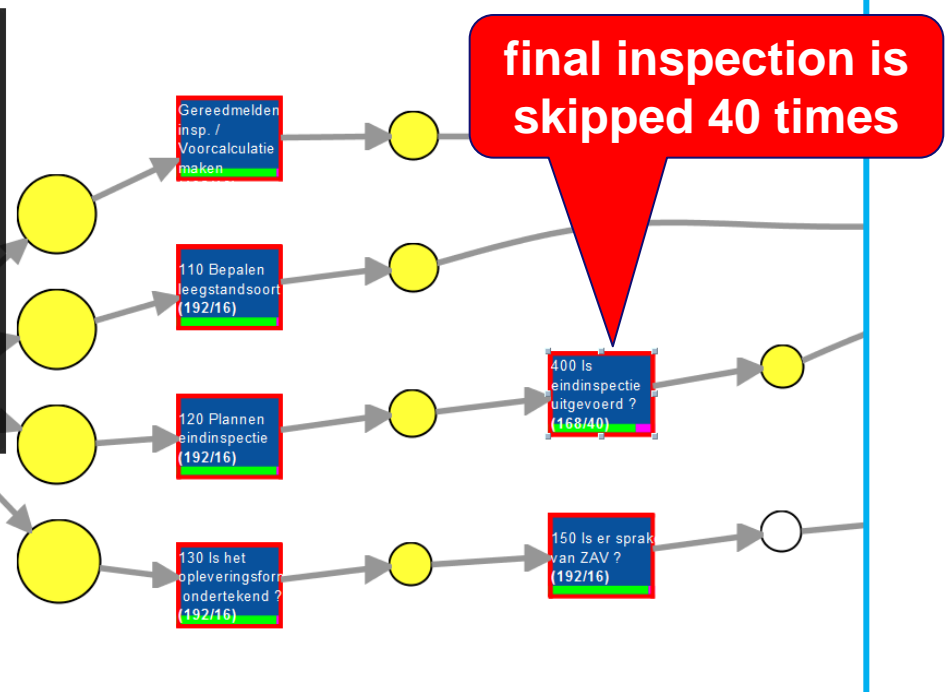
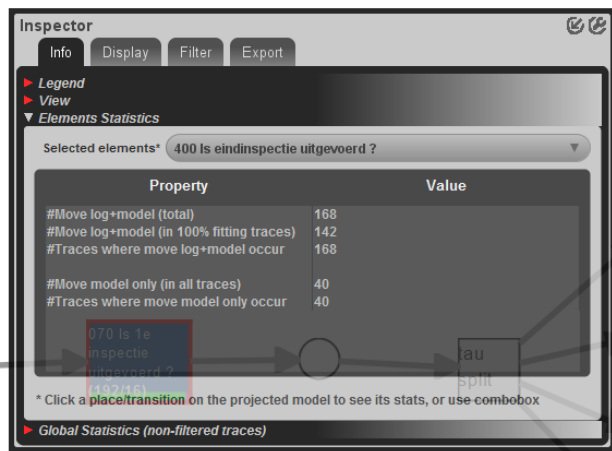


27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toek		
4	185783	050 Plannen afspraak		
5	185783	060 Aanmaken beves		
6	185783	070 Is 1e inspectie uit		
7	185783	100 Gereedmelden 1e		
8	185783	120 Plannen eindinspe		
9	185783	400 Is eindinspectie u		
10	185783	440 Zijn er nieuwe of		
11	185783	450 Krijgt de huurder		
12	185783	500 Beoordelen/wijzi		
13	185783	110 Bepalen leegstan		
14	185783	510 Is opleveringsform		
15	185783	130 Is het opleverings		
16	185783	140 Aanmaken 1e in g		
17	185783	150 Is er sprake van Z		
18	185783	180 Aanpassen wonin		
19	185783	190 Harmoniseren hu		
20	185783	205 Bepalen kandida		
21	185783	170 Aanpassen platte		
22	185783	520 Aanmaken 2e in g		
23	185783	530 Aanmaken werko		
24	185783	540 Worden er bonus		
25	185783	550 Vastleggen bonus		
26	185783	240 Registreren voor		
27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

*conformance checking*





# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging		
3	185783	030 Vastleggen toekomstige adres		
4	185783	050 Plannen afspraak 1e inspectie		
5	185783	060 Aanmaken bevestigingsbrief / huuropzeg		
6	185783	070 Is 1e inspectie uitgevoerd ?		
7	185783	100 Gereedmelden 1e insp. / Voorcalculatie		
8	185783	120 Plannen eindinspectie		
9	185783	400 Is eindinspectie uitgevoerd ?		
10	185783	440 Zijn er nieuwe of niet herstelde gebreken ?		
11	185783	450 Krijgt de huurder tijd om te herstellen ?		
12	185783	500 Beoordelen/wijzigen leegstandsoort		
13	185783	110 Bepalen leegstandsoort		
14	185783	510 Is opleveringsformulier ondertekend ?		
15	185783	130 Is het oplevering		
16	185783	140 Aanmaken 1e i		
17	185783	150 Is er sprake van		
18	185783	180 Aanpassen wor		
19	185783	190 Harmoniseren i		
20	185783	205 Bepalen kandid		
21	185783	170 Aanpassen plat		
22	185783	520 Aanmaken 2e i		
23	185783	530 Aanmaken werkopdracht		
24	185783	540 Worden er bonussen/ kosten toegekend ?		
25	185783	550 Vastleggen bonussen / kosten		
26	185783	240 Registreren voorl. huurovereenkomst +afdrukken		
27	185783	260 Is contract getekend en geld ontvangen ?		
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)		
29	185783	560 Opstellen eindnota		

move on model  
(something should have  
happened, but did not)

move on log  
(something happened that  
should not happen)

*conformance checking*

Case id(s): 80437

Num. Cases

2

Is Alignment Reliable?

Yes

Trace Fitness

0.96

Alignment

16 events

Alignment

16 events

Alignment

18 events

Case id(s): 192867

Num. Cases

1

Is Alignment Reliable?

Yes

Trace Fitness

0.92

Alignment

18 events

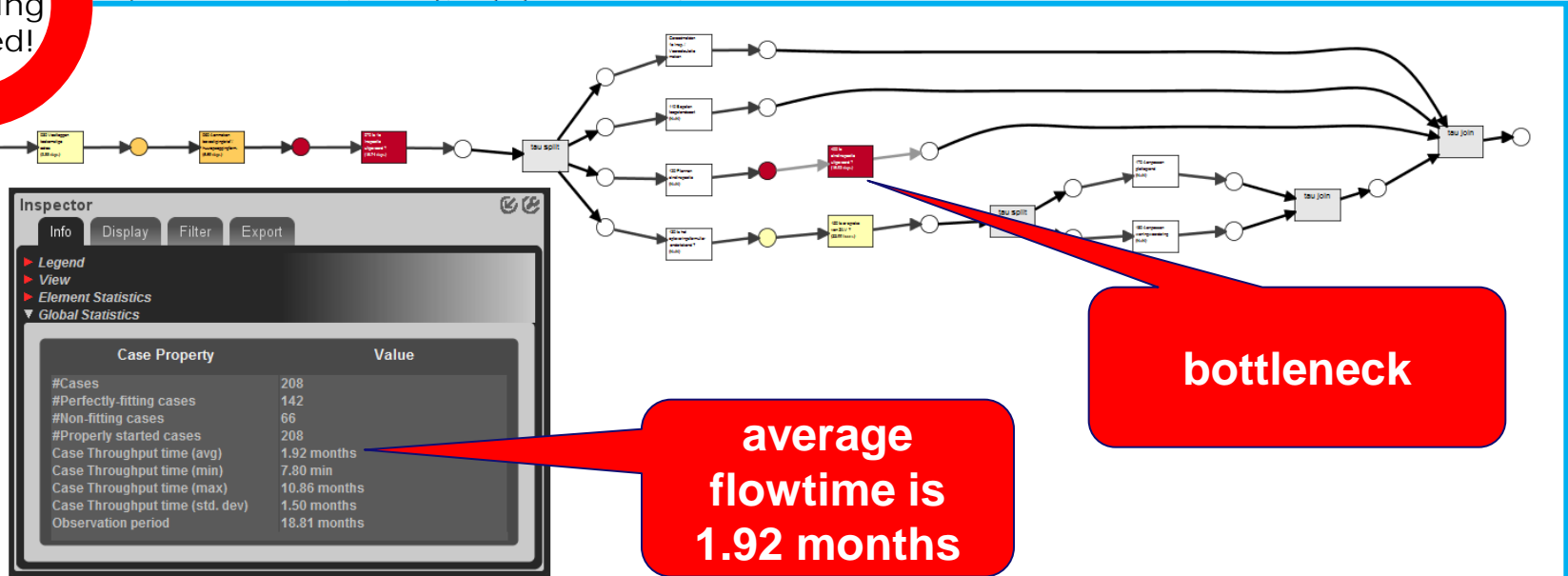


# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID		Resource	Complete Timestamp
2	185783	Opzegging	somebody	2007/09/11 13:21:27.000
3	185783	Onstige adres	somebody	2007/09/11 13:26:29.000
4	185783	Inspectie	somebody	2007/09/11 13:29:34.000

NO  
modeling  
needed!

*performance analysis*



26	185783	240 Registreren voorl. huurovereenkomst + afdrukken	somebody	2007/11/28 12:34:23.000
27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijziging status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/12/12 11:19:41.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/12/12 11:19:41.000
4	185783	050 Plannen afbreken 1e inspectie	somebody	2007/12/12 11:19:41.000
5	185783	060 A		
6	185783	070 Is		
7	185783	100 G		
8	185783	120 P		
9	185783	400 Is		
10	185783	440 Z		
11	185783	450 K		
12	185783	500 B		
13	185783	110 B		
14	185783	510 Is		
15	185783	130 Is		
16	185783	140 A		
17	185783	150 Is		
18	185783	180 A		
19	185783	190 H		
20	185783	205 B		
21	185783	170 A		
22	185783	520 A		
23	185783	530 A		
24	185783	540 V		
25	185783	550 V		
26	185783	240 R		
27	185783	260 Is		
28	185783	300 V		
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

*performance analysis*

waiting time of  
15.74 days

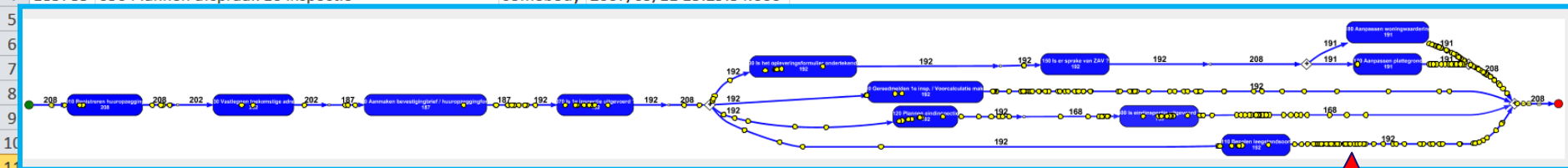
NO  
modeling  
needed!

Inspector						
Info Display Filter Export						
Legend						
View						
Element Statistics						
Selected elements* sink 5						
Property	Min.	Max.	Avg.	Std. Dev.	Freq.	
Waiting time	6.00 seconds	9.10 months	15.74 days	24.52 days	192	
Synchroniza...	0.00 ms	0.00 ms	0.00 ms	0.00 ms	192	
Sojourn time	6.00 seconds	9.10 months	15.74 days	24.52 days	192	

# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Complete Timestamp
2	185783	010 Registreren huuropzegging	somebody	2007/09/11 13:21:27.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/11 13:26:29.000
4	185783	050 Plannen afspraak 1e inspectie	somebody	2007/09/11 13:29:34.000

*animating reality*



11	185783	500 Beoordelen/wijzigen leegstandsoort	somebody	2007/09/24 10:57:02.000
12	185783	110 Bepalen leegstandsoort	somebody	2007/09/24 10:57:42.000
13	185783	510 Is opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:08.000
14	185783	130 Is het opleveringsformulier ondertekend ?	somebody	2007/09/24 10:58:19.000
15	185783	140 Aanmaken 1e in gebreke stelling	somebody	2007/09/24 11:01:58.000
16	185783	150 Is er sprake van ZAV ?	somebody	2007/09/24 11:37:33.000
17	185783	180 Aanpassen woningwaardering	somebody	2007/09/24 11:37:44.000
18	185783	190 Harmoniseren huurprijs	somebody	2007/09/24 11:40:01.000
19	185783	205 Bepalen kandidaat huurder	somebody	2007/09/24 11:47:42.000
20	185783	170 Aanpassen plattegrond	somebody	2007/09/24 12:10:58.000
21	185783	520 Aanmaken 2e in gebreke stelling	somebody	2007/10/30 11:45:53.000
22	185783	530 Aanmaken werkopdracht	somebody	2007/10/30 11:46:09.000
23	185783	540 Worden er bonussen/ kosten toegekend ?	somebody	2007/10/30 11:46:36.000
24	185783	550 Vastleggen bonussen / kosten	somebody	2007/10/30 11:53:00.000
25	185783	240 Registreren voorl. huurovereenkomst +afdrukken	somebody	2007/11/28 12:34:23.000
26	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
27	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
28	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

real cases

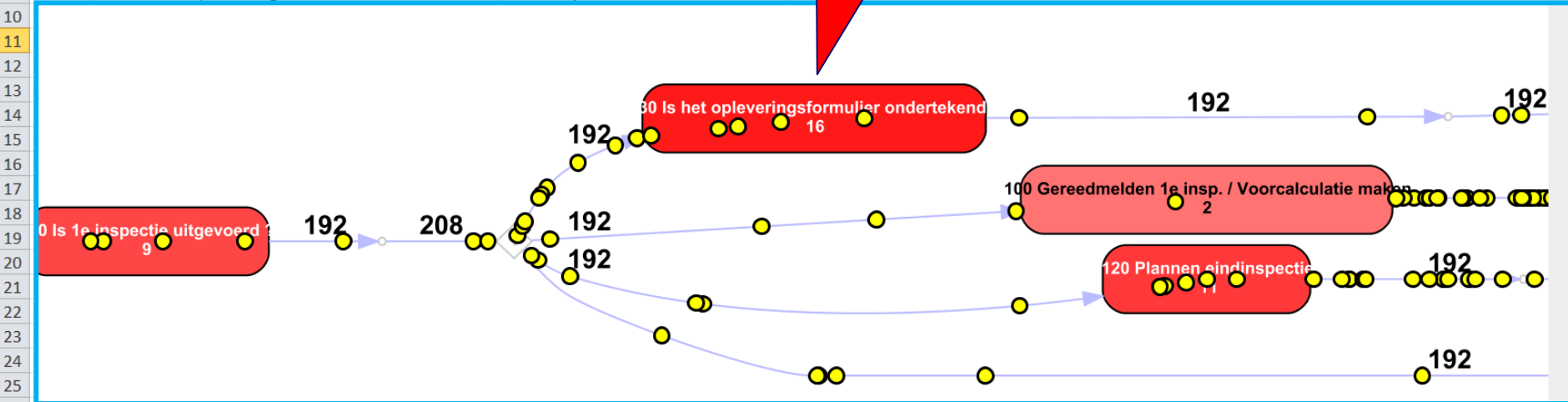
NO  
modeling  
needed!

# Process Mining: Spreadsheet for behavior

	A	B	C	D
1	Case ID	Activity	Resource	Completed
2	185783	010 Registreren huuropzegging	somebody	2007/09/24 08:41:26.000
3	185783	030 Vastleggen toekomstige adres	somebody	2007/09/24 08:51:00.000
4	185783	050 Plannen afspraak 1e inspectie	somebody	2007/09/24 10:55:56.000
5	185783	060 Aanmaken bevestigingsbrief / huuropzeggingform.	somebody	
6	185783	070 Is 1e inspectie uitgevoerd ?	somebody	
7	185783	100 Gereedmelden 1e insp. / Voorcalculatie maken	somebody	2007/09/24 08:41:26.000
8	185783	120 Plannen eindinspectie	somebody	2007/09/24 08:51:00.000
9	185783	400 Is eindinspectie uitgevoerd ?	somebody	2007/09/24 10:55:56.000

*animating reality*

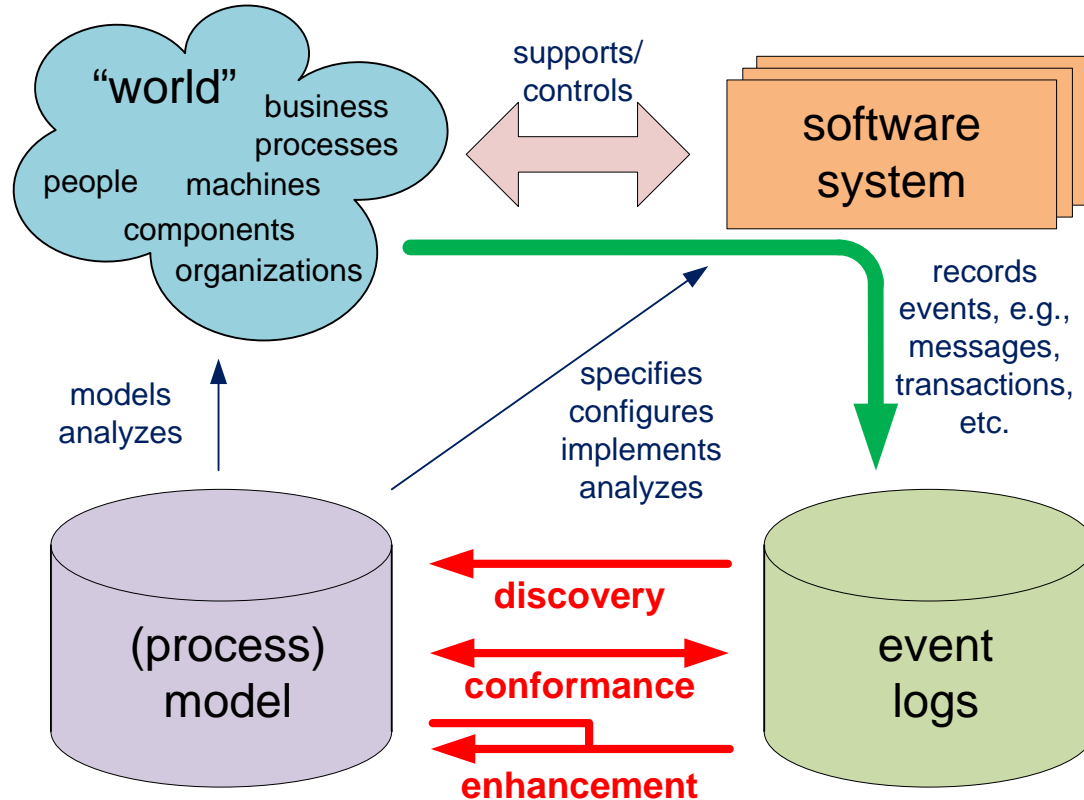
16 cases are queueing



26	185783	240 Registreren voorl. huurovereenkomst +afdrukken	somebody	2007/11/28 12:34:23.000
27	185783	260 Is contract getekend en geld ontvangen ?	somebody	2007/12/10 10:44:06.000
28	185783	300 Wijzigen status WMS (definitief geaccepteerd)	somebody	2007/12/11 16:26:14.000
29	185783	560 Onstellen eindnota	somebody	2007/12/12 11:19:41.000

# *Process models and event logs*

# Getting the "right" data ...



# Event log

- We assume the existence of an **event log** where each **event** refers to a **case**, an **activity**, and a point in **time**.
- An **event log** can be seen as a **collection of cases**.
- A **case** can be seen as a **trace/sequence of events**.

# Event data may come from ...

- a database system (e.g., patient data in a hospital),
- a comma-separated value file or spreadsheet,
- a transaction log (e.g., from a database system),
- a business suite/ERP system (SAP, Oracle, etc.),
- a message log (e.g., from IBM middleware),
- software development tools (e.g., repository mining),
- instrumented software (see Maikel's part),
- etc.

Focus BSR and  
second part of this  
presentation



# An example log

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8
...	...	...	...

**case id**

**activity name**

**timestamp**

**other data**

# Another event log: order handling

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
...	...	...	...	...	...

**case id**

**activity name**

**timestamp**

**resource**

**other data**

# Another event log: patient treatment

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	radiation therapy	23-1-2014@13.57	Dr. Jones	61	140.00
5541	radiation therapy	23-1-2014@13.08	Dr. Jones	61	140.00
...	...	...	...	...	...

**case id**

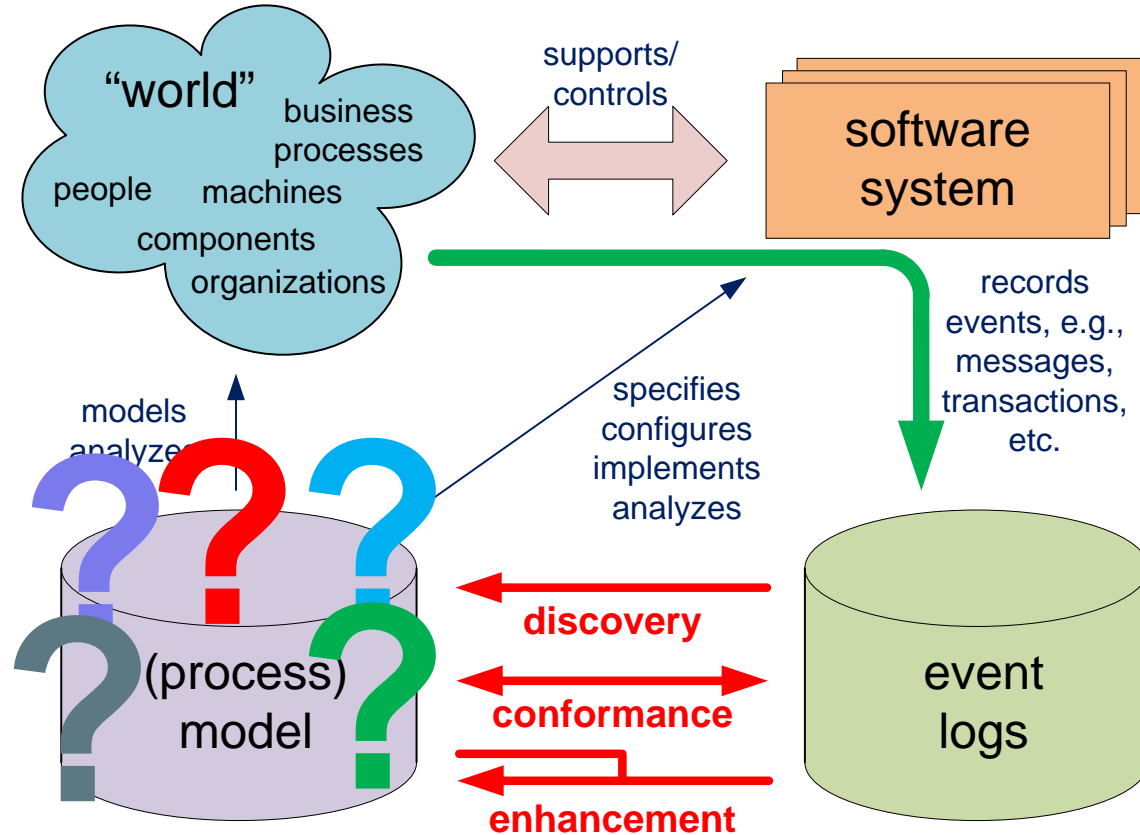
**activity name**

**timestamp**

**resource**

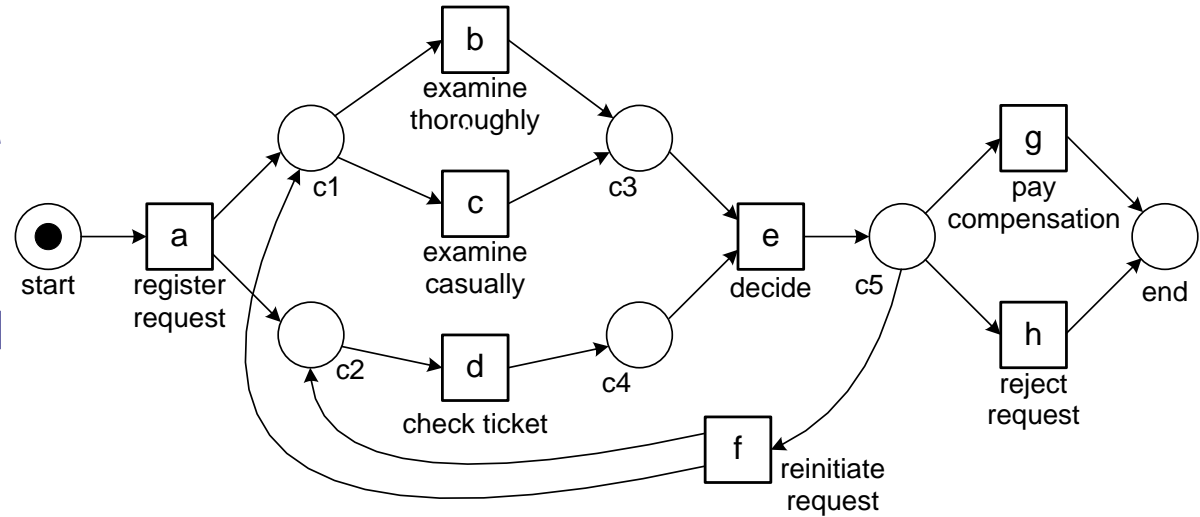
**other data**

# Selecting the "right" representation...

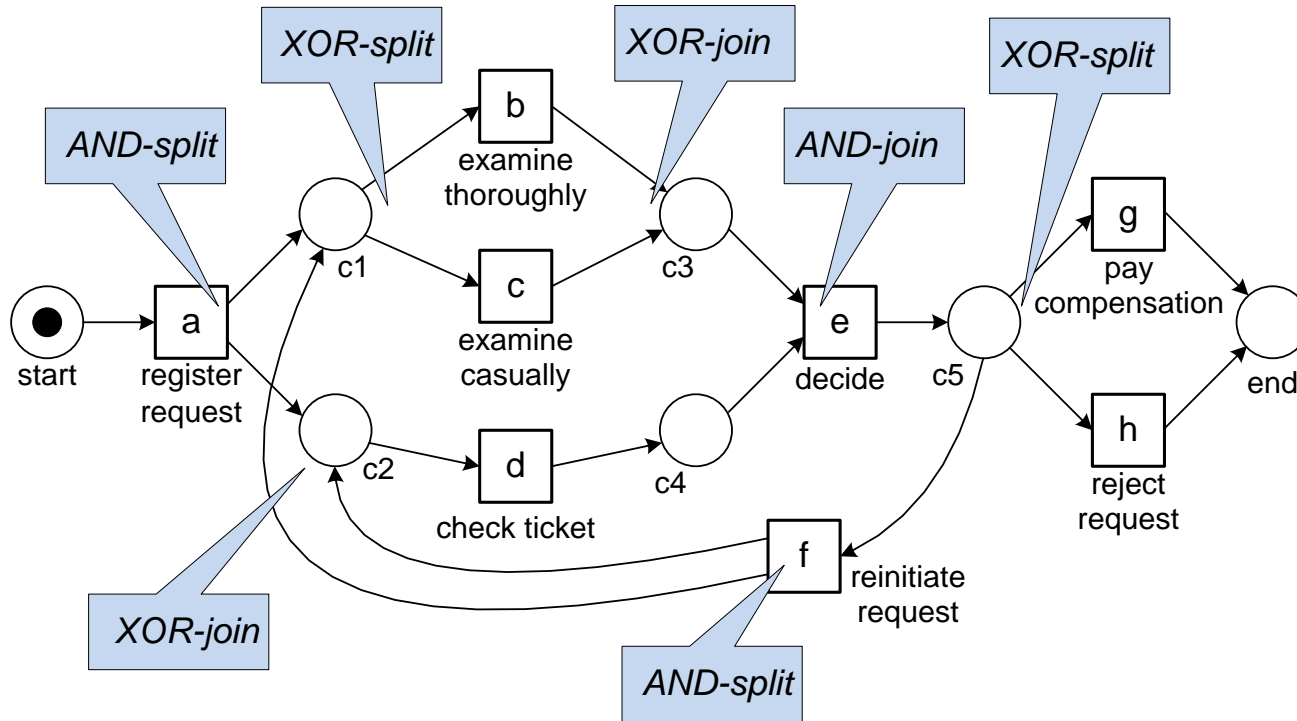


# Just control-flow (represented as a Petri net)

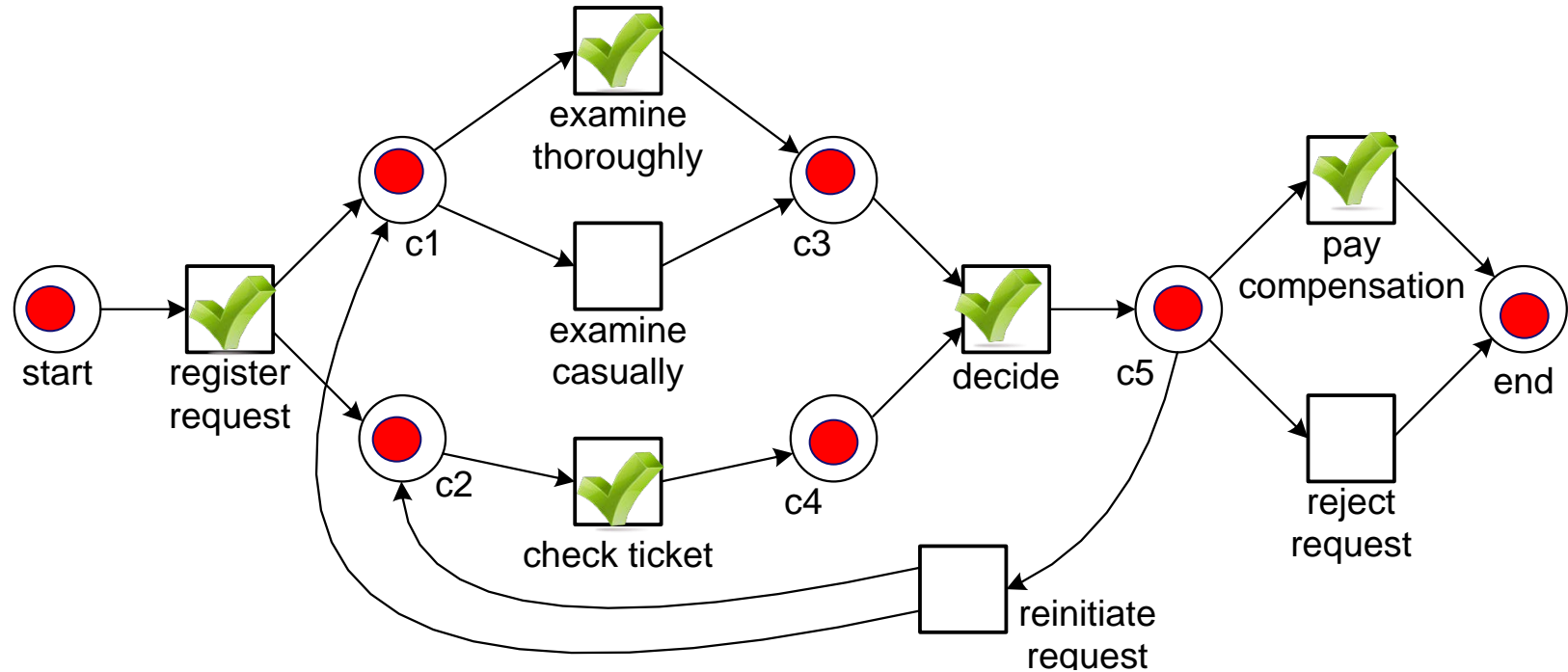
- Case starts with **a** and ends with **g** or **h**.
- Activity **d** is concurrent with **b** or **c**.
- Activity **e** has to wait until (**d** and **b**) or (**d** and **c**) have completed.
- There are three possible decisions: **f**, **g**, or **h**.
- ...



# Petri nets: semantics & terminology

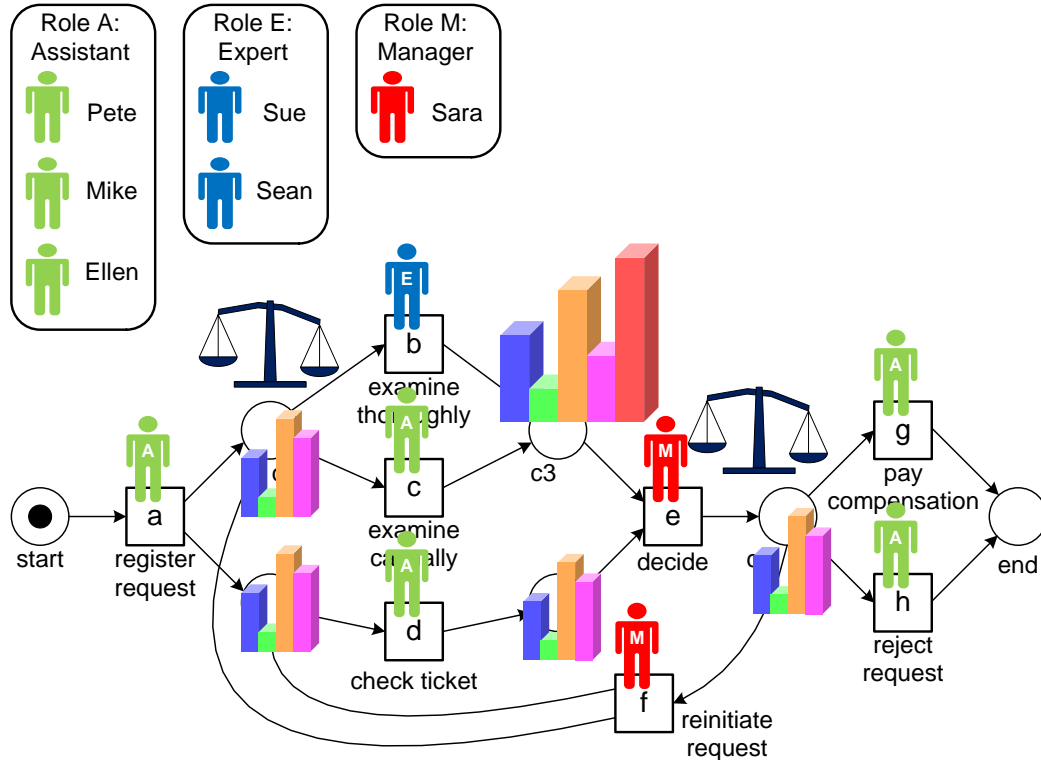


# Example run of the model



Only one of infinitely many possible firing sequences! **nd]**

# Additional perspectives

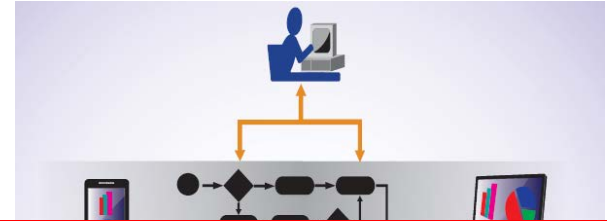


- **control-flow**
- **data-flow**
- **time**
- **resources**
- **costs**
- **risks**
- ...



# Possible control-flow notations

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

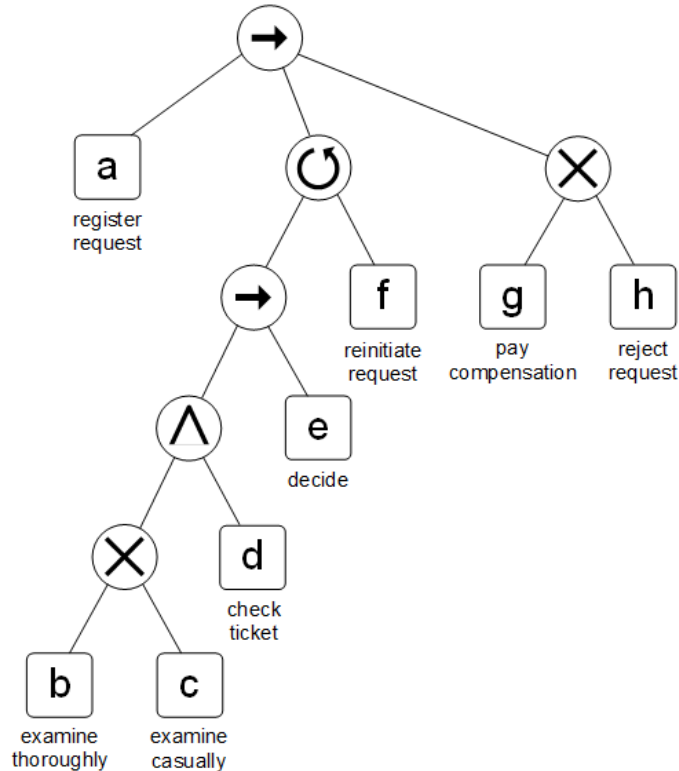
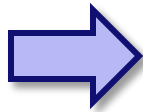
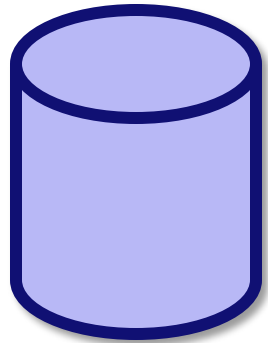


## Notation:

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

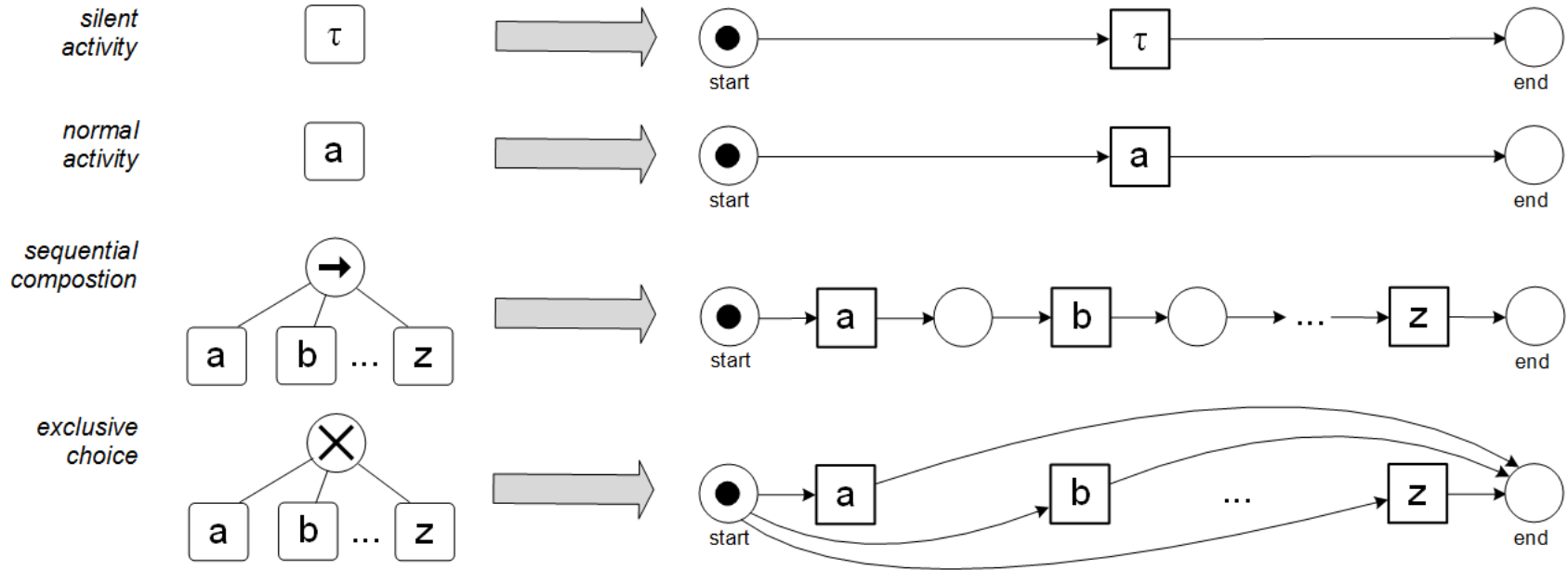
*An example of a  
discovery algorithm*

# Process trees (to ensure soundness)



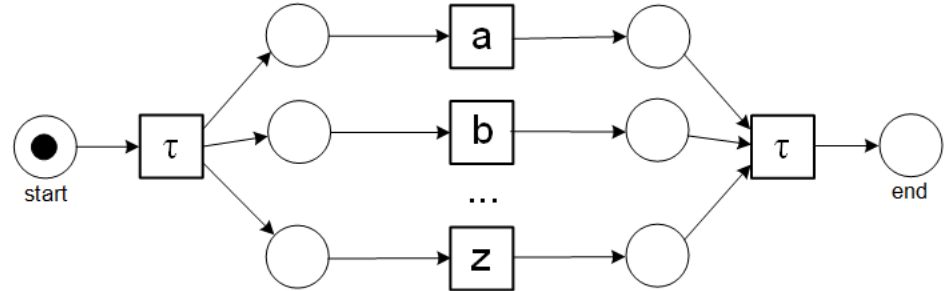
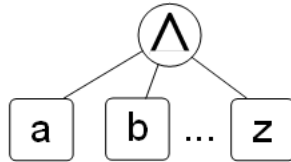
<i>sequential composition</i>	
<i>exclusive choice</i>	
<i>parallel composition</i>	
<i>redo loop</i>	
<i>normal activity</i>	
<i>silent activity</i>	

# Process trees (semantics)

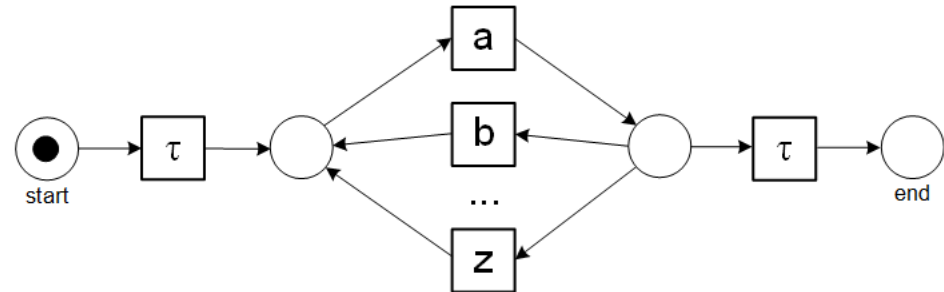
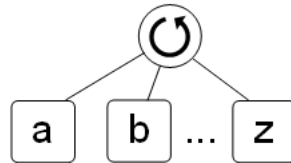


# Process trees (semantics)

*parallel composition*



*redo loop*



# Split event logs based on activity labels

**abdef**  
**acdef**  
**adbef**  
**adcef**  
**abdeg**  
**acdeg**  
**adbeg**  
**adceg**

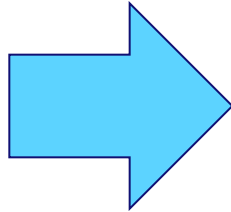


# Split $\{a,b,c,d,e,f,g,h\}$ into $\{a,b,c,d\}$ and $\{e,f,g\}$ using sequence decomposition

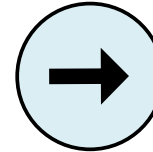
**abdef  
acdef  
adbef  
adcef  
abdeg  
acdeg  
adbeg  
adceg**

# Result

abdef  
acdef  
adbef  
adcef  
abdeg  
acdeg  
adbeg  
adceg

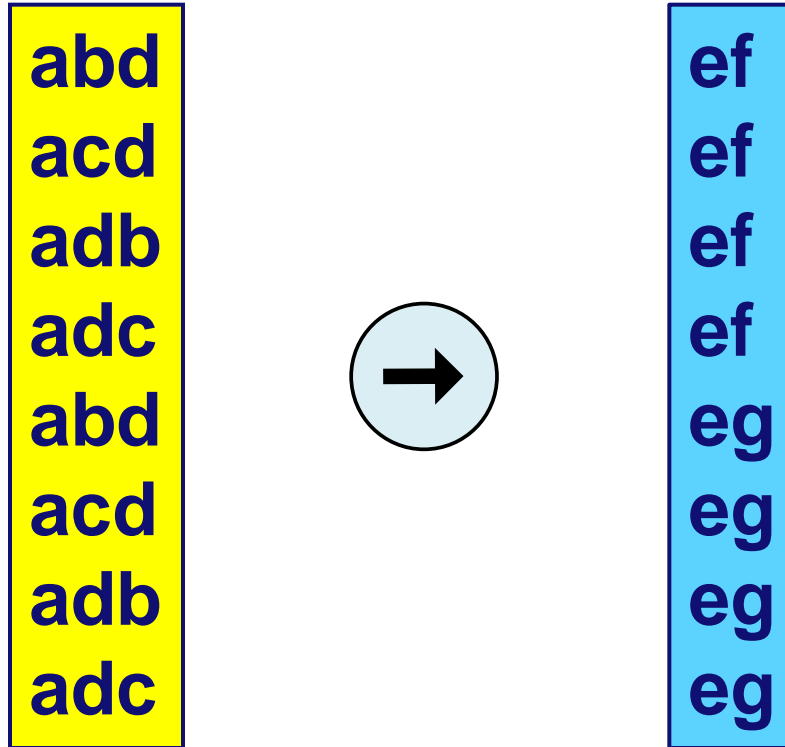


abd  
acd  
adb  
adc  
abd  
acd  
adb  
adc

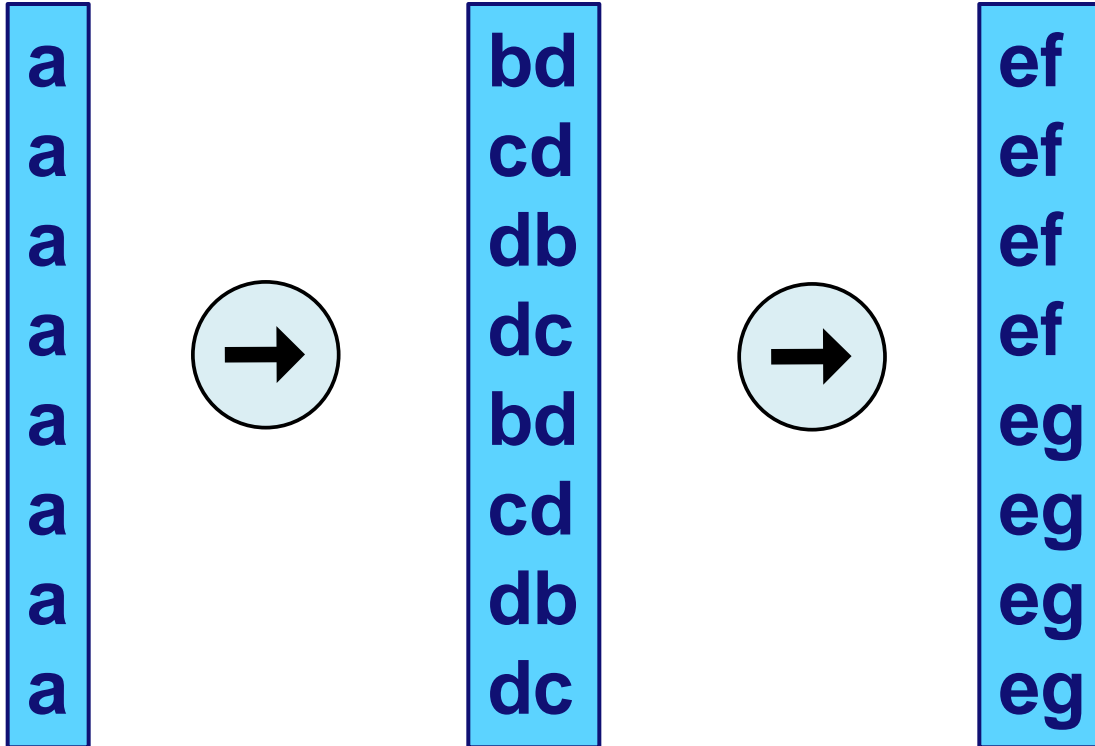


ef  
ef  
ef  
ef  
eg  
eg  
eg  
eg

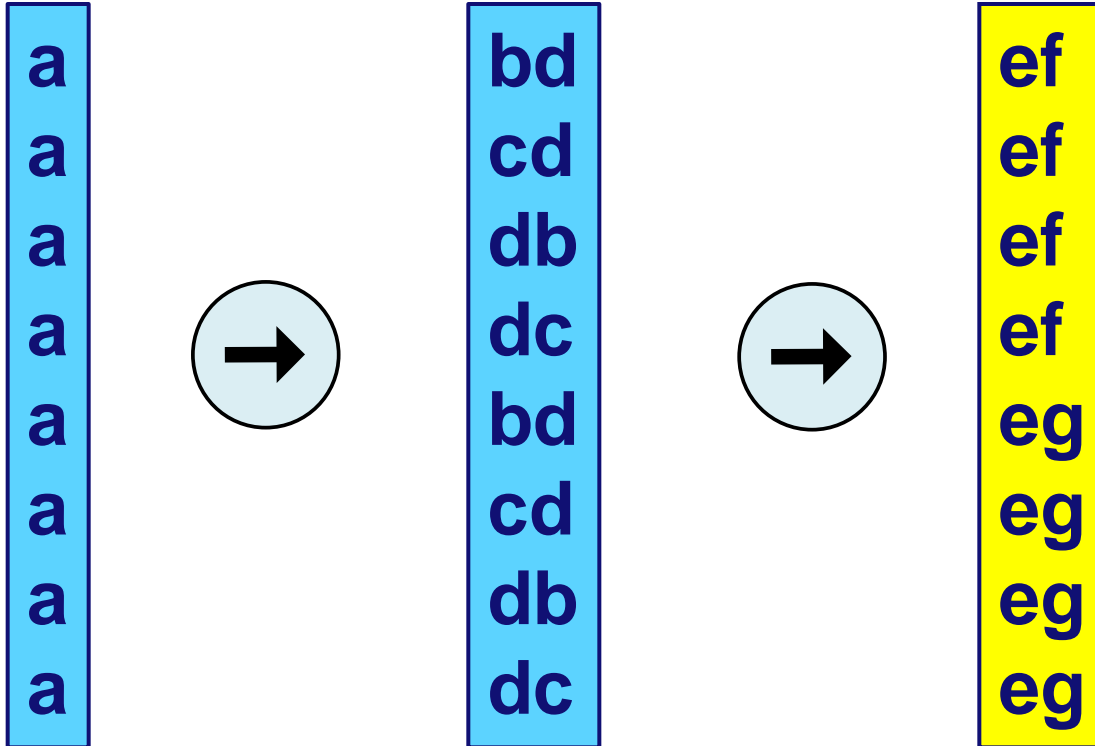
# Split $\{a,b,c,d\}$ into $\{a\}$ and $\{b,c,d\}$ using sequence decomposition



# Result

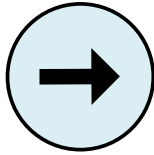


# Split $\{e,f,g\}$ into $\{e\}$ and $\{f,g\}$ using sequence decomposition

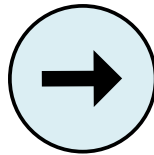


# Result

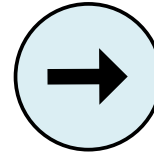
a  
a  
a  
a  
a  
a  
a  
a



bd  
cd  
db  
dc  
bd  
cd  
db  
dc



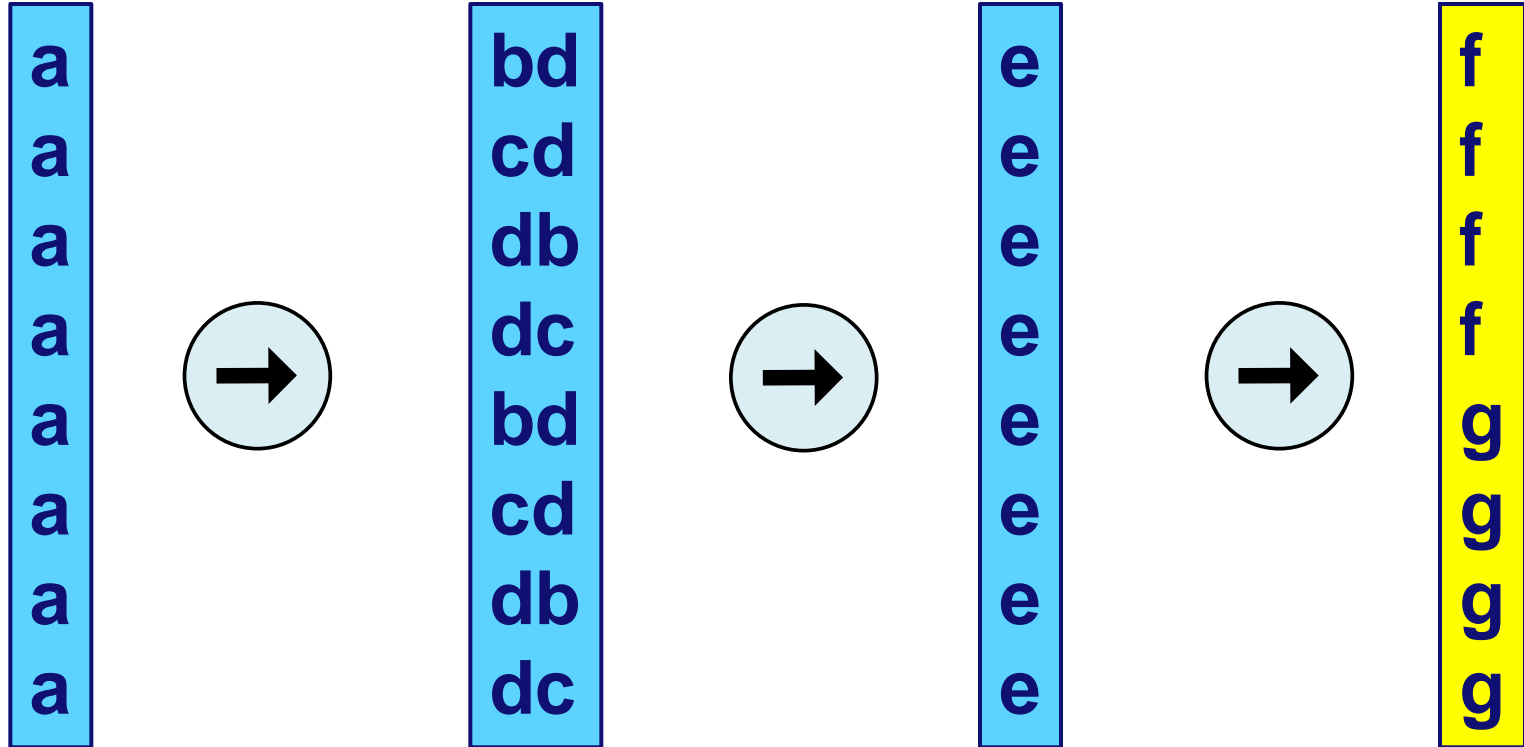
e  
e  
e  
e  
e  
e  
e  
e



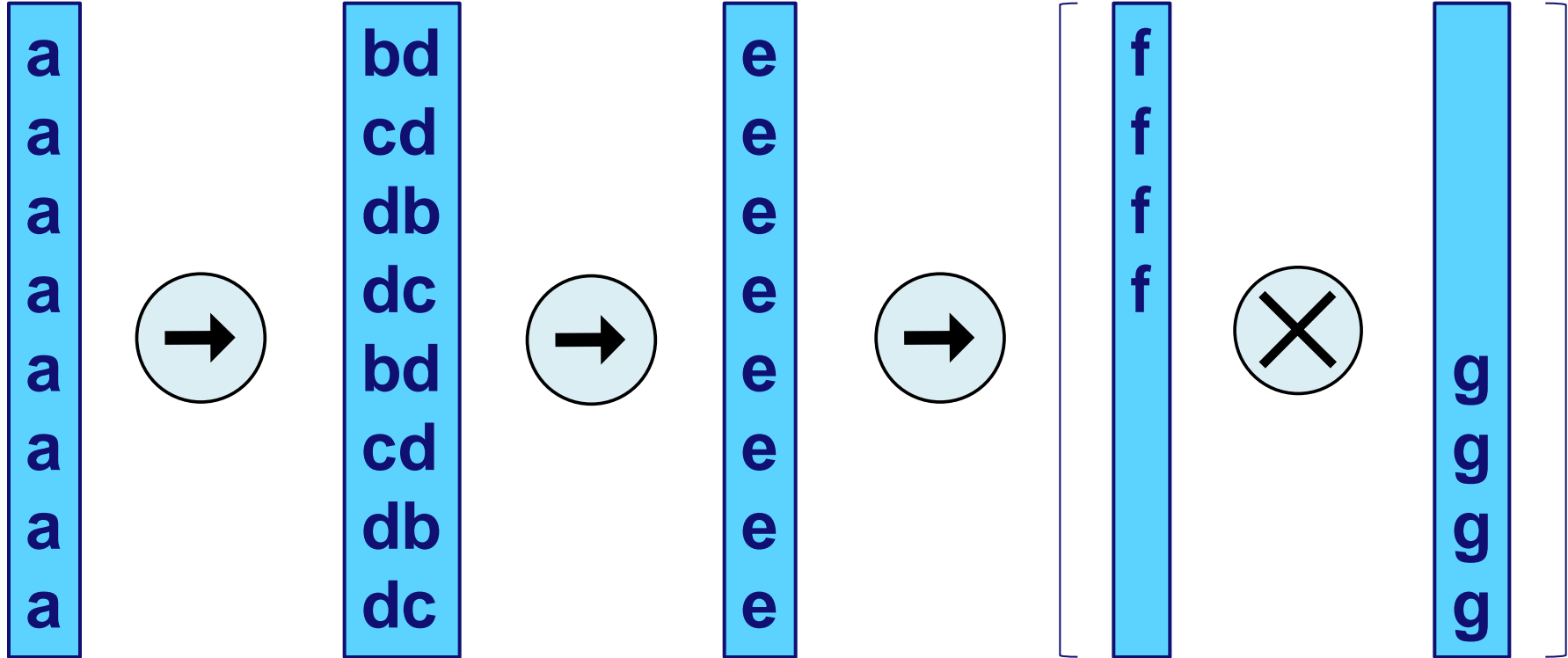
f  
f  
f  
f  
g  
g  
g  
g



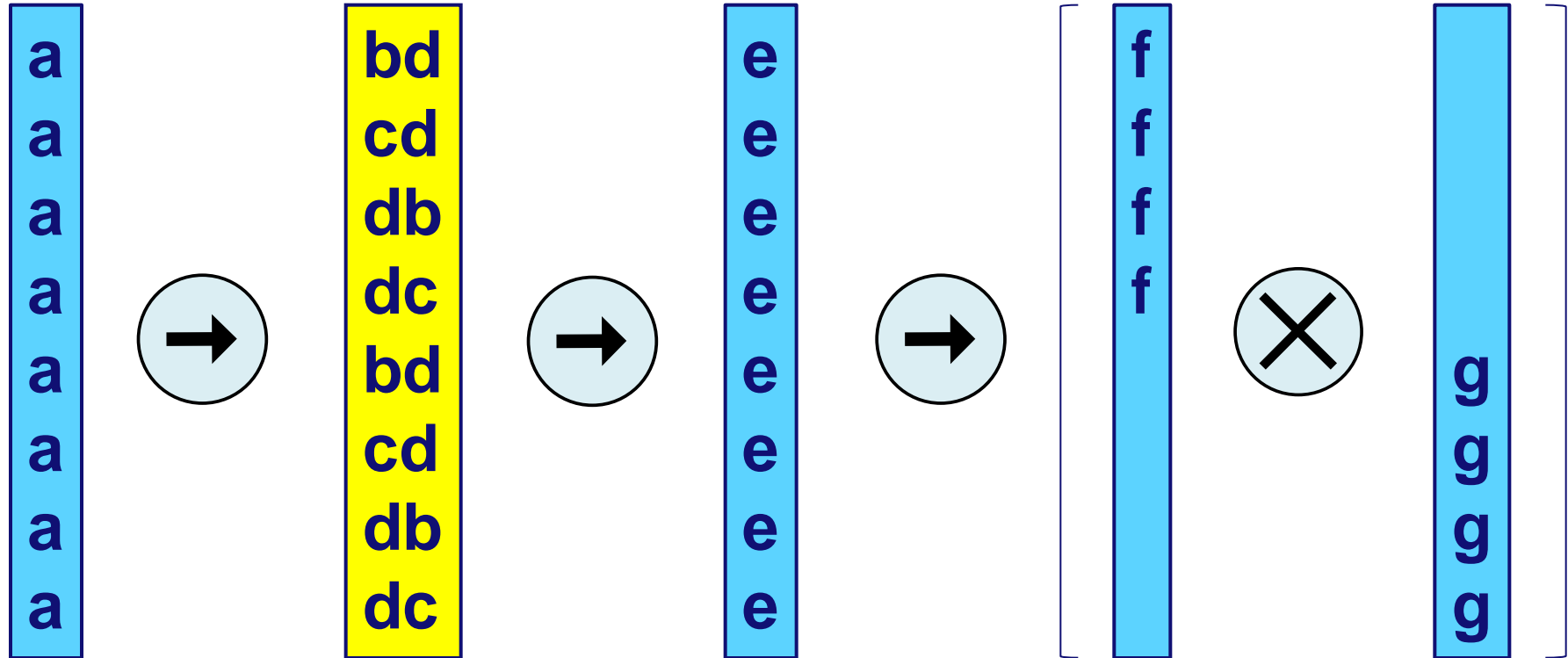
# Split $\{f,g\}$ into $\{f\}$ and $\{g\}$ using XOR decomposition



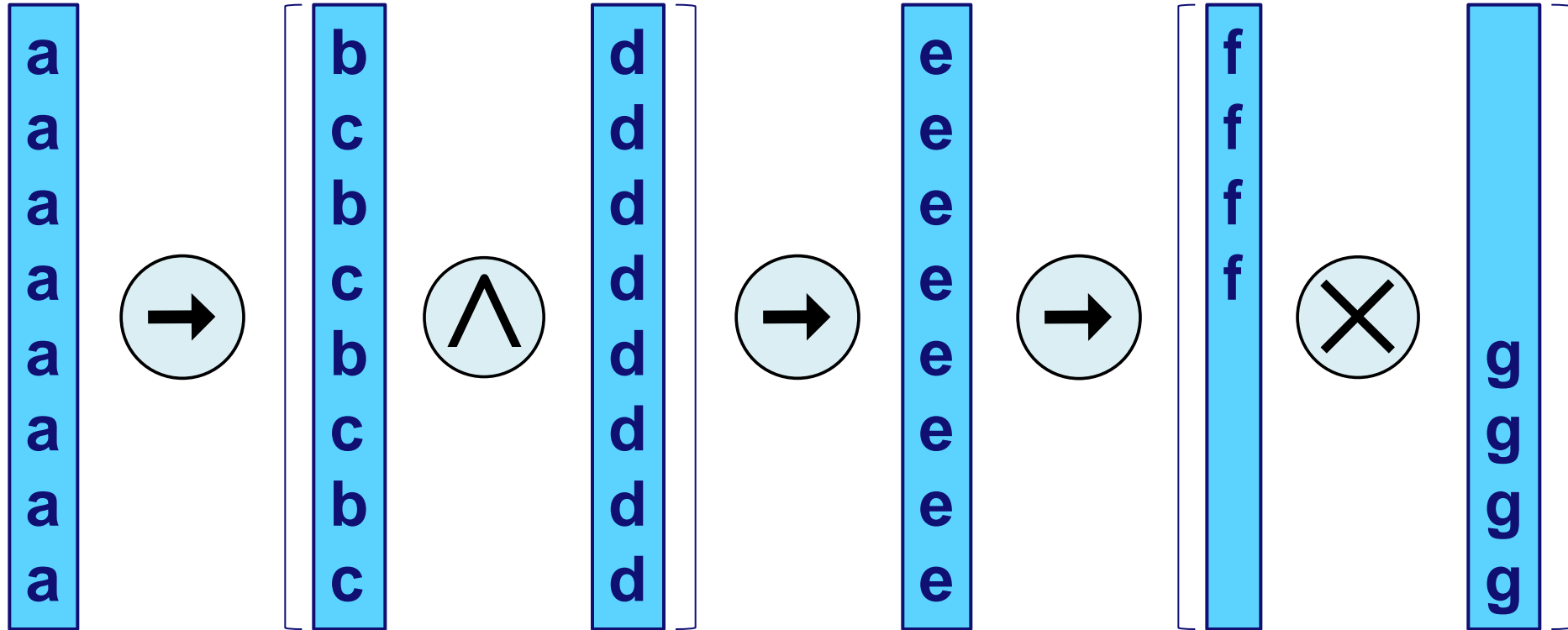
# Result



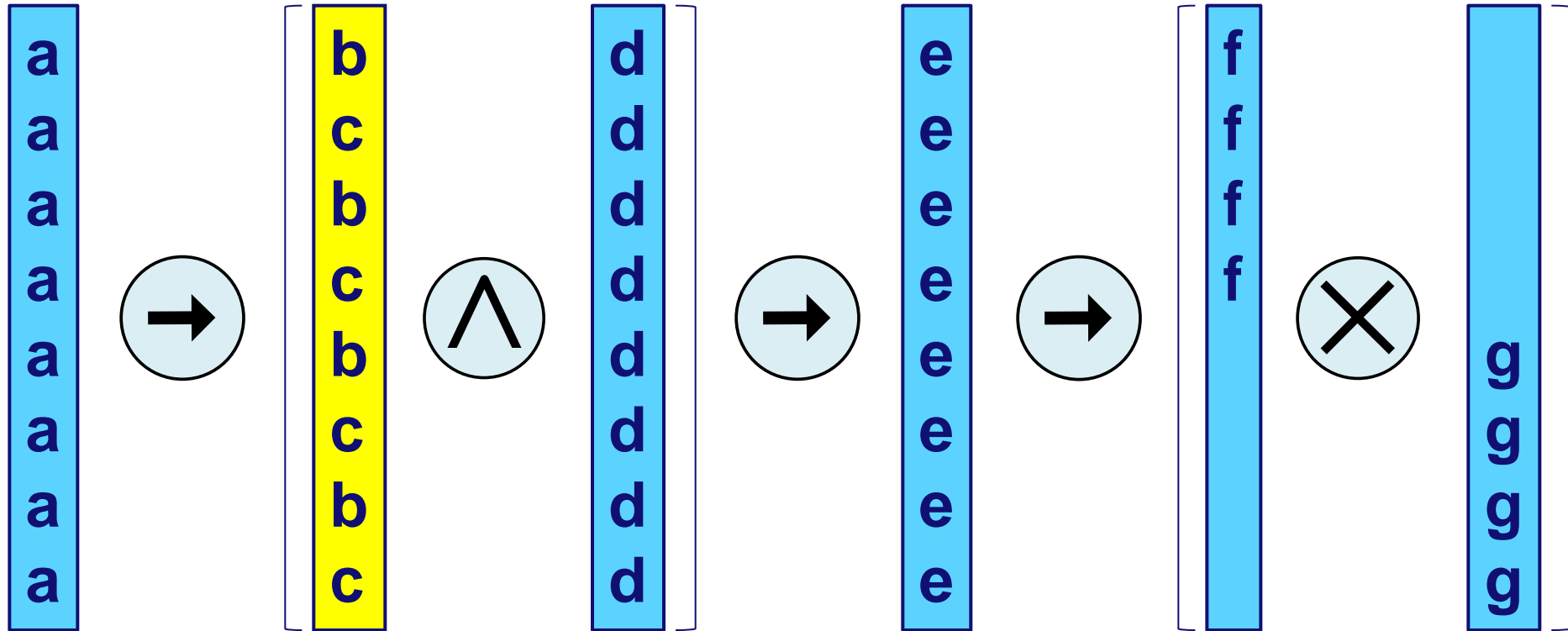
# Split $\{b,c,d\}$ into $\{b,c\}$ and $\{d\}$ using AND decomposition



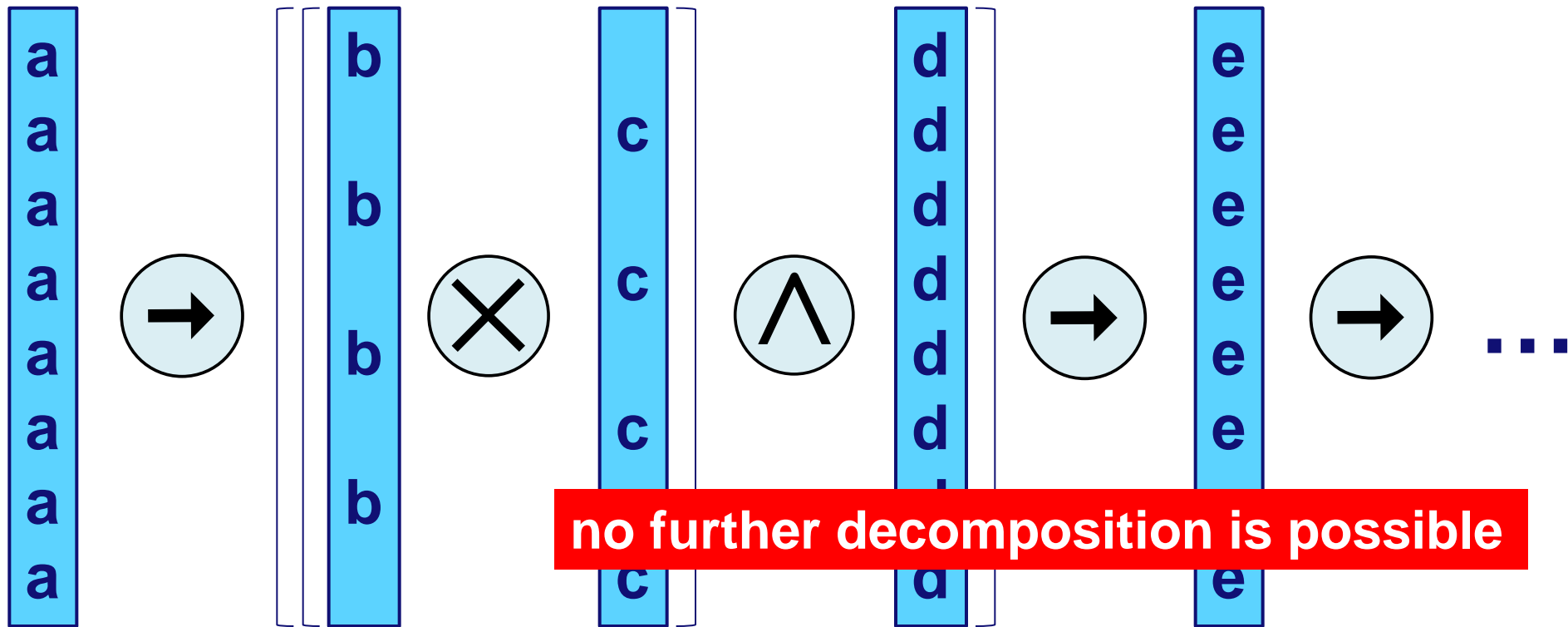
# Result



# Split $\{b,c\}$ into $\{b\}$ and $\{c\}$ using XOR decomposition

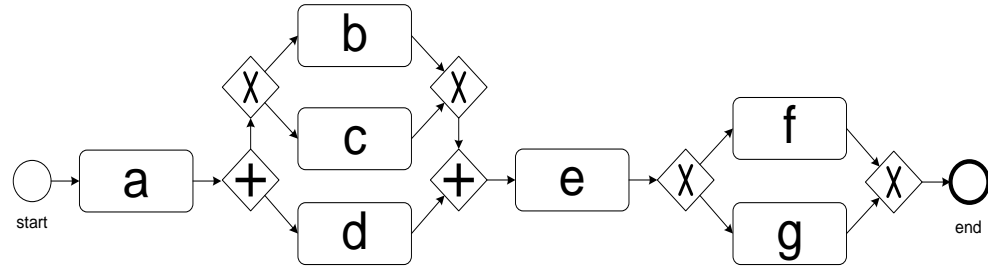
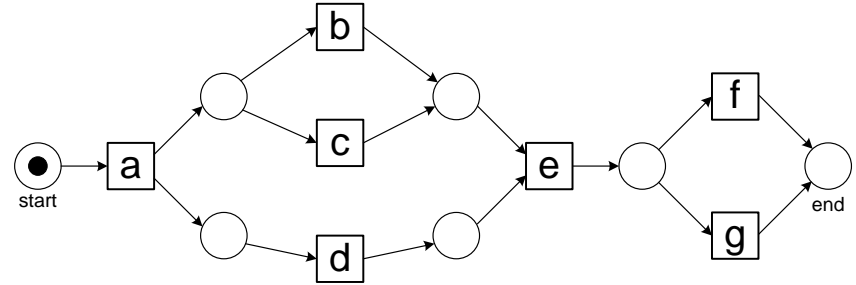
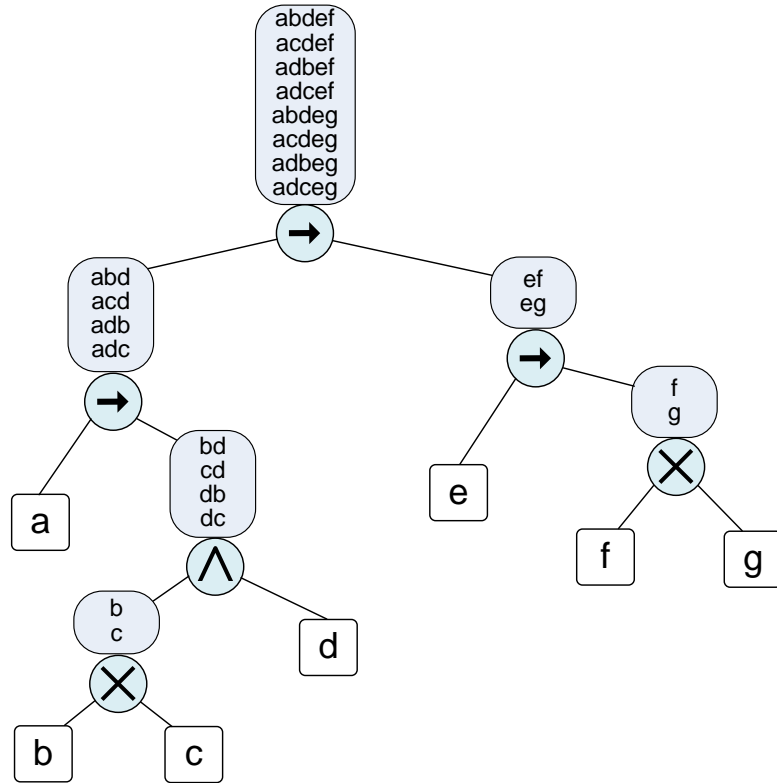


# Result





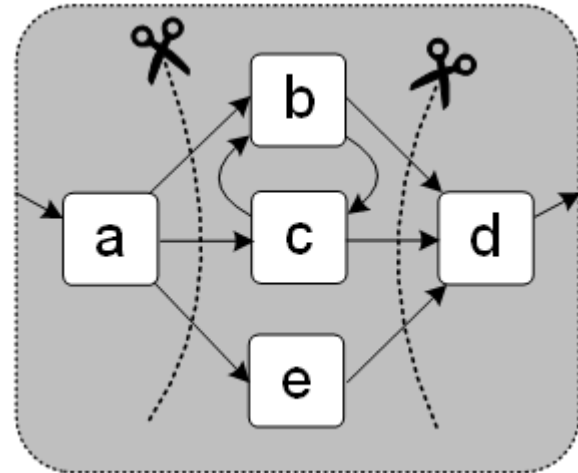
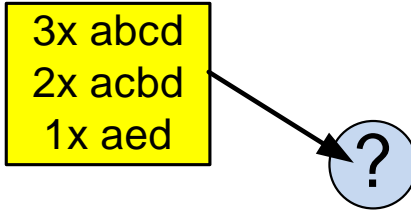
# Process tree



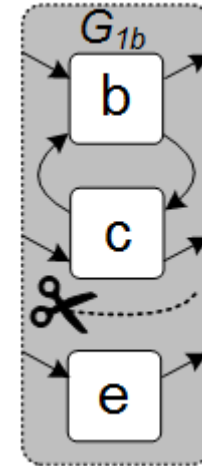
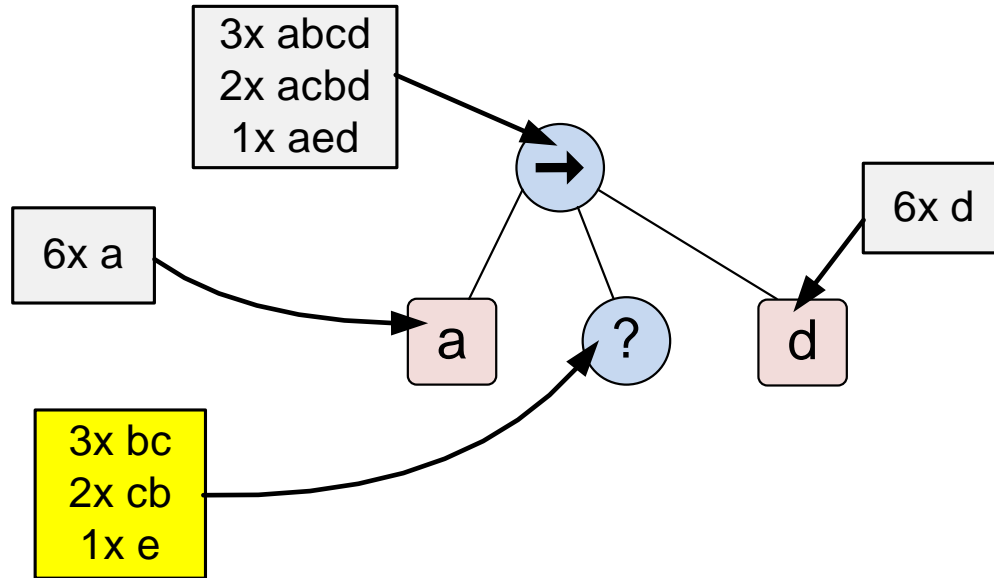
# An example log (6 traces, 23 events)

3x abcd  
2x acbd  
1x aed

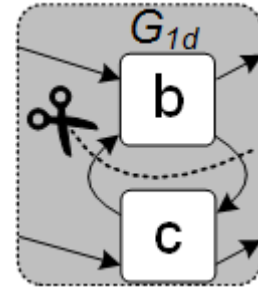
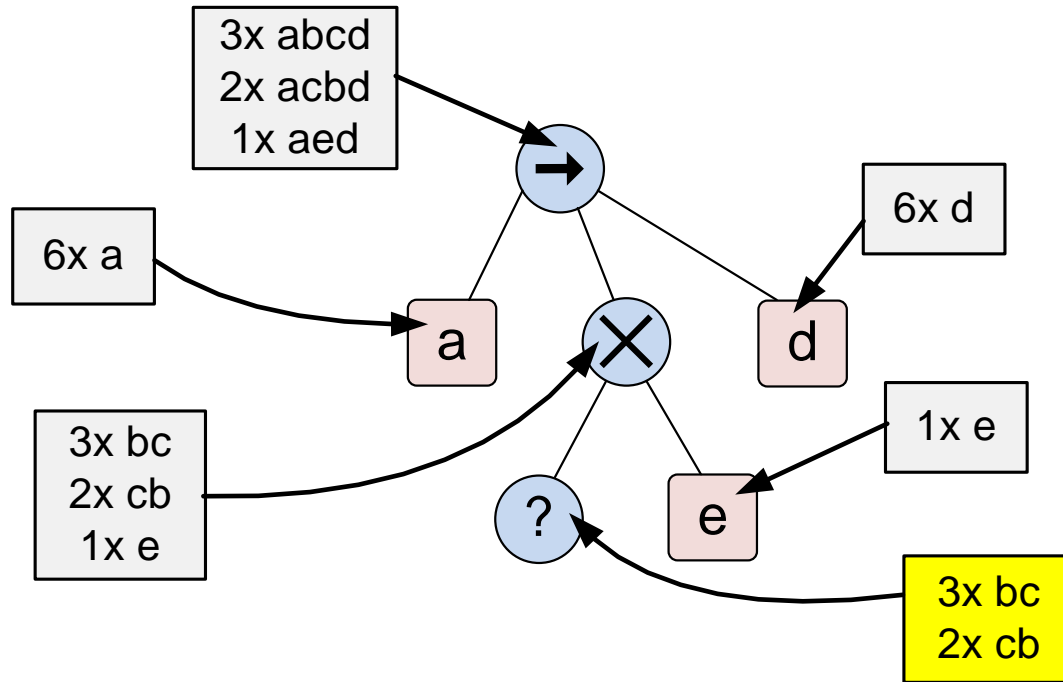
# How to split this event log?



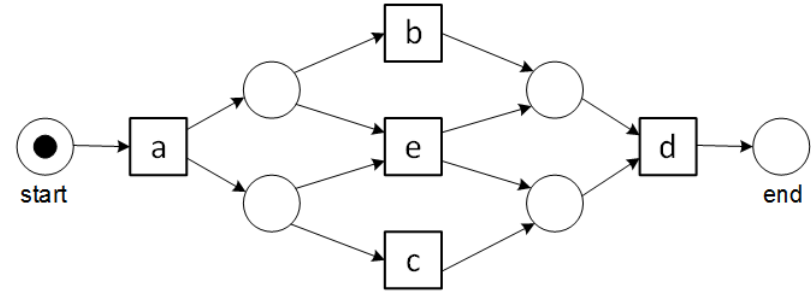
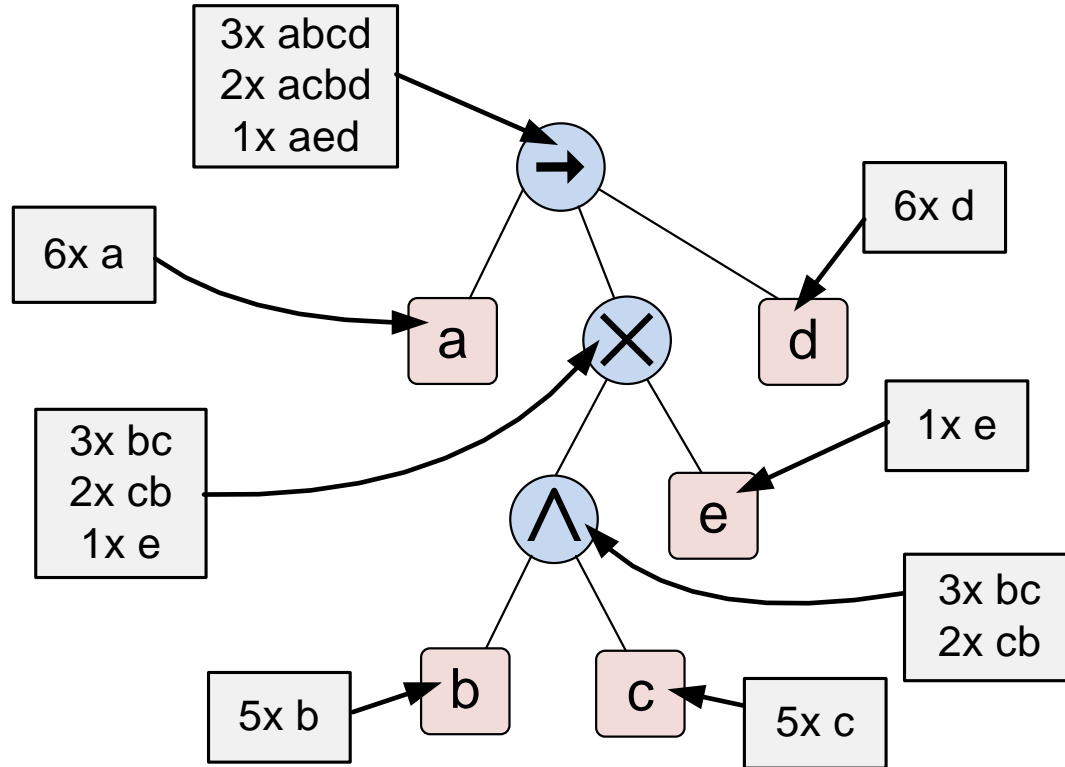
# How to split this event log?



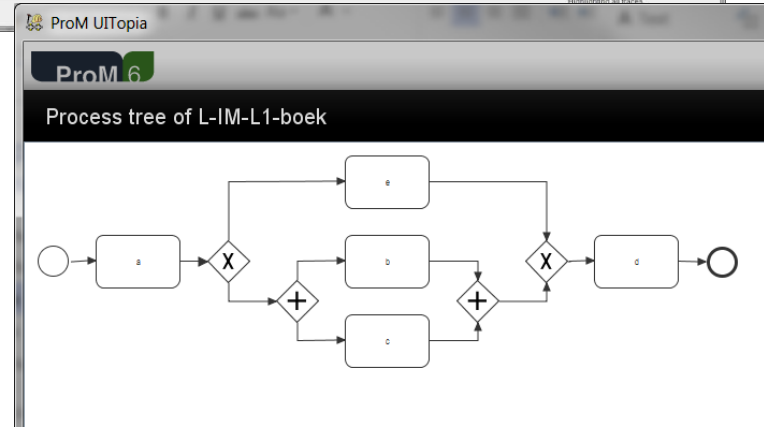
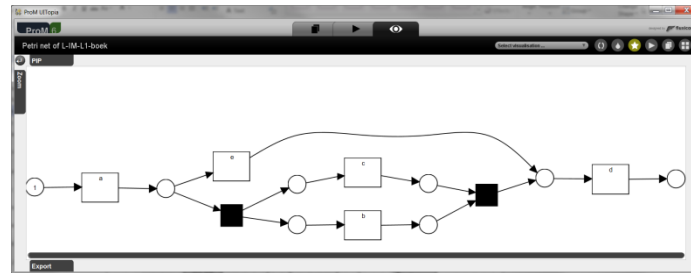
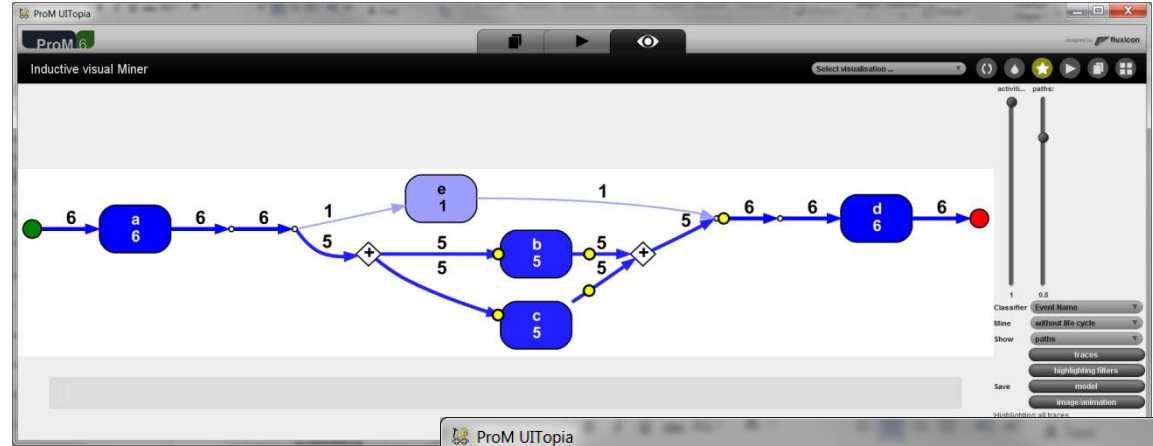
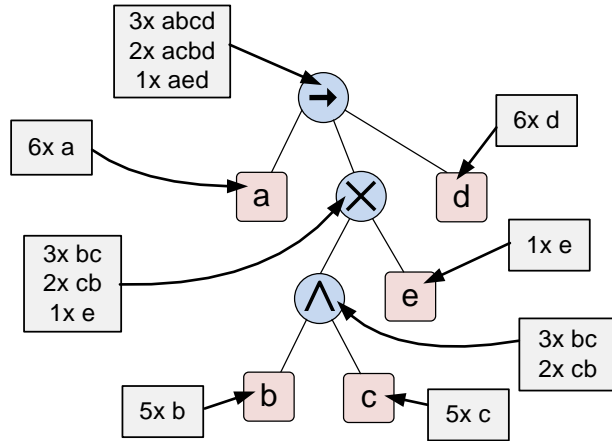
# How to split this event log?



# Final result



# In ProM



# Example log with loops (13 traces, 80 events)

3x abcd

4x acbd

2x abcefbcd

2x acbefbcd

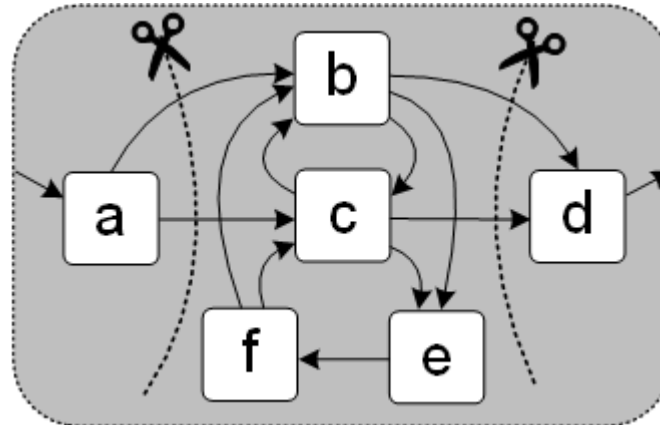
1x abcefcdb

1x acbefbcefcdb

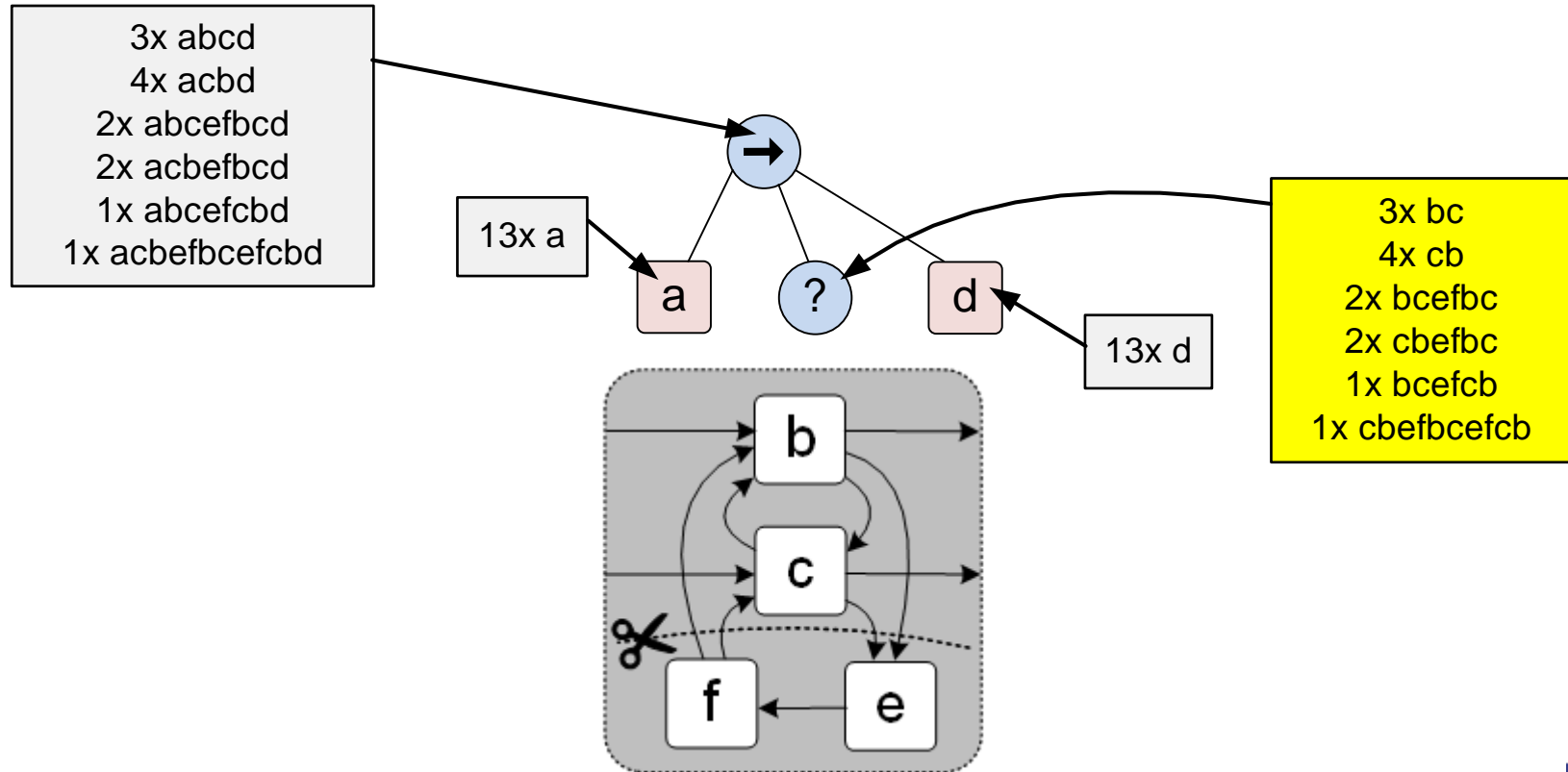


# How to split this event log?

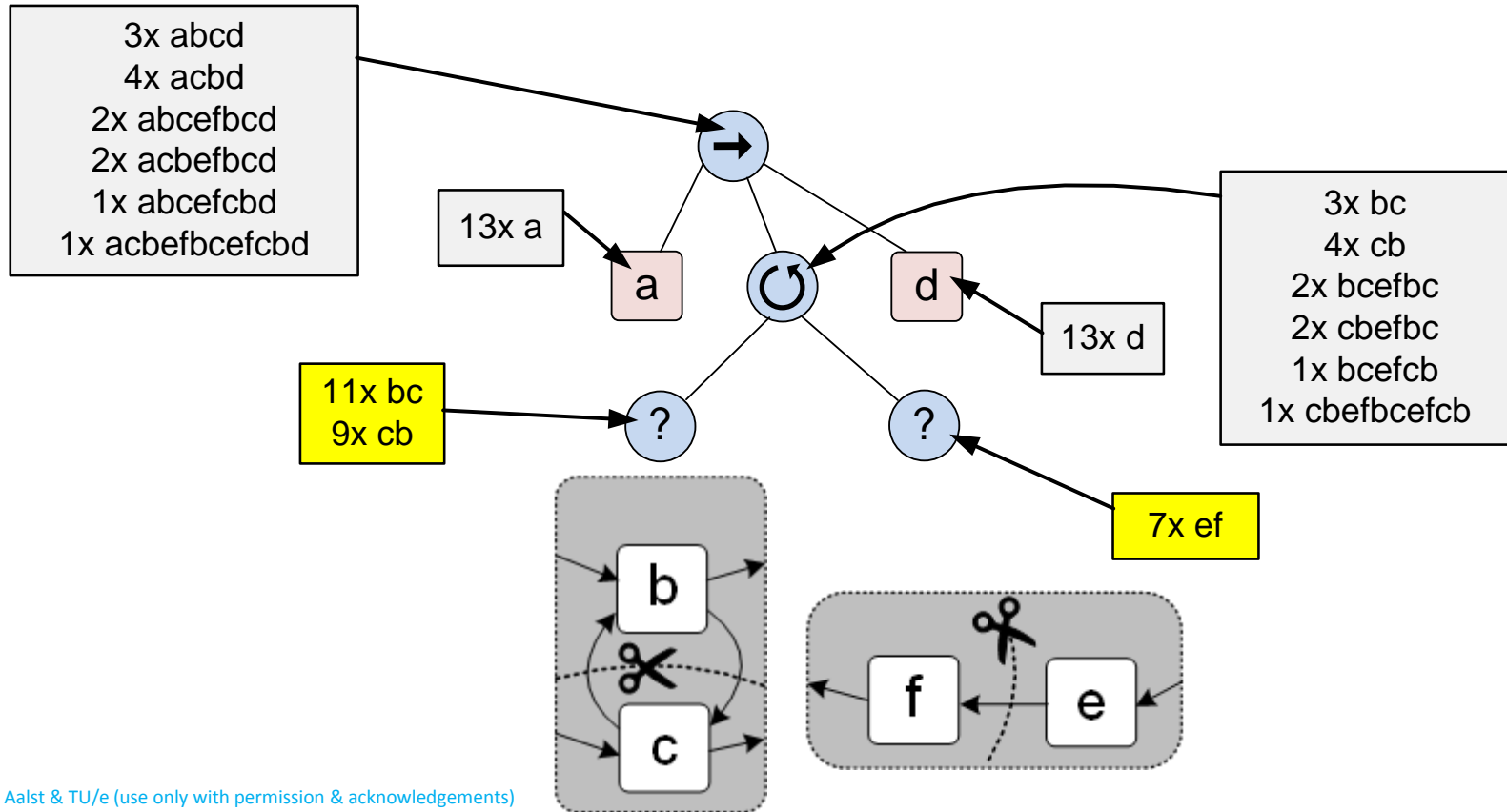
3x abcd  
4x acbd  
2x abcefbcd  
2x acbefbcd  
1x abcefcdb  
1x acbefbcefcdb



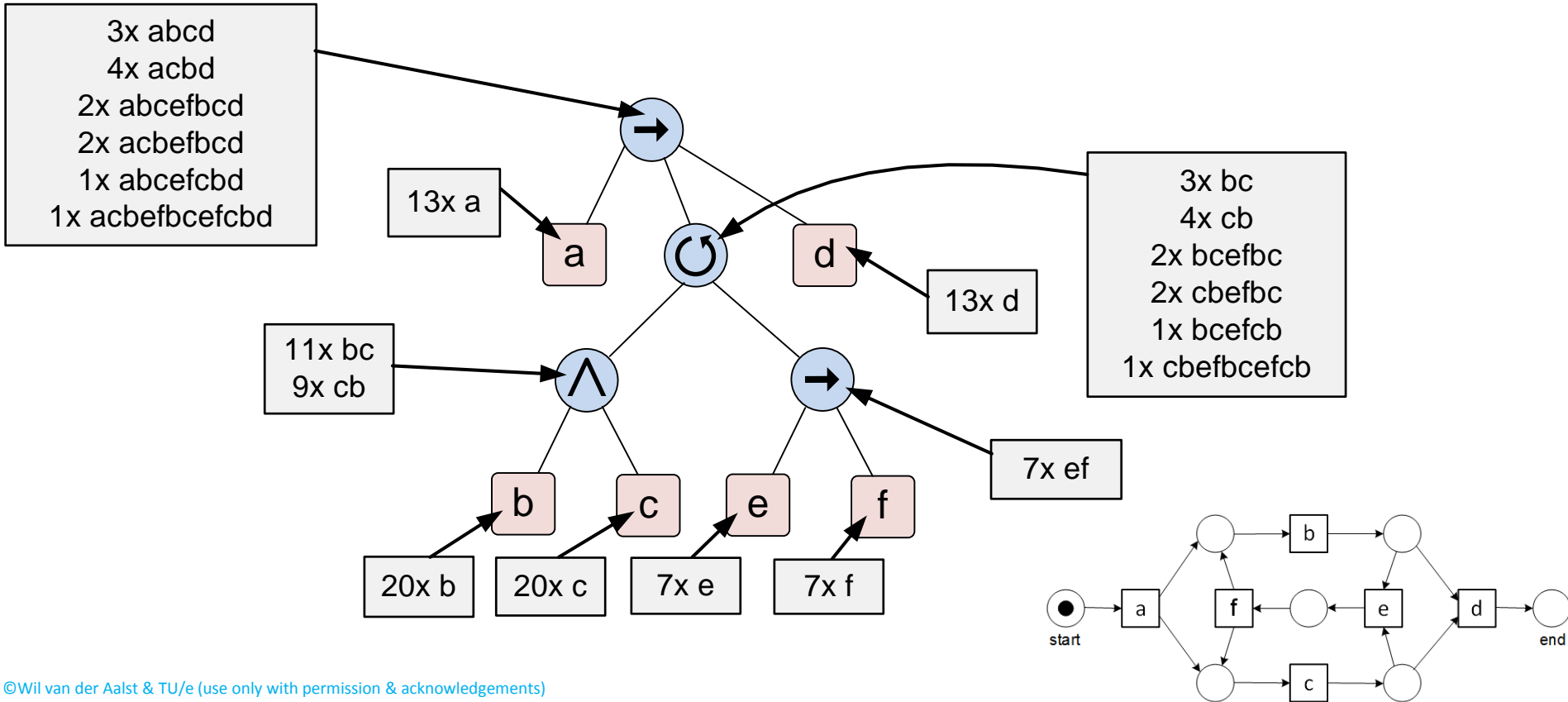
# How to split this event log?



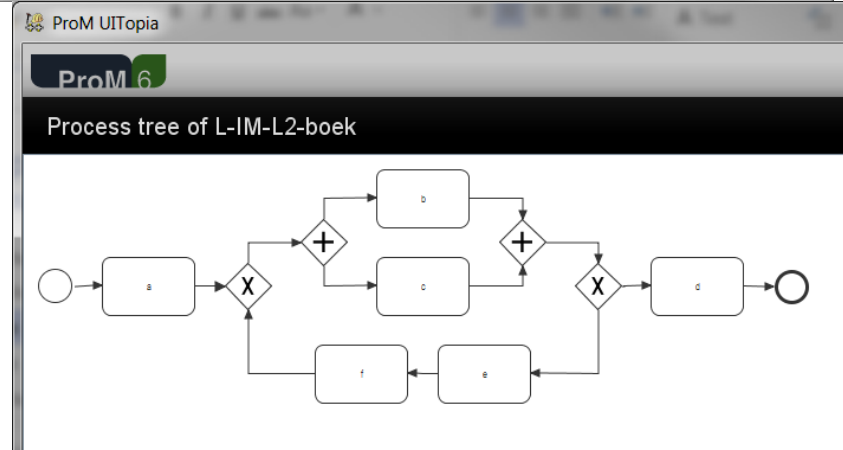
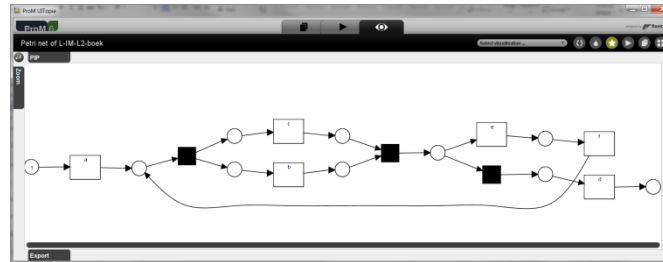
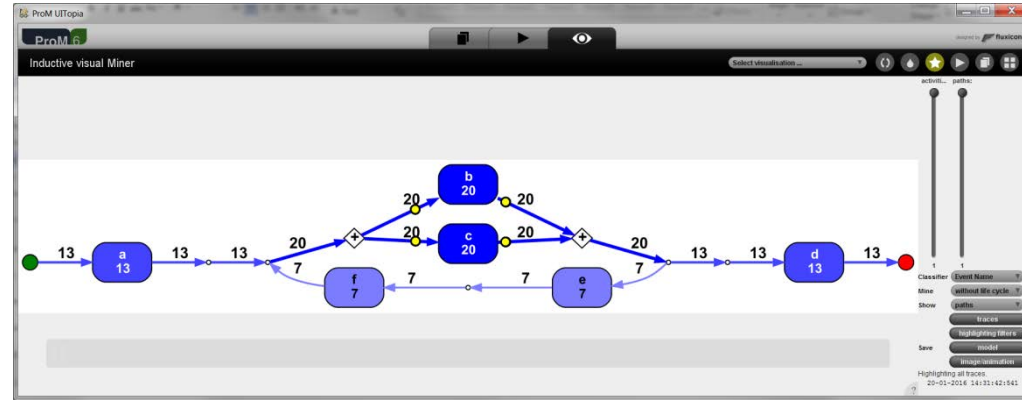
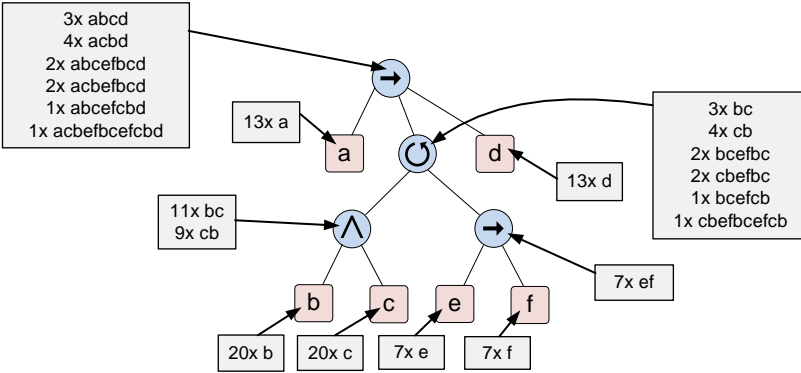
# How to split this event log?



# Result



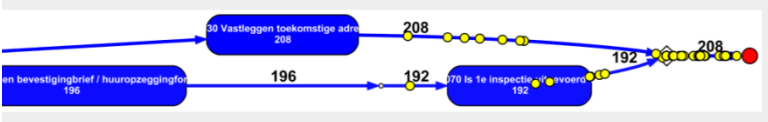
# In ProM



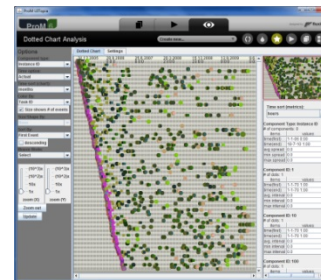
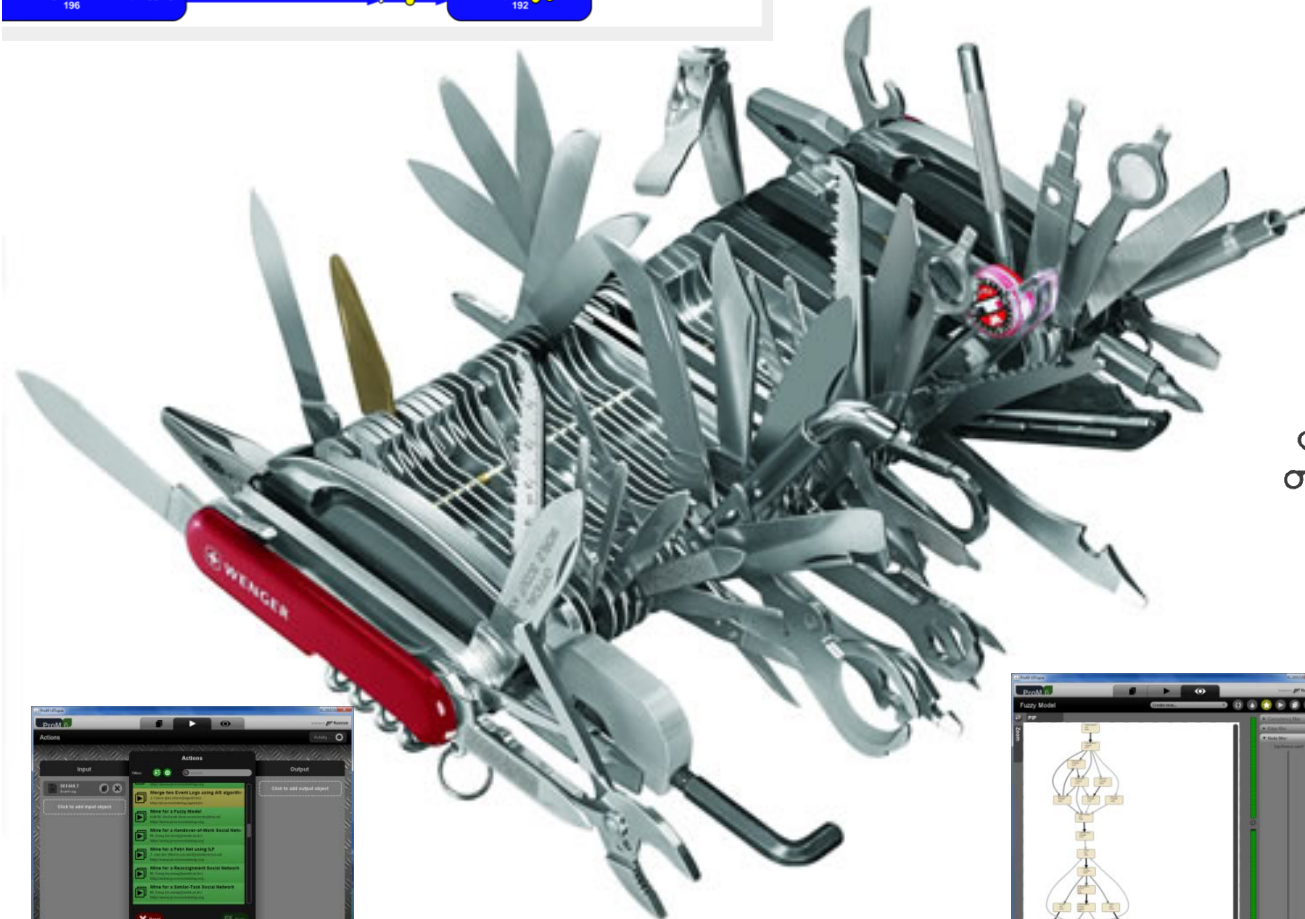
*Tool support*

# Process Mining Software



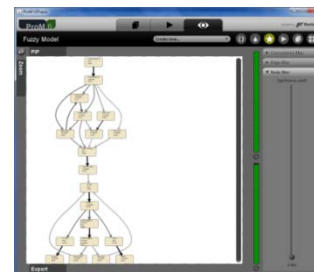


1500+ plug-ins available covering the whole process mining spectrum



100% FREE

>130k downloads

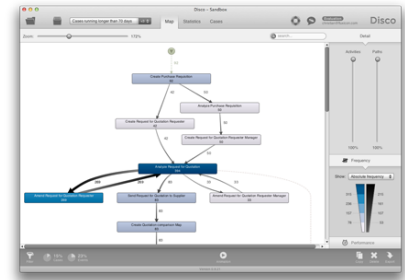






Disco

perceptive software  
a Lexmark company



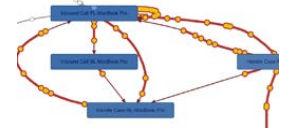
minit



Rialto

QPR  
Quality. Processes. Results.

celonis  
process mining



MAGNAVUE

ICARO  
TECH

XMPRO  
GET BETTER AT GETTING WORK DONE

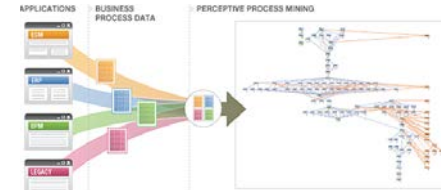
Software AG  
Process Performance Manager

Discover, analyze and monitor:  
the road to process improvement



my i nvenio

FUJITSU



The Transformation Company

TU/e



## Statistics views

Overview  
Global statisticsActivity  
Activity classesproduct  
Other attributeprod-price  
Other attribute

## Overview

Global statistics

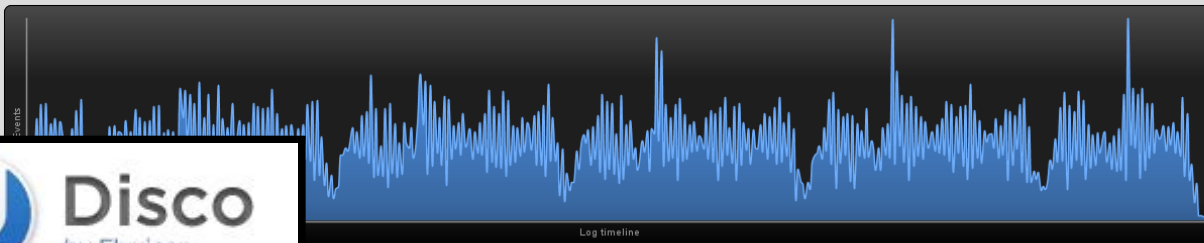
Events over time

Active cases over time

Case variants

Events per case

Case duration



Events 63,763

Cases 10,000

Activities 8

Median case duration 13.9 d

Mean case duration 14.9 d

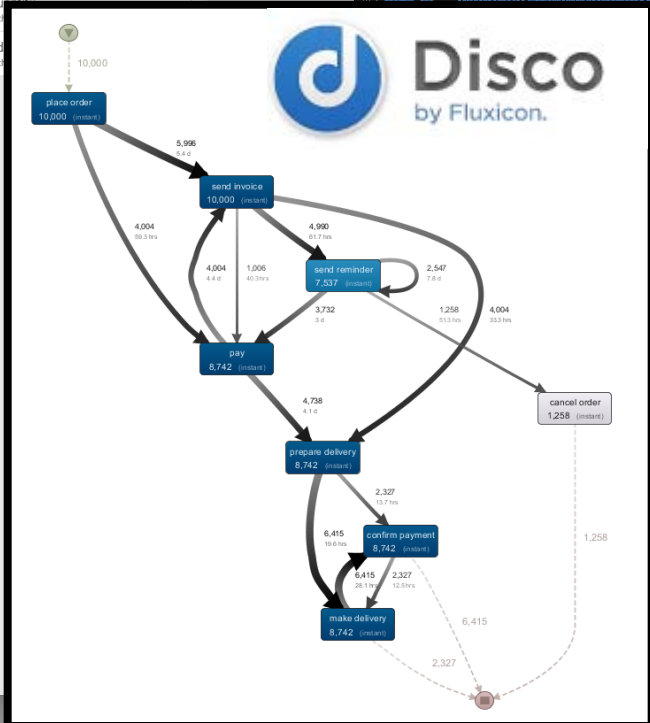
Start 05.01.2015 09:00:07

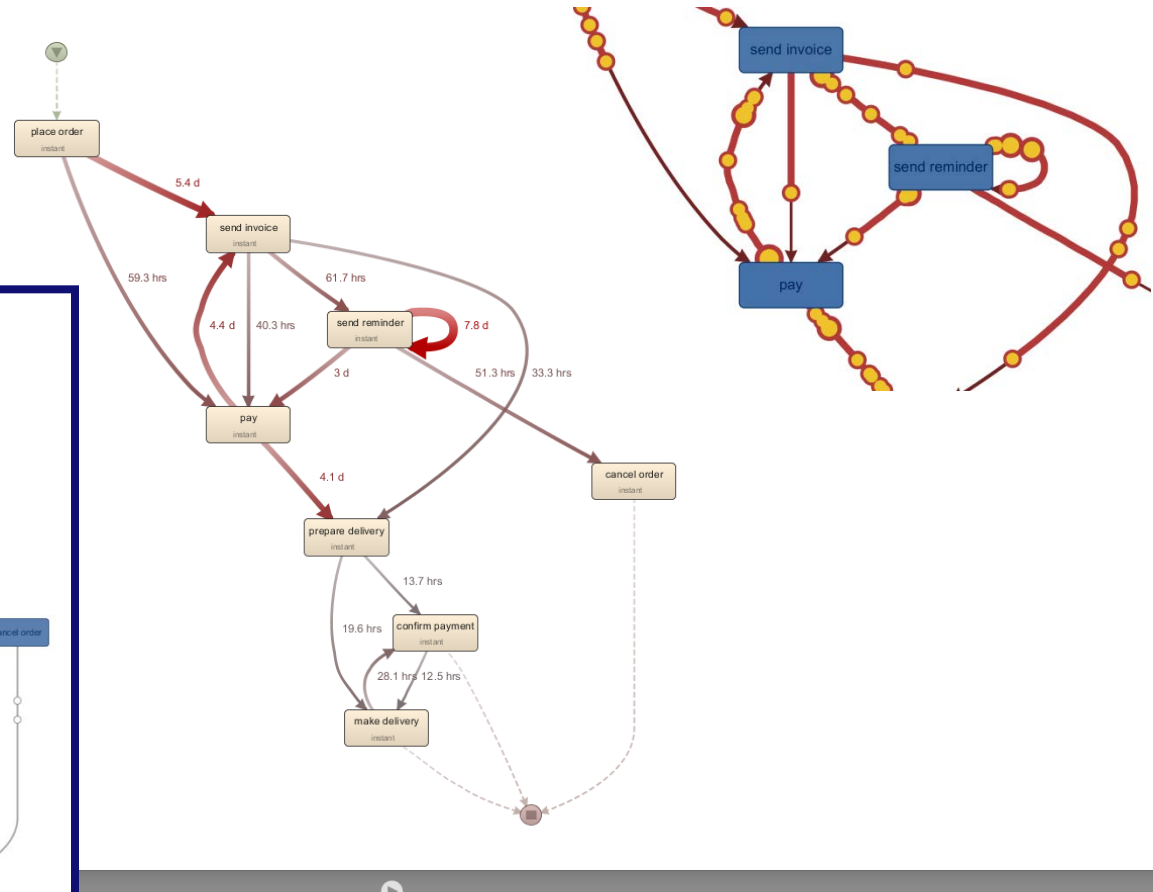
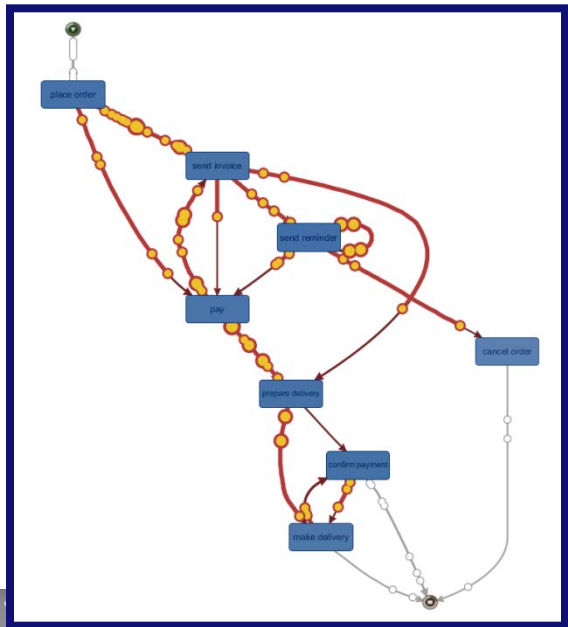
End 31.12.2019 14:46:02

Cases (10000)

Variants (9)

	Variant	Started	Finished	Duration
7		05.01.2015 09:00:07	26.01.2015 16:42:28	21 days, 7 hours
1		05.01.2015 10:18:21	15.01.2015 15:52:30	10 days, 5 hours
4		05.01.2015 11:54:49	09.01.2015 18:38:58	4 days, 6 hours
3		05.01.2015 14:07:45	22.01.2015 13:18:30	16 days, 23 hours
1		05.01.2015 15:33:38	12.01.2015 17:27:36	7 days, 1 hour
5		05.01.2015 17:25:23	02.02.2015 12:31:09	27 days, 19 hours
4		05.01.2015 19:08:53	15.01.2015 14:56:54	9 days, 19 hours
9		05.01.2015 21:54:00	13.01.2015 15:49:53	7 days, 17 hours
4		06.01.2015 07:25:13	15.01.2015 11:27:50	9 days, 4 hours
1		06.01.2015 10:09:51	15.01.2015 19:15:18	9 days, 9 hours
1		06.01.2015 11:37:49	14.01.2015 09:14:28	7 days, 21 hours
4		06.01.2015 13:33:45	14.01.2015 11:30:05	7 days, 21 hours
4		06.01.2015 15:25:38	13.01.2015 12:25:34	6 days, 20 hours
2		06.01.2015 17:09:23	22.01.2015 18:59:10	16 days, 1 hour
3		06.01.2015 18:36:53	22.01.2015 14:39:39	15 days, 20 hours
8		06.01.2015 21:26:54	26.01.2015 17:16:02	19 days, 19 hours
1		07.01.2015 04:42:36	16.01.2015 10:17:14	9 days, 5 hours
3		07.01.2015 10:10:58	21.01.2015 17:31:29	14 days, 7 hours
8		07.01.2015 11:40:04	28.01.2015 10:27:12	20 days, 22 hours
9		07.01.2015 13:38:15	13.01.2015 13:22:15	5 days, 23 hours
1		07.01.2015 15:34:37	19.01.2015 09:11:23	11 days, 17 hours
1		07.01.2015 17:27:21	16.01.2015 09:09:25	8 days, 15 hours
5		07.01.2015 19:12:50	03.02.2015 14:34:33	26 days, 19 hours
6		07.01.2015 22:01:54	19.01.2015 13:15:02	11 days, 15 hours
8		08.01.2015 07:12:36	28.01.2015 10:41:14	20 days, 3 hours
3		08.01.2015 09:55:59	26.01.2015 15:52:42	18 days, 5 hours
6		08.01.2015 12:10:05	15.01.2015 13:54:59	7 days, 1 hour
1		08.01.2015 13:38:17	14.01.2015 12:30:26	5 days, 22 hours
5		08.01.2015 15:34:42	02.02.2015 14:10:36	24 days, 22 hours
2		08.01.2015 17:27:31	29.01.2015 11:26:06	20 days, 17 hours
3		08.01.2015 19:13:09	26.01.2015 14:16:02	17 days, 19 hours
4		08.01.2015 22:02:32	20.01.2015 10:35:40	11 days, 12 hours

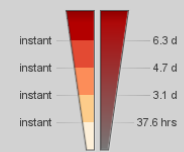




Frequency

Performance

Show: Mean duration



Add secondary metrics



Animation

Version 1.9.1



Copy

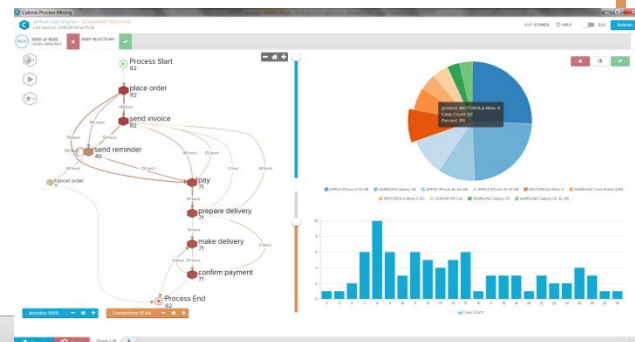
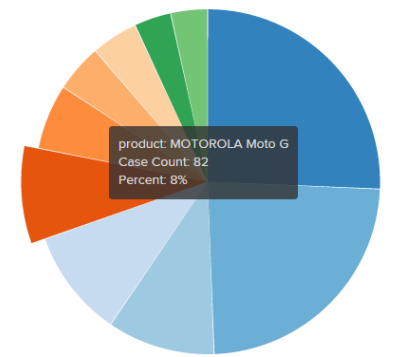
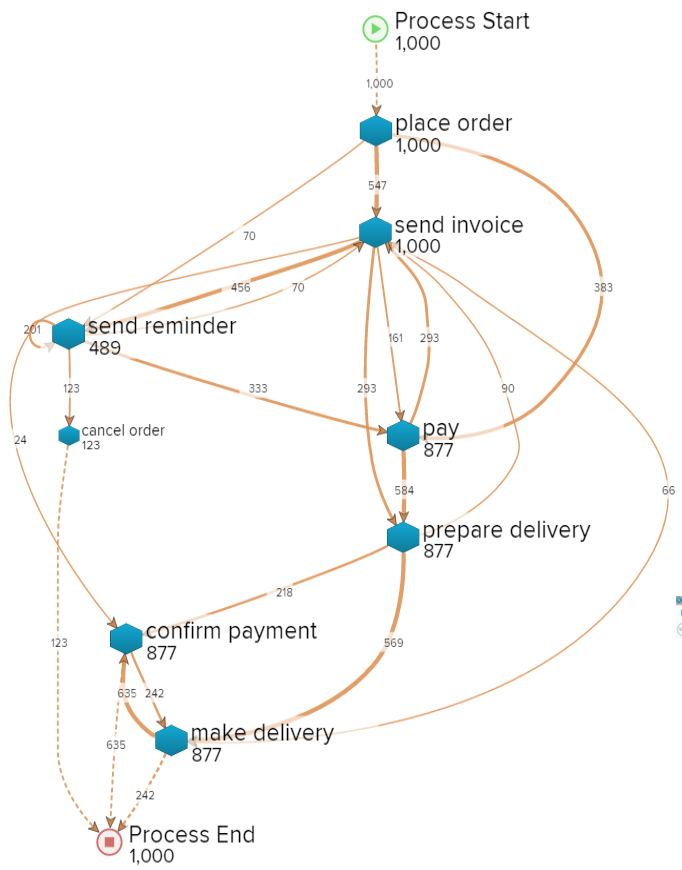
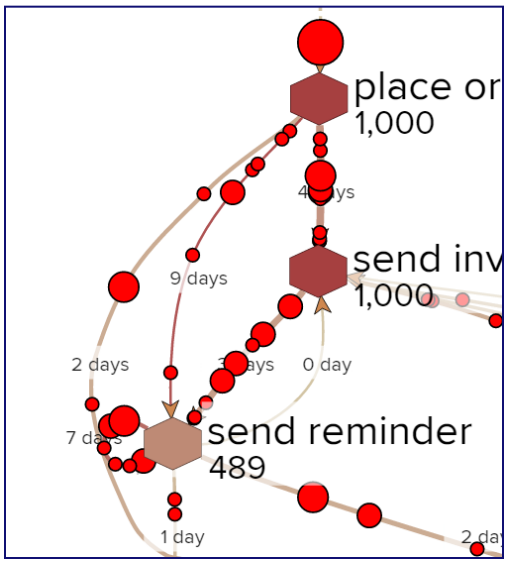
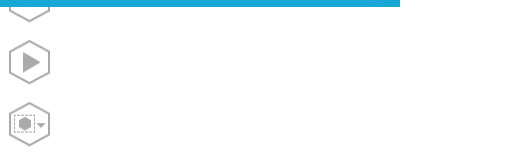


Delete



Export







## Process map

event log 10000 ▾



Manage view



Project hub



Repository



Settings

vanderAalstWil  
Trial ▾

minit

cases  
100%events  
100%

10k place order

4004

5996

8742 pay

1006

4004

3732

10k send invoice

4990

7537 send reminder

2547

1258

1258 cancel order

4738

4004

8742 prepare delivery

2327

8742 confirm payment

8742 make delivery

6415

2327

6415

## Customize

Process view

Social view

Missing attribute of type Resource.

Show terminal nodes



Snap to backbone



Left to right



Highlight predecessor/successor activities



Activities

100 %

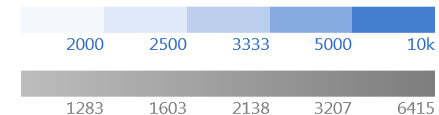
Paths

100 %

Frequency

Performance

Event count ▾



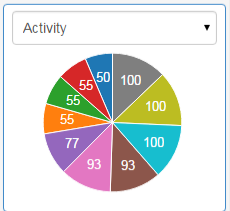
Visualize

reviewing

100 cases 2278 events

Resolution Tokens

Sat Dec 31, 2005 23:00:00



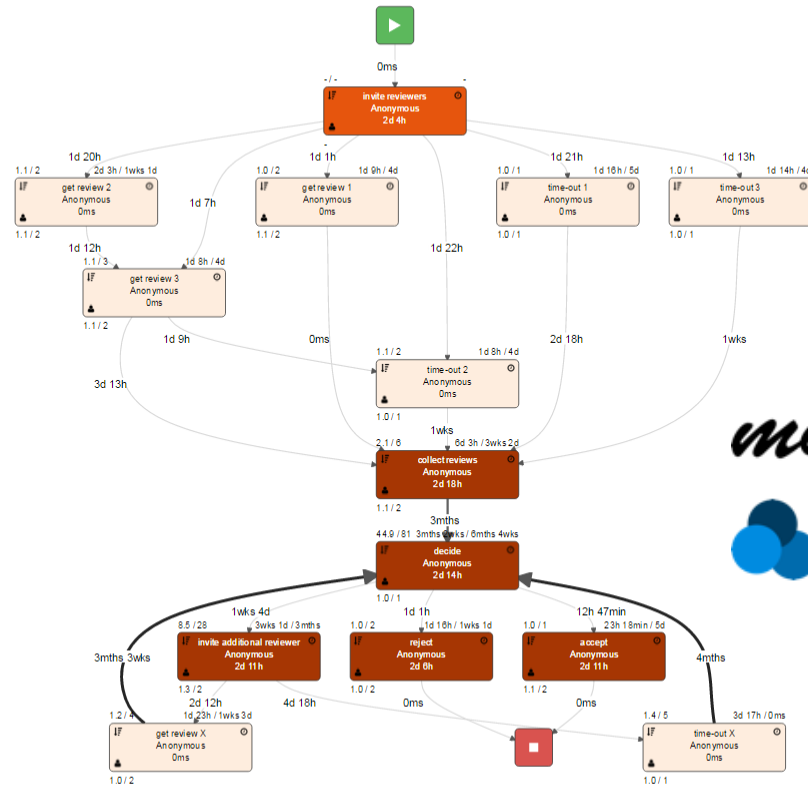
Model's details

Activities

Relations

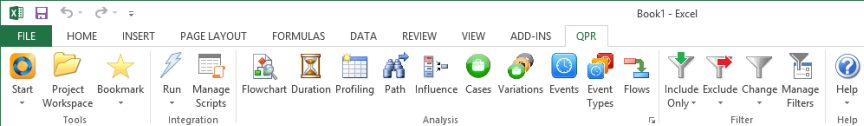
Resource focus

Portrait Landscape



0ms  
13h 20min  
1d 2h  
1d 45h

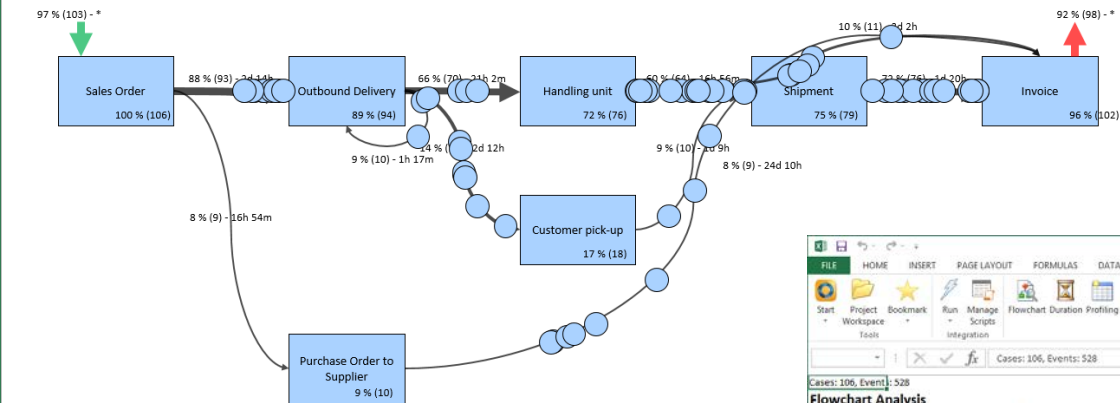




Cases: 106, Event: 528

## Flowchart Analysis

Median duration: 7d 19h, Average duration: 14d 10h



## Settings

### Flowchart Animation

#### Controls



0 103d 10h

5d 8h

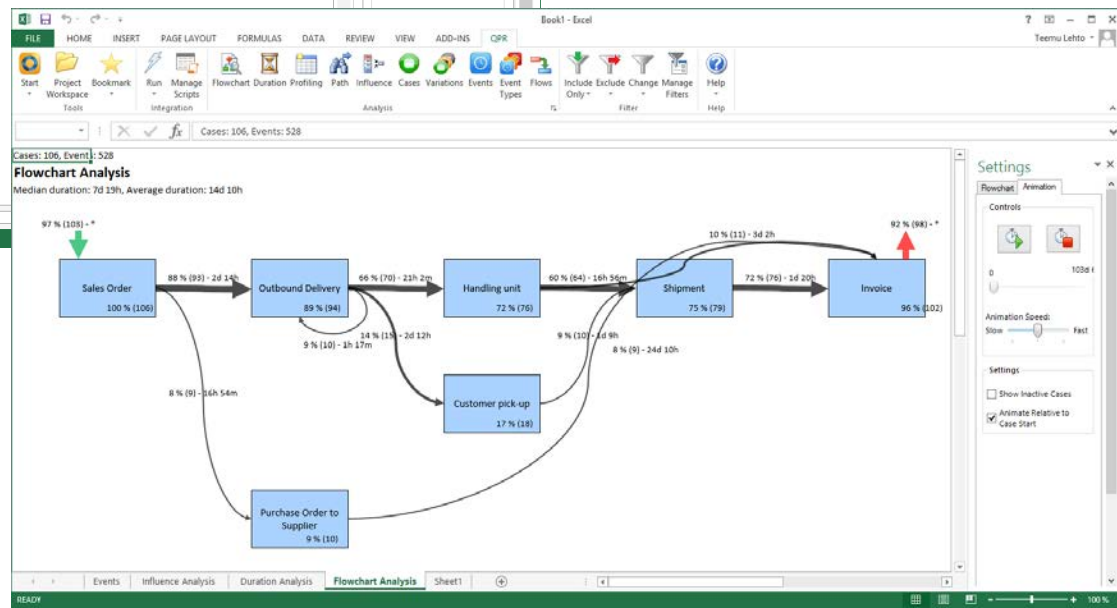
Animation Speed:

Slow Fast

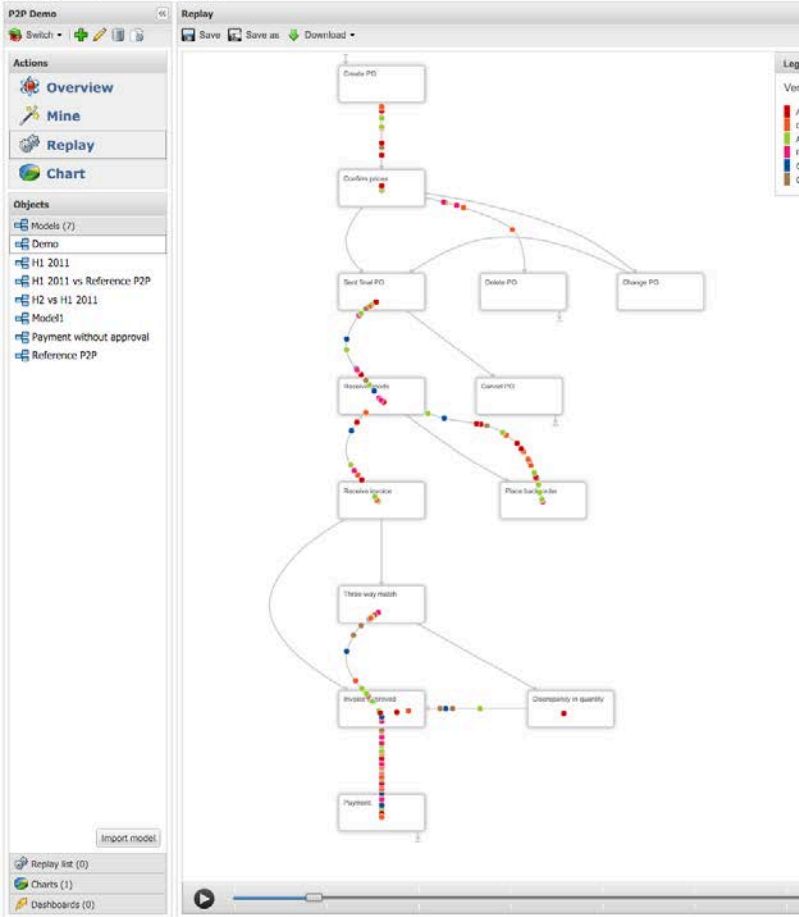
Settings

☐ Show Inactive Cases

READY



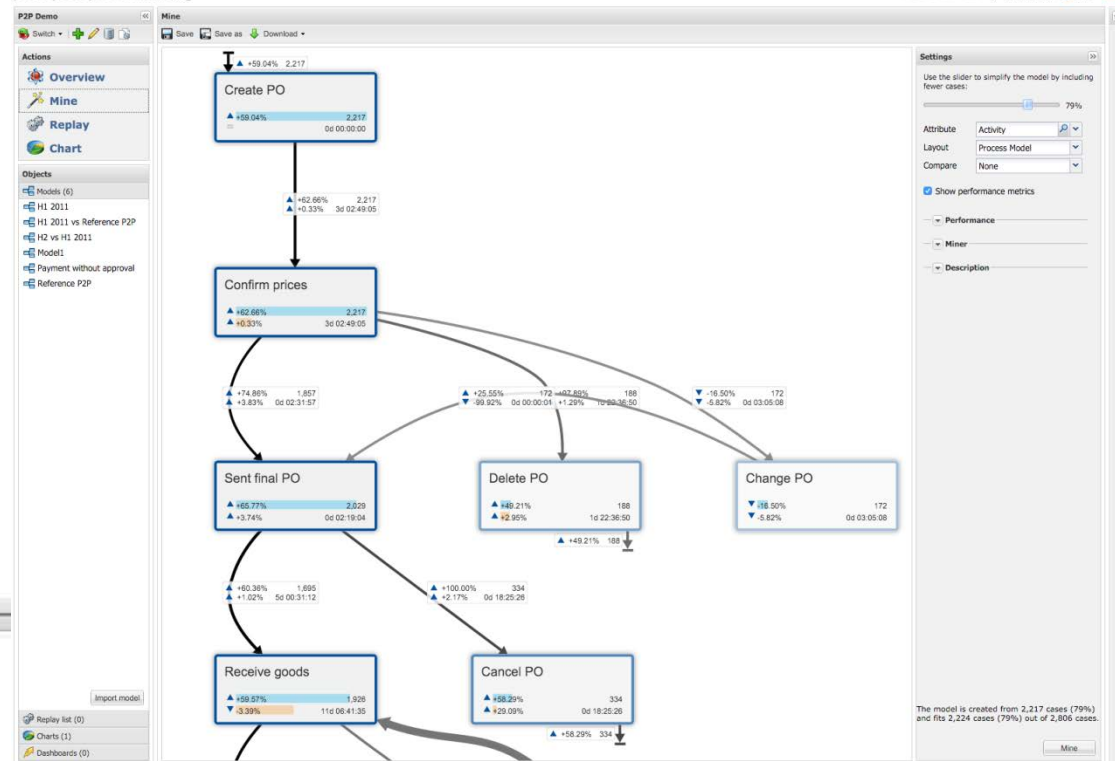




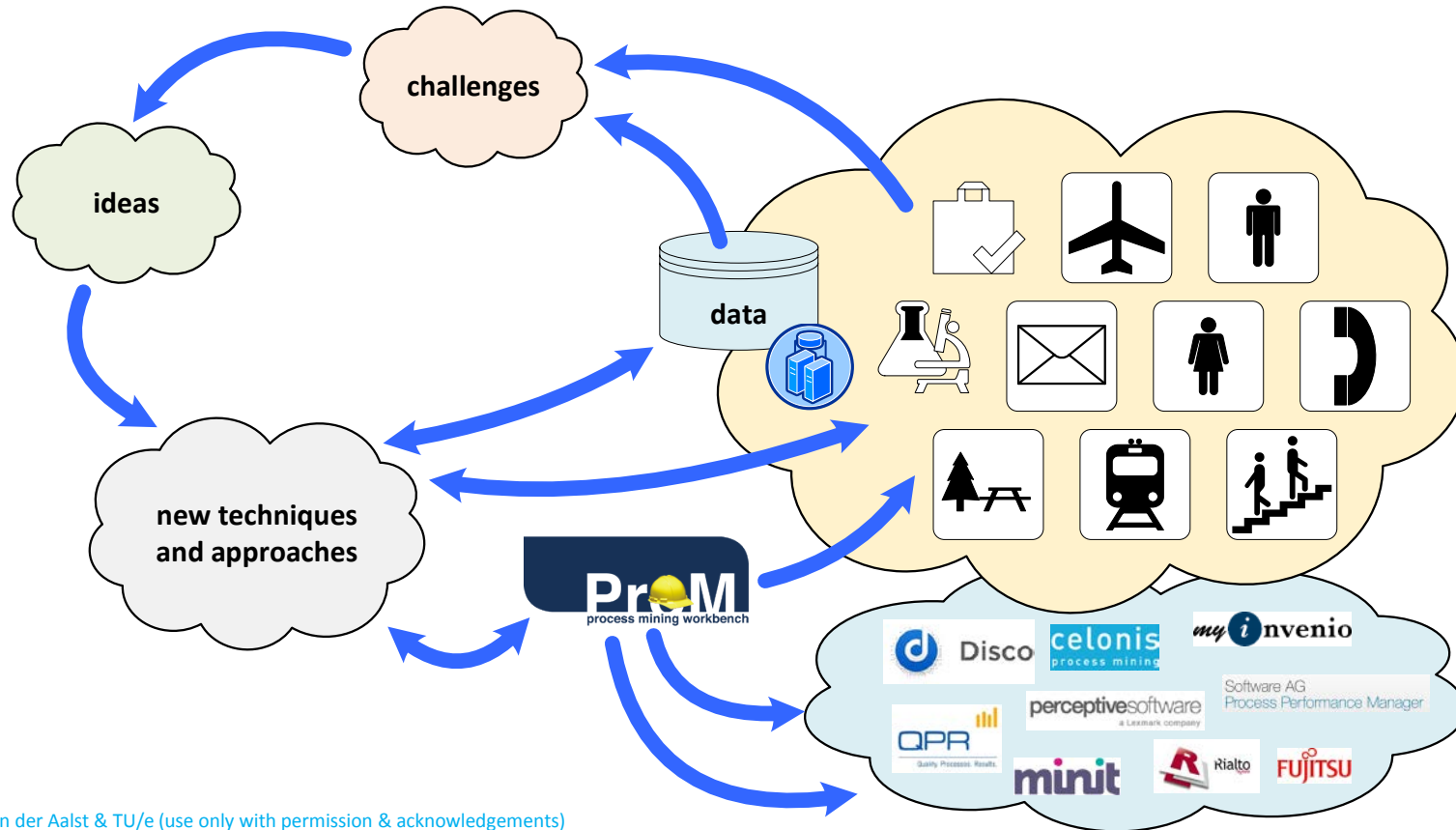
Lexmark™

Perceptive Process Mining

## perceptive process mining

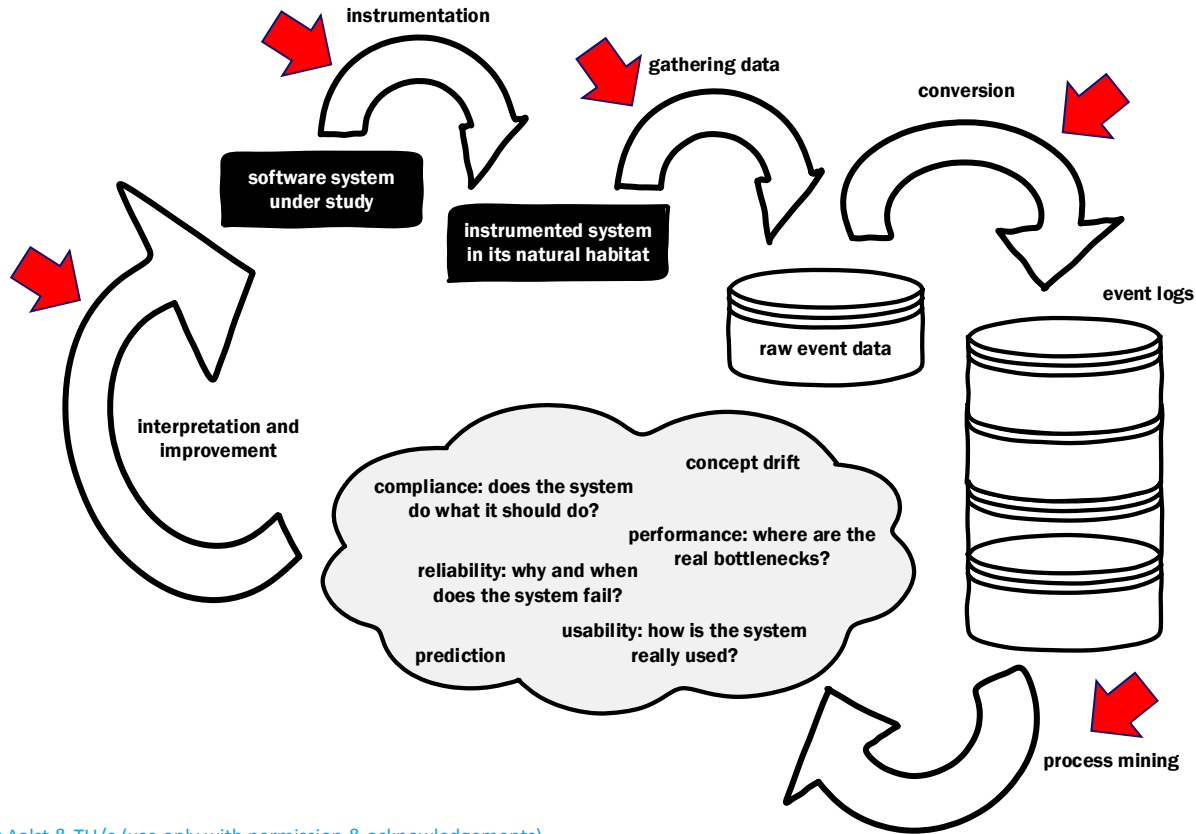


# Interaction with industry

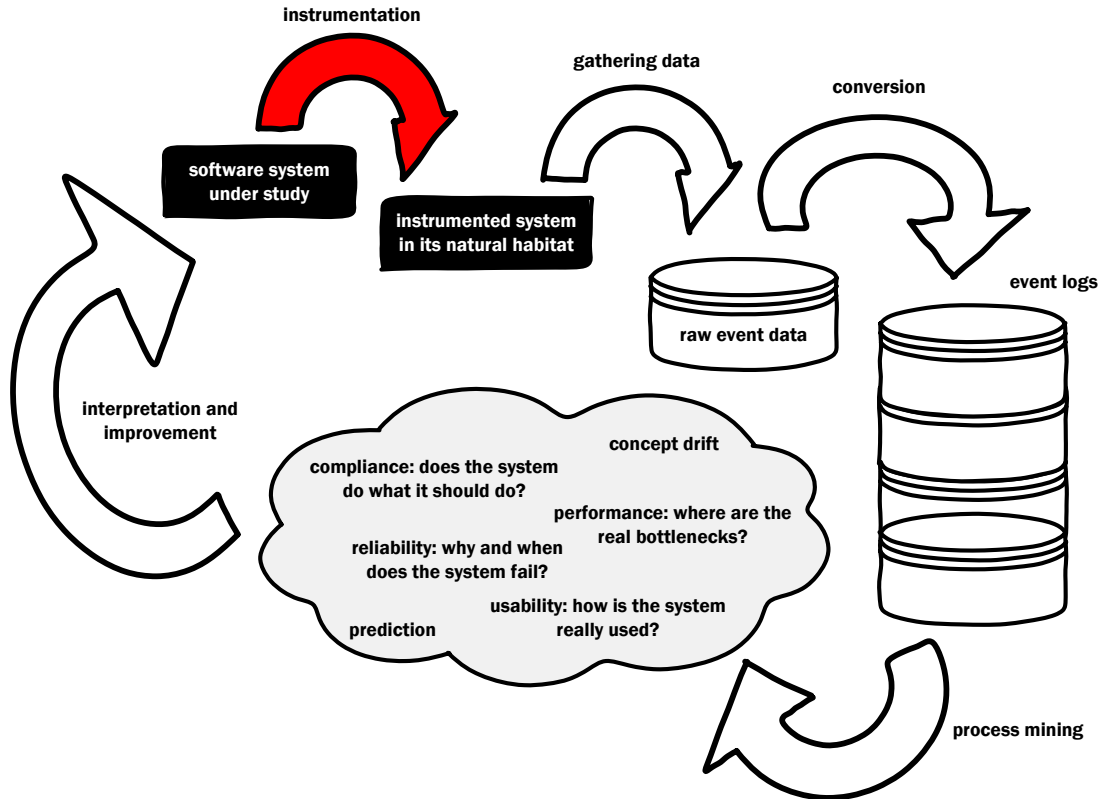


*Transition*

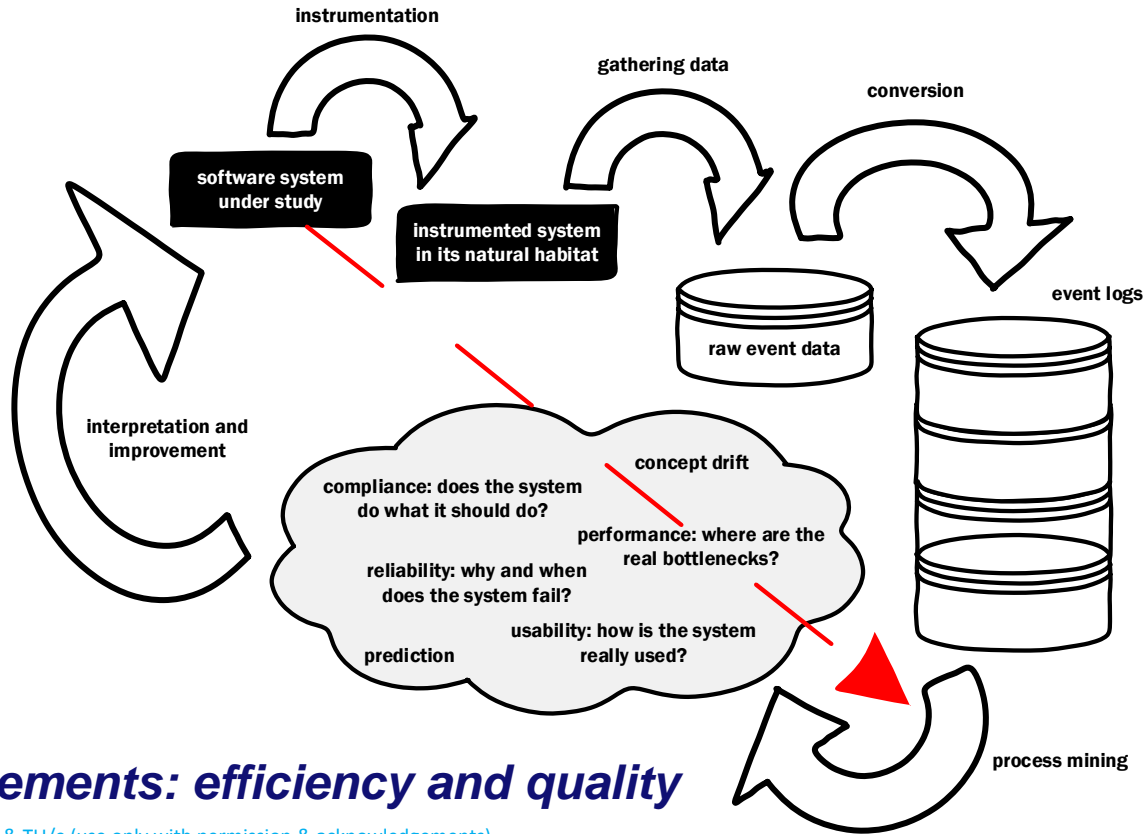
# What is specific for software?



# Opportunity: Control what is recorded



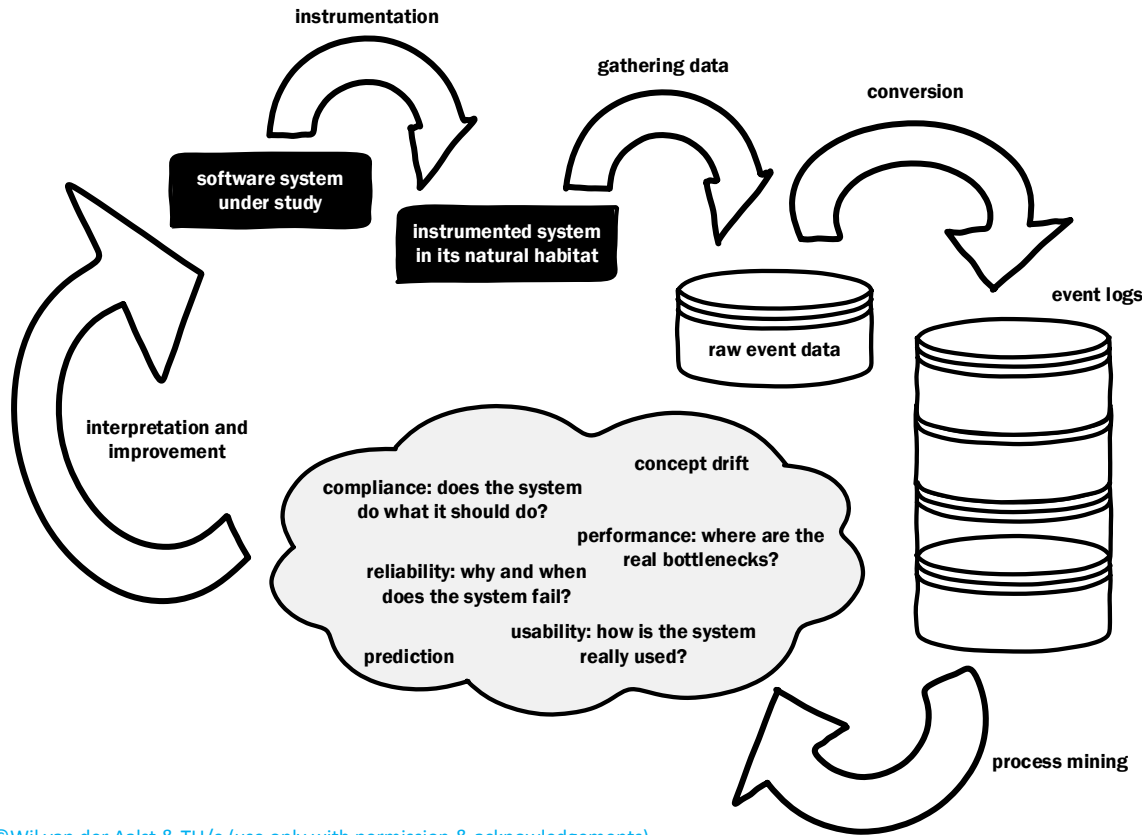
# Opportunity: Exploiting known structure



- **components**
- **services**
- **interfaces**
- ...

W.M.P. van der Aalst, A. Kalenkova, V. Rubin, and E. Verbeek. Process Discovery Using Localized Events. In *Petri Nets 2015*, volume 9115 of *Lecture Notes in Computer Science*, pages 287–308. Springer-Verlag, Berlin, 2015.

# Challenges



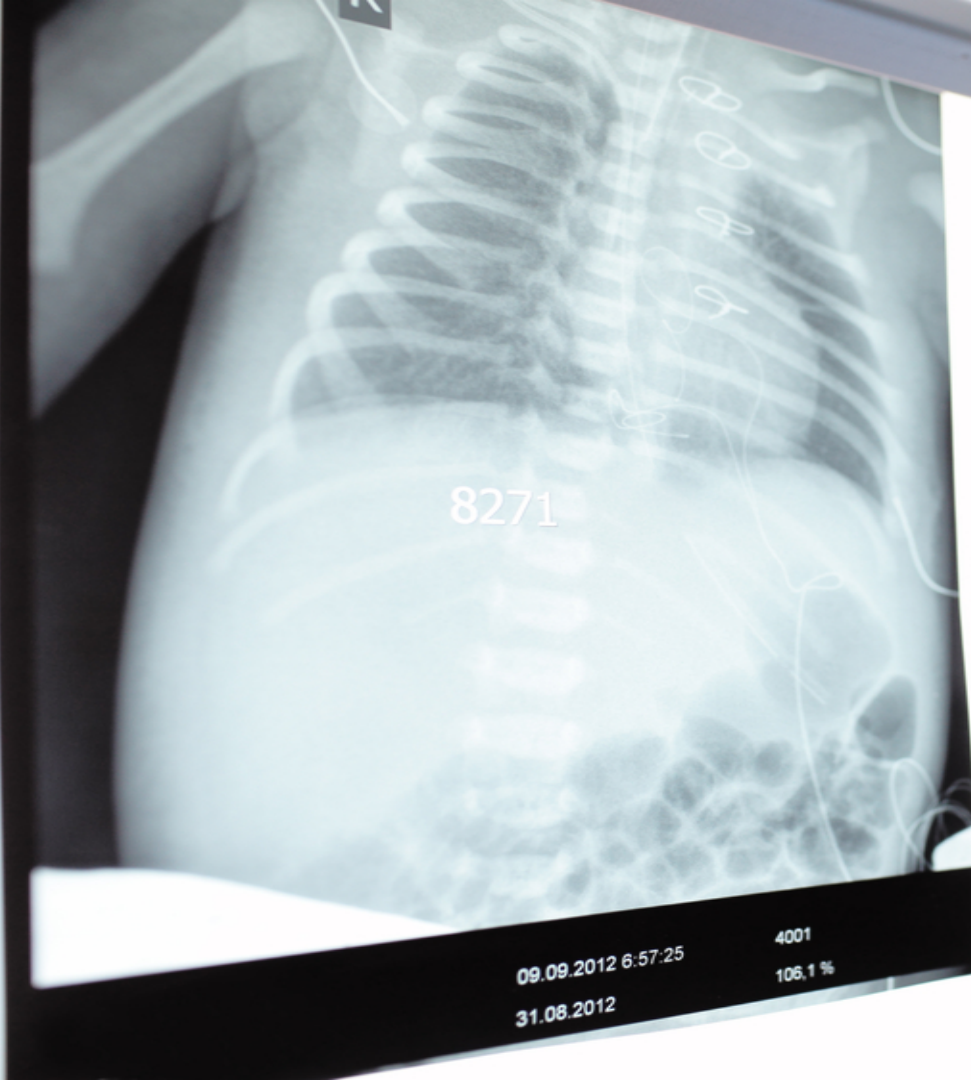
- **Scoping and correlation:**  
What are my process instances?
- **Granularity:**  
Coarse-grained or fine-grained?
- **Privacy issues**
- **Linking back to development processes**

## Big Software on the Run (BSR): Understanding software in its natural habitat



**Process mining: Understanding  
processes inside software**





**X-ray your software  
“in vivo”**

compliance

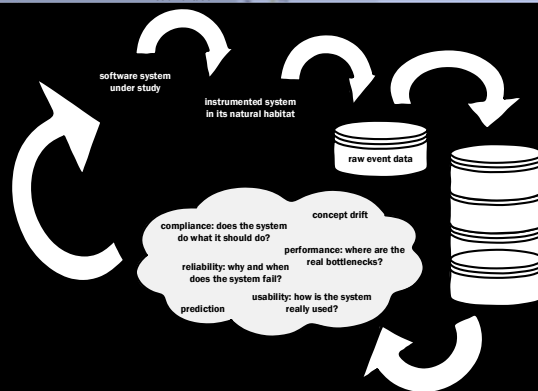
prediction

performance

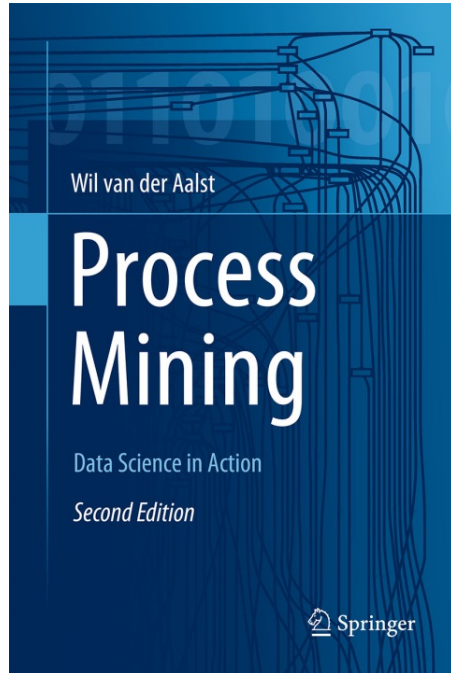
Of course also applicable to software development processes!

usability

reliability



# More information



- <http://www.processmining.org/>
- <http://www.promtools.org/>
- <http://www.promtools.org/doku.php?id=promlite11>
- <http://www.processmining.org/book/start>
- <http://www.coursera.org/learn/process-mining/>
- <http://www.futurelearn.com/courses/process-mining>
- <http://vdaalst.com/>

**W.M.P. van der Aalst. Process Mining: Data Science in Action.**  
**Springer-Verlag, Berlin, 2016.** <http://www.springer.com/978-3-662-49850-7>



## Process Mining Data Science in Action

Starts  
every 4  
weeks

Info MOOC: [www.coursera.org/course/procmin](https://www.coursera.org/course/procmin)