

Do you want to do a process mining project?

*What are the requirements?
How do I start?*

prof.dr.ir. Wil van der Aalst
RWTH Aachen University
W: vdaalst.com T: @wvdaalst

Many people and organizations contact me to apply process mining. They all have data and processes. However, often the prerequisites of process mining are unclear.

On the one hand, process mining is super generic and can be applied in any domain, just like spreadsheets are used in any organization. Spreadsheets can do anything with numbers. Process mining can do anything with events.

On the other hand, event data are not just any type of data and the notion of process is very broad.

These slides aim to clarify this. You need to check:

- 1. Do my events have a case id, activity name, and timestamp?**
- 2. Can I sketch the expected process model in terms of the activities in the event log?**

(The process will be very different, but you should have some expectations, otherwise it is pointless to talk about processes.)

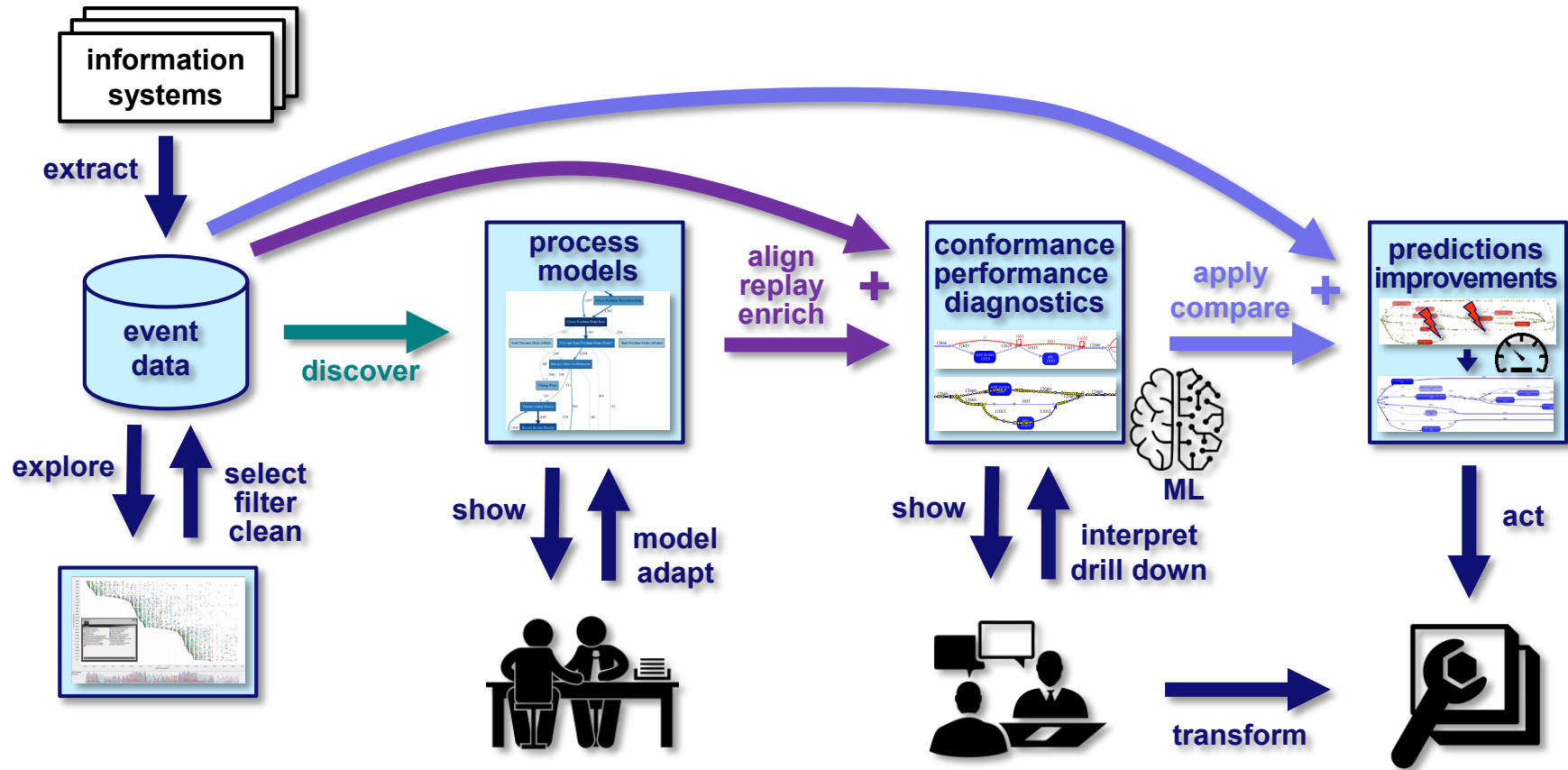
Wil van der Aalst

Did you read the previous slide?

Please do if you want to work with us. If you cannot or do not want to answer these questions, then do not ask for a meeting to collaborate!

What is it?

“event data are everywhere”



Starting point: Event data

Case ID	Activity	Resource	Timestamp	product	prod-price	quantity	address
...
6350	place order	Aiden	2018/02/13 14:29:45.000	APPLE iPhone 6 16 GB	639,00 €	5	NL-7751DG-21
6283	pay	Lily	2018/02/13 14:39:25.000	SAMSUNG Galaxy S6 32 GB	543.99	3	NL-7828AM-11a
6253	prepare delivery	Sophia	2018/02/13 15:01:33.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7887AC-13
6257	prepare delivery	Aiden	2018/02/13 15:03:43.000	SAMSUNG Galaxy S6 32 GB	543.99	1	NL-9521KJ-34
6185	confirm payment	Emily	2018/02/13 15:05:36.000	SAMSUNG Galaxy S4	329,00 €	1	NL-9521GC-32
6218	confirm payment	Emily	2018/02/13 15:08:11.000	APPLE iPhone 6s Plus 64 GB	969,00 €	2	NL-7948BX-10
6245	make delivery	Michael	2018/02/13 15:14:04.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7905AX-38
6272	pay	Emily	2018/02/13 15:20:36.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7821AC-3
6269	pay	Charlotte	2018/02/13 15:25:21.000	SAMSUNG Galaxy S4	329,00 €	1	NL-7907EJ-42
6212	prepare delivery	Sophia	2018/02/13 15:43:39.000	HUAWEI P8 Lite	234,00 €	1	NL-7905AX-38
6323	send invoice	Alexander	2018/02/13 15:46:08.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7833HT-15
6246	confirm payment	Jack	2018/02/13 15:56:03.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7833HT-15
6347	send invoice	Jack	2018/02/13 15:57:42.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7905AX-38
6351	place order	Zoe	2018/02/13 16:17:37.000	APPLE iPhone 5s 16 GB	449,00 €	3	NL-9521GC-32
6204	prepare delivery	Sophia	2018/02/13 16:31:28.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6265	confirm payment	Lily	2018/02/13 16:55:55.000	SAMSUNG Galaxy S4	329,00 €	4	NL-9521GC-32
6250	confirm payment	Jack	2018/02/13 17:03:26.000	MOTOROLA Moto G	199,00 €	4	NL-7942GT-2
6328	send invoice	Lily	2018/02/13 17:30:16.000	APPLE iPhone 6s 64 GB	858,00 €	4	NL-9514BV-16
6352	place order	Aiden	2018/02/13 17:53:22.000	APPLE iPhone 6 16 GB	639,00 €	2	NL-9514BV-16
6317	send invoice	Jack	2018/02/13 18:45:30.000	APPLE iPhone 6s 64 GB	858,00 €	5	NL-7907EJ-42
6353	place order	Sophia	2018/02/13 20:16:20.000	APPLE iPhone 5s 16 GB	449,00 €	4	NL-7751AR-19
...



71,043 events
12,666 cases
7 activities

Starting point: Event data

Case ID	Activity	Resource	Timestamp	product	prod-price	quantity	address
...
6350	place order	Aiden	2018/02/13 14:29:45.000	APPLE iPhone 6 16 GB	639,00 €	5	NL-7751DG-21
6283	pay	Lily	2018/02/13 14:39:25.000	SAMSUNG Galaxy S6 32 GB	543,99	3	NL-7828AM-11a
6253	prepare delivery	Sophia	2018/02/13 15:01:33.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7887AC-13
6257	prepare delivery	Aiden	2018/02/13 15:03:43.000	SAMSUNG Galaxy S6 32 GB	543,99	1	NL-9521KJ-34
6185	confirm payment	Emily	2018/02/13 15:05:36.000	SAMSUNG Galaxy S4	329,00 €	1	NL-9521GC-32
6218	confirm payment	Emily	2018/02/13 15:08:11.000	APPLE iPhone 6s Plus 64 GB	969,00 €	2	NL-7948BX-10
6245	make delivery	Michael	2018/02/13 15:14:04.000	APPLE iPhone 6s 16 GB	639,00 €	3	NL-7905AX-38
6272	pay	Emily	2018/02/13 15:20:36.000	APPLE iPhone 6s 16 GB	639,00 €	1	NL-7821AC-3
6269	pay	Charlotte	2018/02/13 15:25:21.000	SAMSUNG Galaxy S4	329,00 €	1	NL-7907EJ-42
6212	prepare delivery	Sophia	2018/02/13 15:43:39.000	HUAWEI P8 Lite	234,00 €	1	NL-7905AX-38
6323	send invoice	Alexander	2018/02/13 15:46:08.000	APPLE iPhone 6s 16 GB	639,00 €	1	NL-7833HT-15
6246	confirm payment	Jack	2018/02/13 15:56:03.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7833HT-15
6347	send invoice	Jack	2018/02/13 15:57:42.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7905AX-38
6351	place order	Zoe	2018/02/13 16:17:37.000	APPLE iPhone 5s 16 GB	449,00 €	3	NL-9521GC-32
6204	prepare delivery	Sophia	2018/02/13 16:31:28.000	SAMSUNG Galaxy S4	329,00 €	4	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Galaxy S4	329,00 €	2	NL-7828AM-11a
6265	confirm payment	Lily	2018/02/13 16:55:55.000	SAMSUNG Galaxy S4	329,00 €	4	NL-9521GC-32
6250	confirm payment	Jack	2018/02/13 17:03:26.000	MOTOROLA Moto G	199,00 €	4	NL-7942GT-2
6328	send invoice	Lily	2018/02/13 17:30:16.000	APPLE iPhone 6s 64 GB	858,00 €	4	NL-9514BV-16
6352	place order	Aiden	2018/02/13 17:53:22.000	APPLE iPhone 6 16 GB	639,00 €	2	NL-9514BV-16
6317	send invoice	Jack	2018/02/13 18:45:30.000	APPLE iPhone 6s 64 GB	858,00 €	5	NL-7907EJ-42
6353	place order	Sophia	2018/02/13 20:16:20.000	APPLE iPhone 5s 16 GB	449,00 €	4	NL-7751AR-19
...

event =
case +
activity +
timestamp +

Let's look at orders 6350, 6351, and 6352

Case ID	Activity	Timestamp
6350	place order	2018/02/13 14:29:45.000
6351	place order	2018/02/13 16:17:37.000
6352	place order	2018/02/13 17:53:22.000
6352	send invoice	2018/02/19 09:20:28.000
6351	send invoice	2018/02/19 16:08:07.000
6350	send invoice	2018/02/21 09:38:16.000
6350	pay	2018/03/02 12:39:37.000
6352	pay	2018/03/05 15:46:47.000
6351	cancel order	2018/03/06 10:17:01.000
6350	prepare delivery	2018/03/07 13:50:35.000
6350	make delivery	2018/03/07 16:41:01.000
6350	confirm payment	2018/03/07 16:53:00.000
6352	prepare delivery	2018/03/07 17:05:59.000
6352	confirm payment	2018/03/07 17:59:55.000
6352	make delivery	2018/03/08 09:54:36.000

Let's look at orders 6350, 6351, and 6352

Case ID	Activity	Timestamp
6350	place order	2018/02/13 14:29:45.000
6351	place order	2018/02/13 16:17:37.000
6352	place order	2018/02/13 17:53:22.000
6352	send invoice	2018/02/19 09:20:28.000
6351	send invoice	2018/02/19 16:08:07.000
6350	send invoice	2018/02/21 09:38:16.000
6350	pay	2018/03/02 12:39:37.000
6352	pay	2018/03/05 15:46:47.000
6351	cancel order	2018/03/06 10:17:01.000
6350	prepare delivery	2018/03/07 13:50:35.000
6350	make delivery	2018/03/07 16:41:01.000
6350	confirm payment	2018/03/07 16:53:00.000
6352	prepare delivery	2018/03/07 17:05:59.000
6352	confirm payment	2018/03/07 17:59:55.000
6352	make delivery	2018/03/08 09:54:36.000

order 6350

place order

send invoice

pay

prepare delivery
make delivery
confirm payment

order 6351

place order

send invoice

cancel order

order 6352

place order
send invoice

pay

prepare delivery
confirm payment
make delivery



Using the whole event log

Case ID	Activity	Timestamp
6350	place order	2018/02/13 14:29:45.000
6351	place order	2018/02/13 16:17:37.000
6352	place order	2018/02/13 17:53:22.000
6352	send invoice	2018/02/19 09:20:28.000
6351	send invoice	2018/02/19 16:08:07.000
6350	send invoice	2018/02/21 09:38:16.000
6350	pay	2018/03/02 12:39:37.000
6352	pay	2018/03/05 15:46:47.000
6351	cancel order	2018/03/06 10:17:01.000
6350	prepare delivery	2018/03/07 13:50:35.000
6350	make delivery	2018/03/07 16:41:01.000
6350	confirm payment	2018/03/07 16:53:00.000
6352	prepare delivery	2018/03/07 17:05:59.000
6352	confirm payment	2018/03/07 17:59:55.000
6352	make delivery	2018/03/08 09:54:36.000

place order
send invoice
pay
prepare delivery
make delivery
confirm payment

8016 x

place order
send invoice
cancel order

1651 x

place order
send invoice
pay
prepare delivery
confirm payment
make delivery

2962 x

place order
pay
send invoice
prepare delivery
make delivery
confirm payment

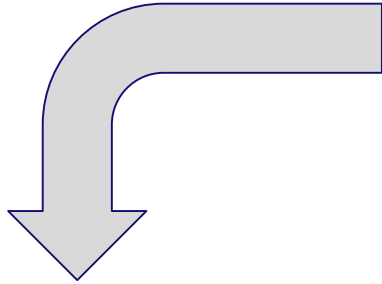
30 x

place order
pay
send invoice
prepare delivery
confirm payment
make delivery

7 x



Using the whole event log



place order
send invoice
pay
prepare delivery
make delivery
confirm payment

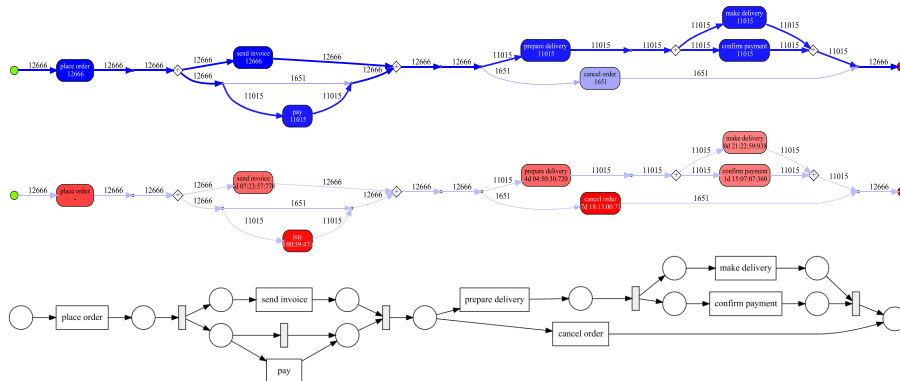
8016 x

place order
send invoice
cancel order

1651 x

place order
send invoice
pay
prepare delivery
confirm payment
make delivery

2962 x



place order
pay
send invoice
prepare delivery
make delivery
confirm payment

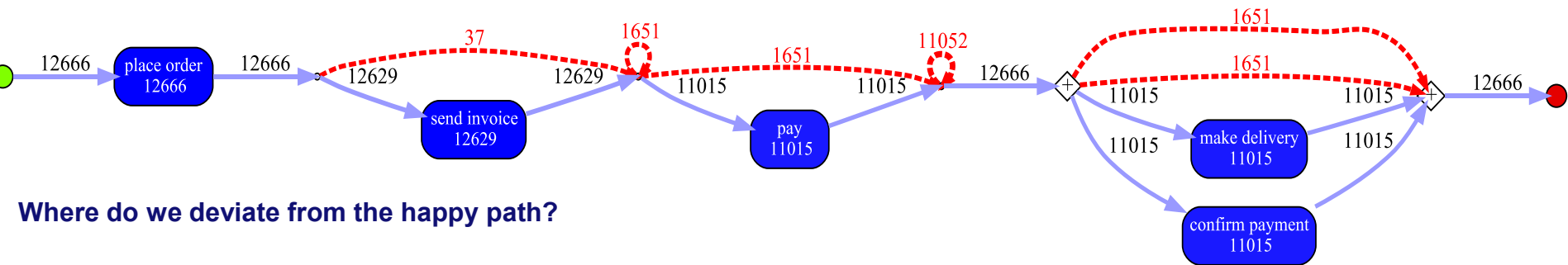
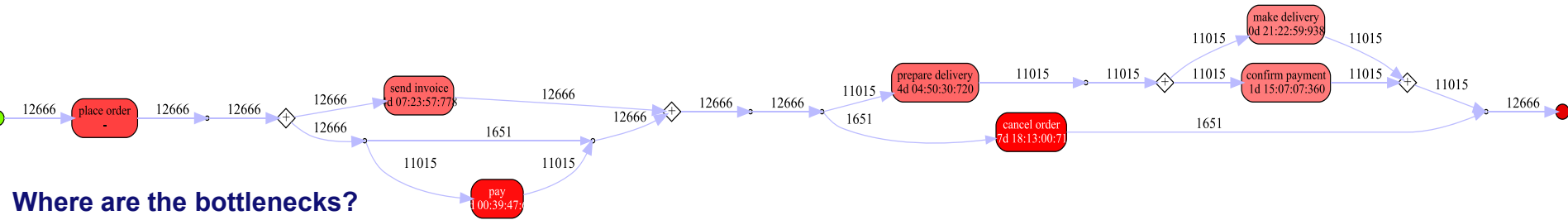
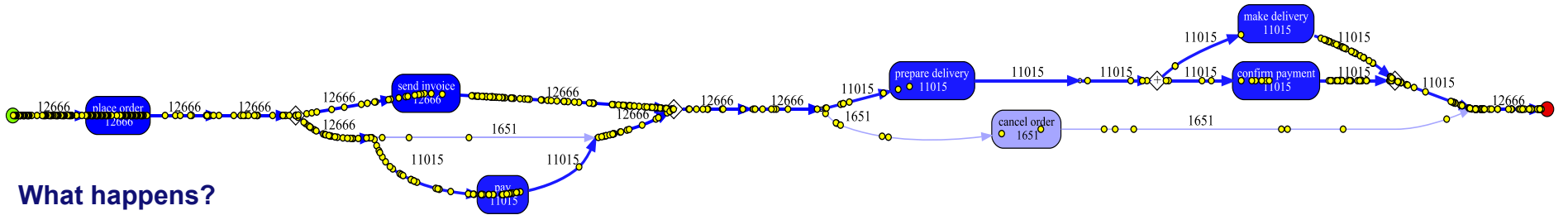
30 x

place order
pay
send invoice
prepare delivery
confirm payment
make delivery

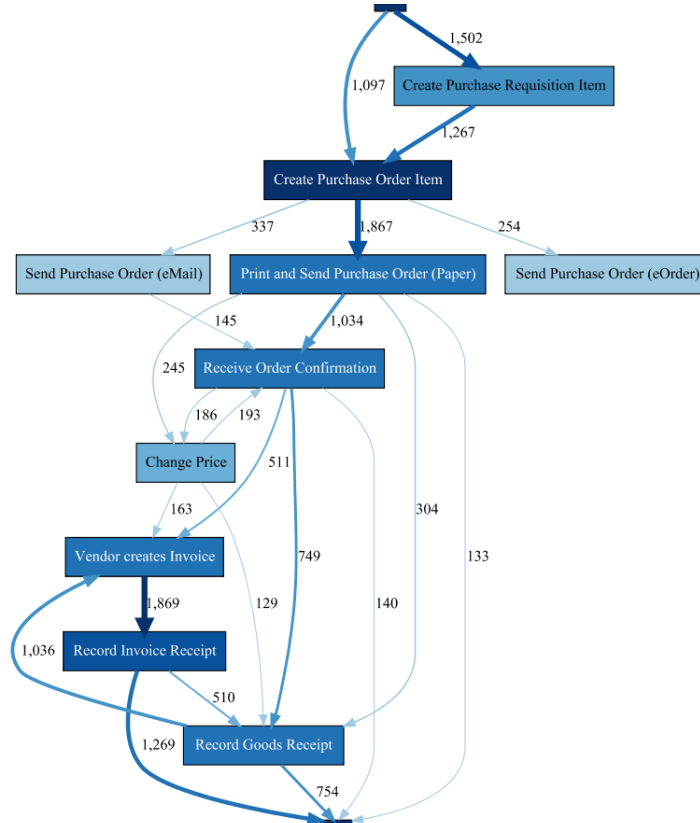
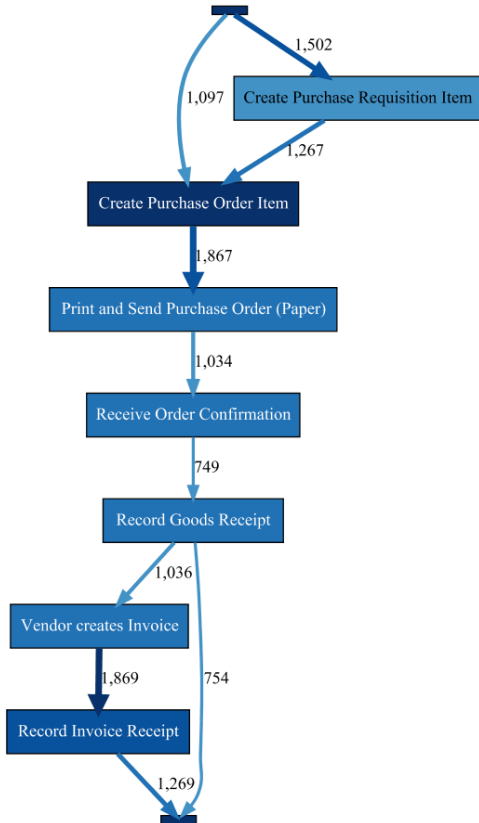
7 x



Performance and Compliance



Reality is not so simple



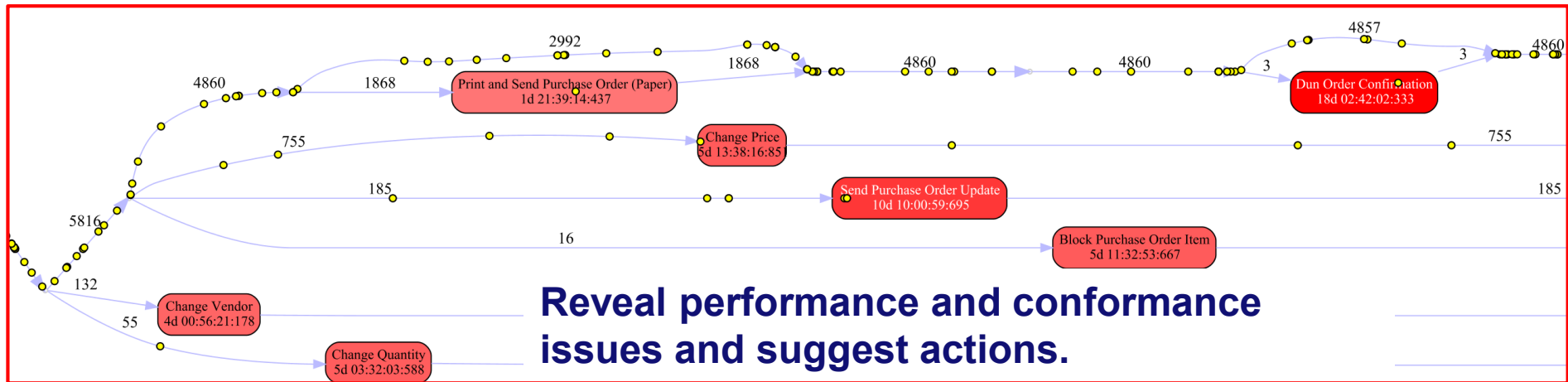
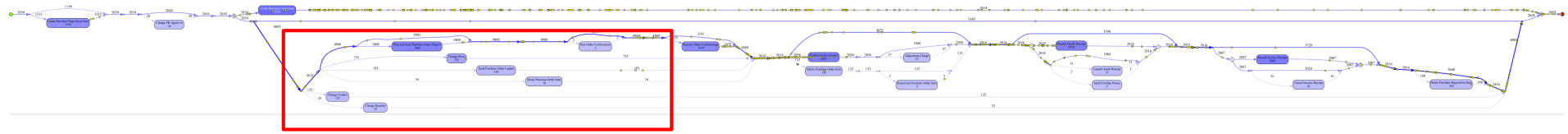
Reality is not so simple



It is common to find thousands of different variants for simple core processes like P2P and O2C!

Caused by hand-offs, rework, duplication, ineffective communication, etc.

Process mining helps organizations to address these problems (to actually realize the economies of scale promised)



More on event data

“are your data really event 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 values (CSV) file or spreadsheet,
- a transaction log (e.g., a trading system),
- a business suite/ERP system (SAP, Oracle, etc.),
- a message log (e.g., from IBM middleware),
- an open API providing data from websites or social media, ...

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

Minimal requirements in terms of a CSV file

- Each row corresponds to an event.
- There are at least three columns:
 - Case id (patient id, order number, claim number, ...)
 - Activity name (approve, reject, request, send, ...)
 - Timestamp (2015-08-18T06:36:40, ...)
- There may be many other (optional) columns: resource, transaction type, age, costs, etc.

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

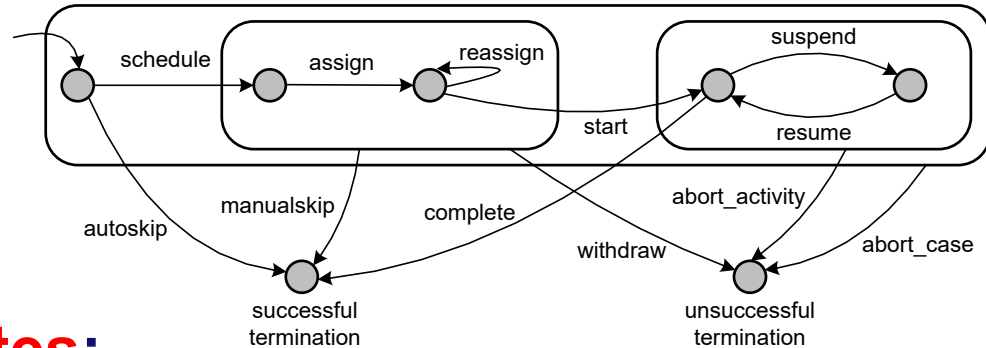
Extensions

- **Transactional information** on activity instances:

An event can represent a **start, complete, suspend, resume, abort, etc.**

- **Case** versus **event attributes**:

- case attributes do not change, e.g., the birth date or gender of a patient,
- event attributes are related to a particular step in the process.



XES (eXtensible Event Stream)

- Adopted by the **IEEE Task Force on Process Mining**.
- The format is supported by tools such as **ProM** and **Disco** (used in this course).
- Predecessors: MXML and SA-MXML.
- Conversion from other formats (CSV) is easy if the right data are available.
- **XML syntax** and **OpenXES library** available.
- See www.xes-standard.org.



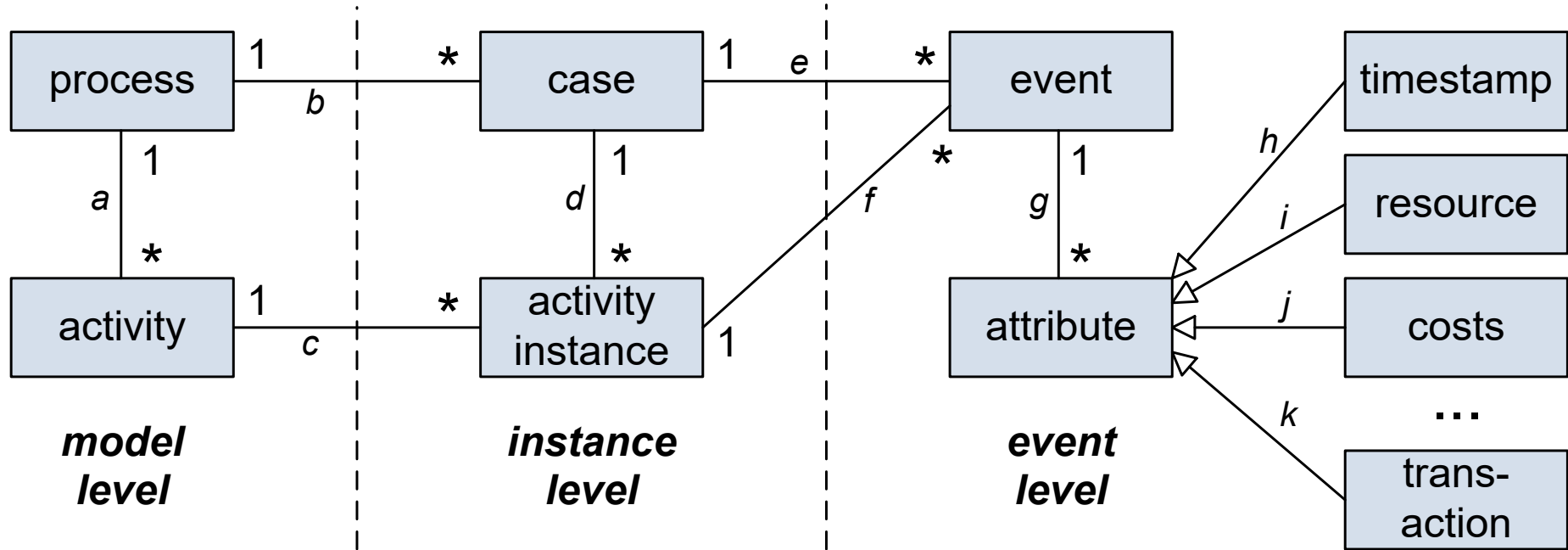
Extensible Event Stream

Simplistic view on event data

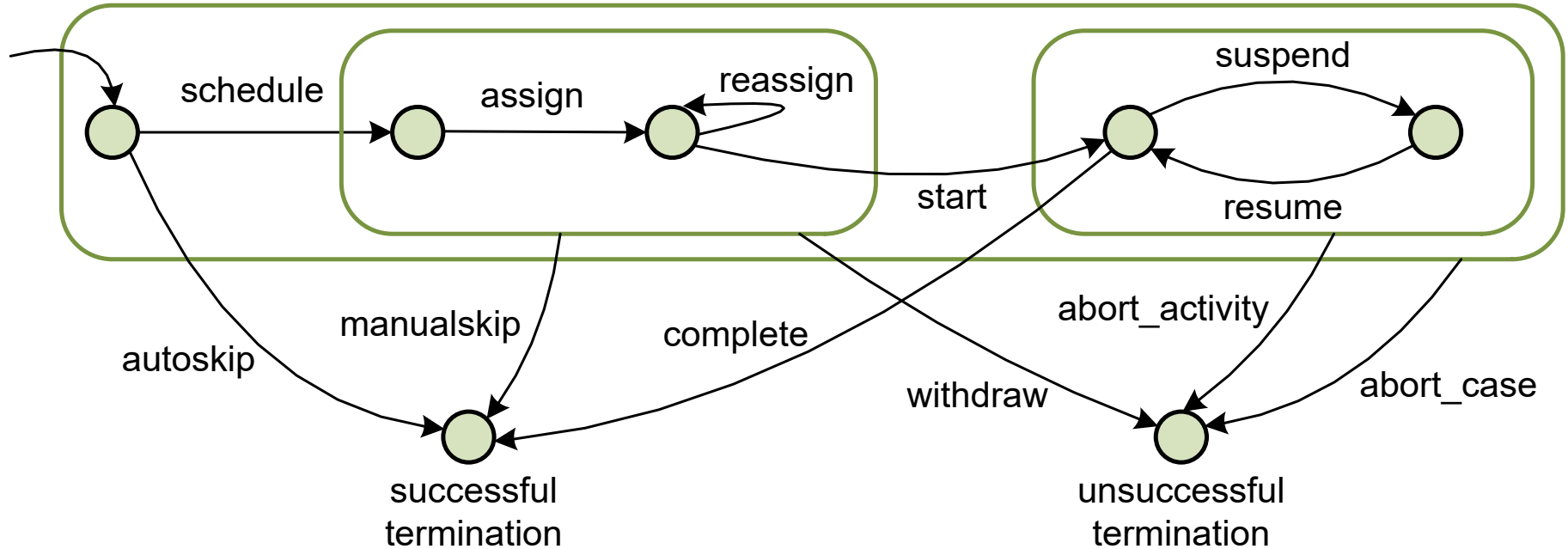
order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	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**

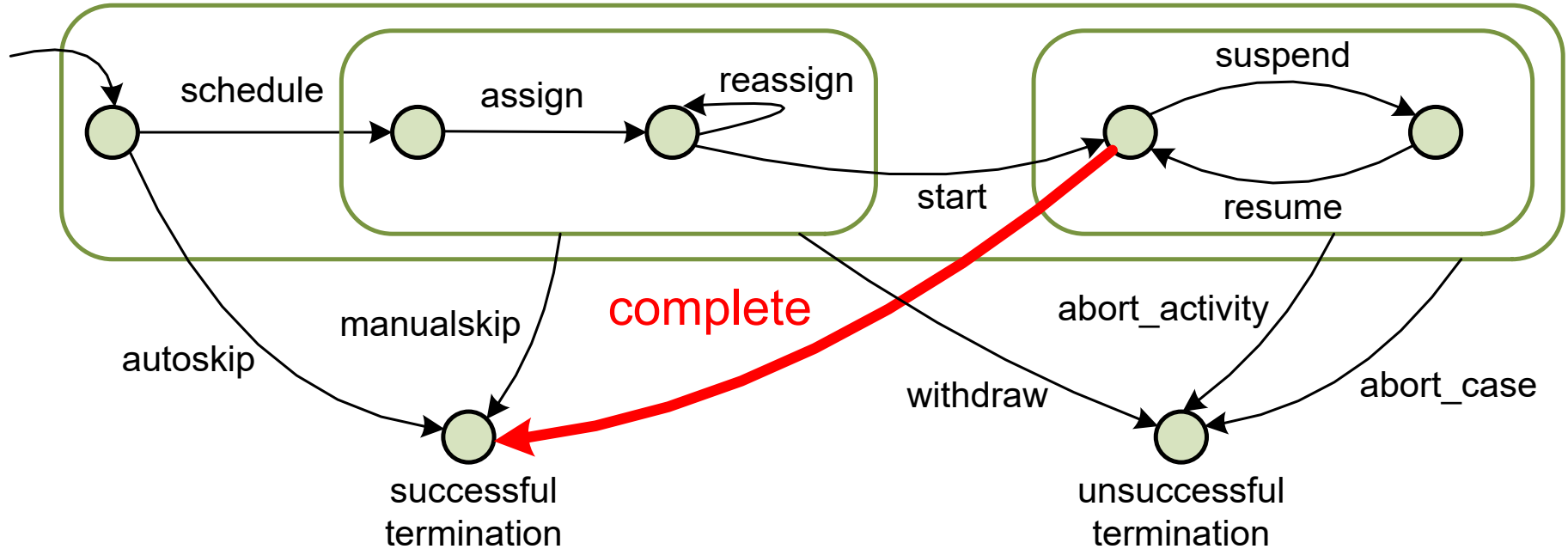
A more refined view



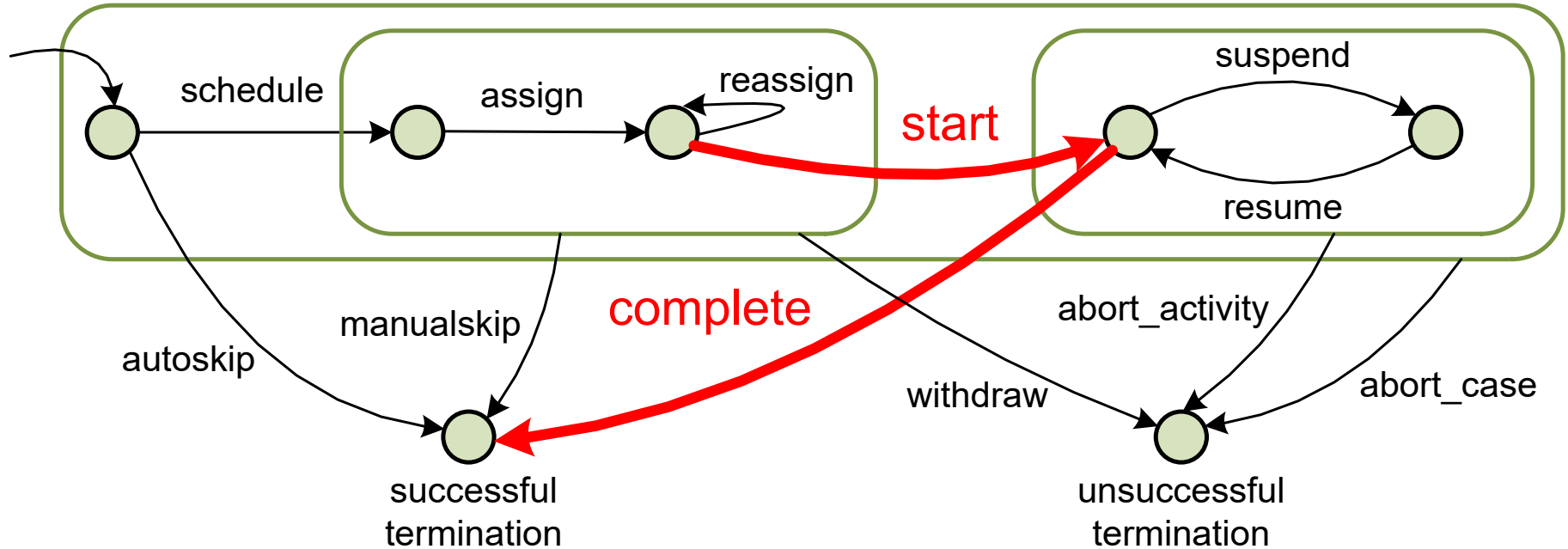
Transactional model for activities



Atomic activities



Activities that have a duration



- Disco - New project

Activity

Activity

Case ID	Activity	Resource	Start Timestamp	Complete Timestamp	Case ID	Result by Resource A	Result by Resource B	Result by Resource X	Result by Resource Y	Activity Result	Activity Result
1	initia review	Mia	20060105 00:00:00	20060105 00:00:00	Simul					complete	
1	case	Carli	20060105 00:00:00	20060105 00:00:00	Simul					complete	
1	get review 1	Pam	20060110 00:00:00	20060110 00:00:00	Simul		reject		reject	complete	
1	case	John	20060111 00:00:00	20060111 00:00:00	Simul					complete	
1	collect reviews	Aime	20060113 00:00:00	20060113 00:00:00	Simul					complete	
1	initia review	Will	20060113 00:00:00	20060113 00:00:00	Simul					complete	
1	initia additional review	Pam	20060113 00:00:00	20060113 00:00:00	Simul					complete	3
1	initia review	Will	20060123 00:00:00	20060123 00:00:00	Simul					accept	1
1	decide	Mia	20060124 00:00:00	20060127 00:00:00	Simul					complete	3
1	initia additional review	Will	20060127 00:00:00	20060127 00:00:00	Simul					complete	1
1	get review 5	Mary	20060131 00:00:00	20060131 00:00:00	Simul					reject	1
1	initia additional review	Will	20060131 00:00:00	20060131 00:00:00	Simul					complete	4
1	decide	Mia	20060128 00:00:00	20060222 00:00:00	Simul					complete	1
1	initia additional review	Sara	20060222 00:00:00	20060222 00:00:00	Simul					complete	4
1	decide	Will	20060408 01:00:00	20060410 01:00:00	Simul					reject	1
1	decide	Will	20060408 01:00:00	20060410 01:00:00	Simul					complete	5
1	initia review	Aime	20060411 01:00:00	20060411 01:00:00	Simul					complete	
1	initia review	Aime	20060303 00:00:00	20060306 00:00:00	Simul					complete	
1	get review 2	Sara	20060306 00:00:00	20060306 00:00:00	Simul					accept	accept
1	case	John	20060306 00:00:00	20060306 00:00:00	Simul					complete	
1	initia review	Will	20060306 00:00:00	20060306 00:00:00	Simul					complete	
1	collect reviews	Will	20060307 00:00:00	20060307 00:00:00	Simul					complete	
1	decide	Mia	20060301 00:00:00	20060301 00:00:00	Simul					complete	2
1	initia additional review	Will	20060304 00:00:00	20060304 00:00:00	Simul					complete	0
1	decide	Will	20060312 00:00:00	20060312 00:00:00	Simul					complete	0
1	decide	Mia	20060514 01:00:00	20060517 01:00:00	Simul					complete	2
1	initia additional review	Will	20060520 01:00:00	20060521 01:00:00	Simul					complete	2
1	initia review	Will	20060521 01:00:00	20060521 01:00:00	Simul					complete	2
1	decide	Will	20060727 01:00:00	20060730 01:00:00	Simul					complete	0
1	initia additional review	Will	20060808 01:00:00	20060812 01:00:00	Simul					complete	2
1	decide	Pam	20060814 01:00:00	20060814 01:00:00	Simul					reject	2
1	decide	Will	20071115 00:00:00	20071122 00:00:00	Simul					complete	1
1	initia additional review	Aime	20070205 00:00:00	20070206 00:00:00	Simul					complete	1
1	initia review	Will	20070205 00:00:00	20070205 00:00:00	Simul					complete	1
1	decide	Mia	20071202 00:00:00	20071202 00:00:00	Simul					complete	2
1	initia additional review	Will	20071202 00:00:00	20071202 00:00:00	Simul					complete	1
1	decide	Carli	20071205 00:00:00	20071205 00:00:00	Simul					accept	3
1	initia review	Will									

**Import CSV file, then apply
“Convert CSV to XES” plug-in.**

Now you should be able to check:

1. Do my events have indeed

- a) a case id,**
- b) activity name, and**
- c) timestamp?**

2. Can I sketch the expected process model in terms of the activities in the event log?

(The process will be very different, but you should have some expectations, otherwise it is pointless to talk about processes.)

Please bring fragments of event data and rough sketches of process models to the meeting. This will help to quickly see whether process mining will be feasible and beneficial.

The later slides provide some additional context.

Wil van der Aalst

Do you want to discuss a joint project or seek advice?

Please ensure that you followed the suggestions on the previous slide. If you cannot or do not want to do this, then please do not contact us for free advice or support. This may seem rude, but we are spending a lot of time interacting with people that have lots of “data” and “processes” (everyone has), but do not have a clue what process mining is (despite the abundance of information).

A Bit of History

“bridging the gap between process science and data science”

“process management by modeling”

Petri nets

Formal methods

Concurrency theory

BPM, WFM, etc.

Simulation

< 1999

≥ 1999

Process mining

Predictive analytics

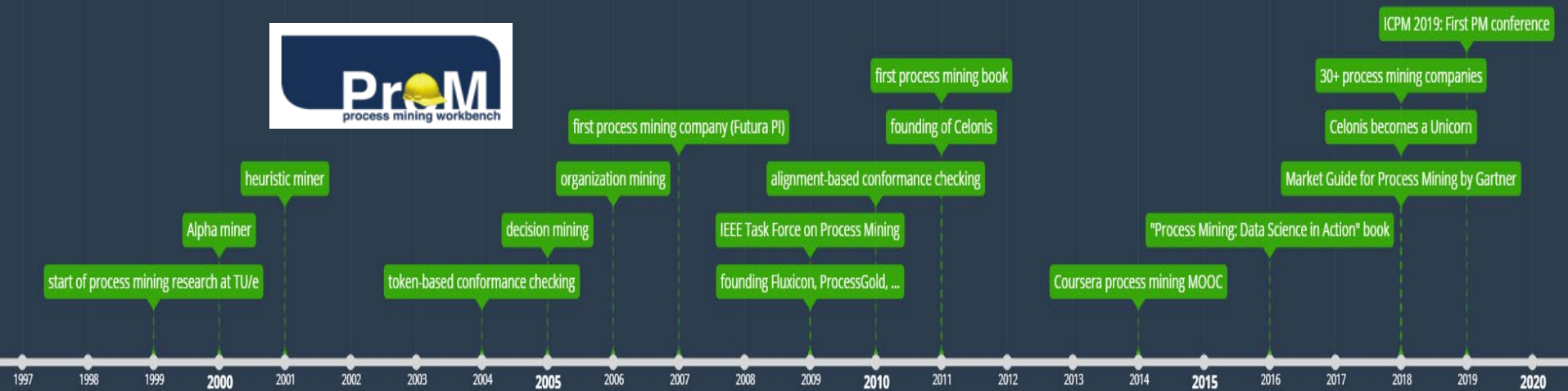
Process discovery

Conformance checking

“process management by mining”

research

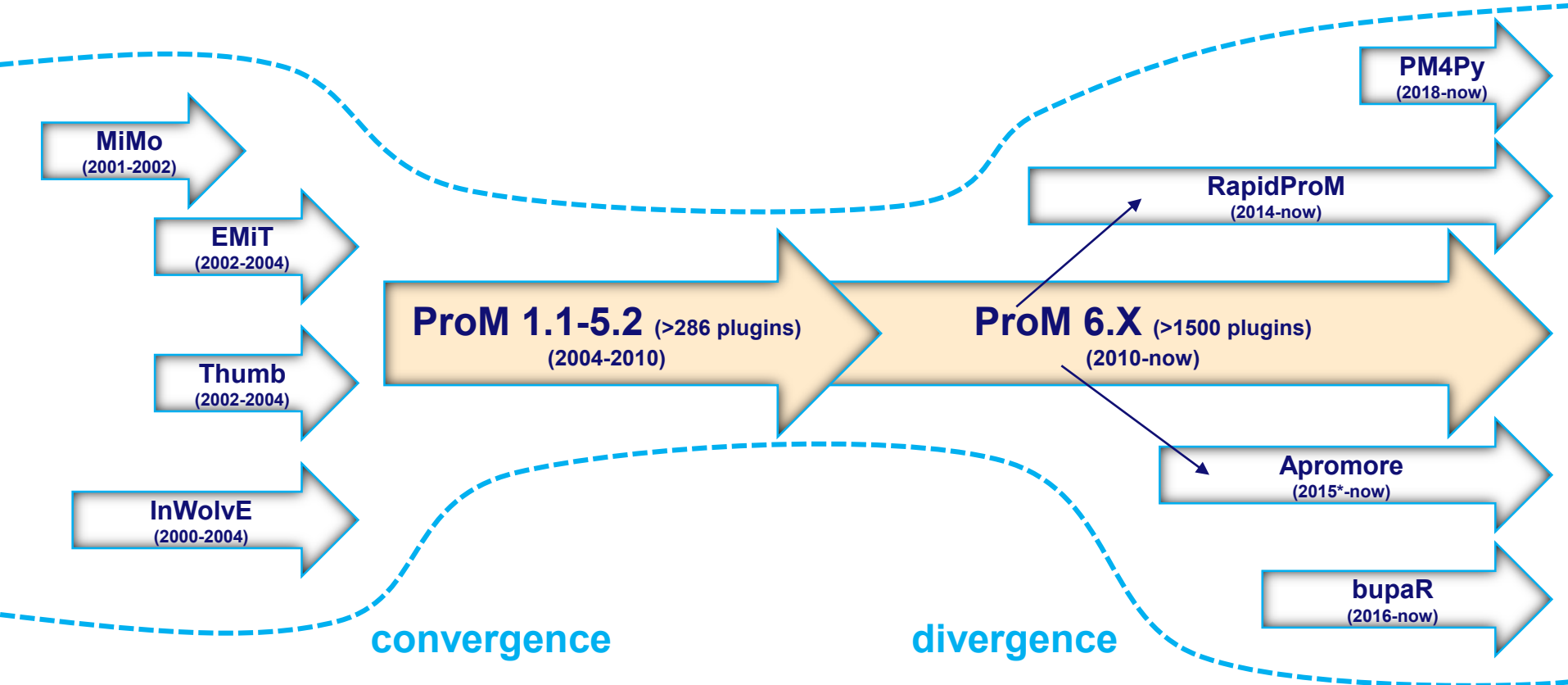
commercial tools



adoption

Open-Source Process Mining Software

(far from complete)



Many process mining tools are available

Vendor	Tool	Website	Acad. ver.
Abbyy	ABBYY Timeline	www.abbyy.com	No
Appian (Lana Labs)	LANA Process Mining	lanalabs.com	No
Apromore	Apromore Enterprise Edition	apromore.org	Yes
bupaR	bupaR	bupar.net	Yes
businessOptix	businessOptix	businessoptix.com	Yes
Celonis	Celonis EMS	celonis.com	Yes
Datricks	Datricks	datricks.com	Yes
DCR	DCR Portal	www.dcrsolutions.net	Yes
Deloitte	Process X-ray	processxray.deloitte.com	No
EverFlow	EverFlow	everflow.ai	No
Fluxicon	Disco	fluxicon.com	Yes
FortressIQ	FortressIQ	fortressiq.com	No
Fraunhofer FIT	PM4Py	pm4py.fit.fraunhofer.de	Yes
Hyland	Onbase	www.hyland.com	No
IBM (myInvenio)	myInvenio	my-invenio.com	No
Integr8	Explora Process	integr8.it	No
Kofax	Kofax Insight	www.kofax.com	No
livejourney	livejourney	www.livejourney.com	No
Logpickr	Logpickr Process Explorer 360	www.logpickr.com	No
Mavim	Mavim	www.mavim.co	No
Mehrwerk GmbH	MPM	mpm-processmining.com	No
Mindzie	mindzie	mindzie.com	Yes
Minit (Microsoft)	Minit	www.minit.io	Yes
Nintex UK Ltd	Nintex	www.nintex.com	No
Oniq	IQ/A	www.oniq.com	No
PAFnow (Celonis)	PAFnow	pafnow.com	No
Process.science	process.science	www.process.science	No
ProcessDiamond	ProcessDiamond	processdiamond.com	Yes
ProcessM	PmBI	processm.com	Yes
Puzzle Data	ProDiscovery	www.puzzledata.com	No
QPR Software	QPR ProcessAnalyzer	www.qpr.com	No
SAP (Signavio)	SAP Signavio	www.signavio.com	Yes
Skan AI	Skan	www.skan.ai	No
Software AG	Aris	aris-process-mining.com	Yes
Soroco	Scout Platform	soroco.com	No
StereoLogic	StereoLogic Process Mining	www.stereologic.com	No
TU/e	Prom	www.promtools.org	Yes
TU/e	RapidProm	www.rapidprom.org	Yes
UI Path	UI Path Process Mining	www.uipath.com	Yes
UltimateSuite	UltimateSuite TM/RPA	www.ultimatesuite.com	No
Upflux	Upflux	upflux.net	No
Worksoft	Worksoft	www.worksoft.com	No

ProcessMining

Introduction Overview Publications Courses Event Data **Software** Links

Process Mining

The bridge between process science and data science.

LEARN MORE

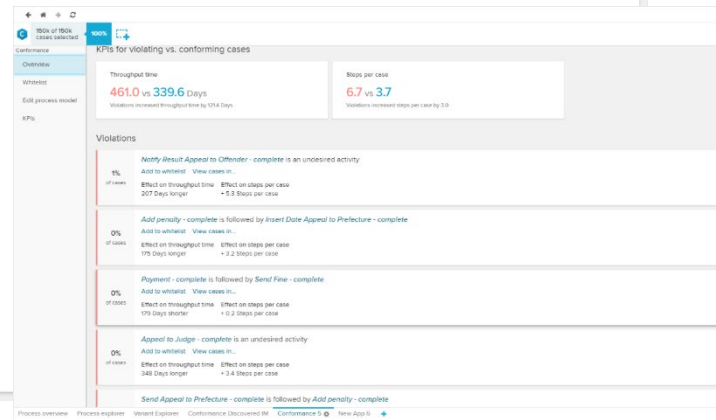
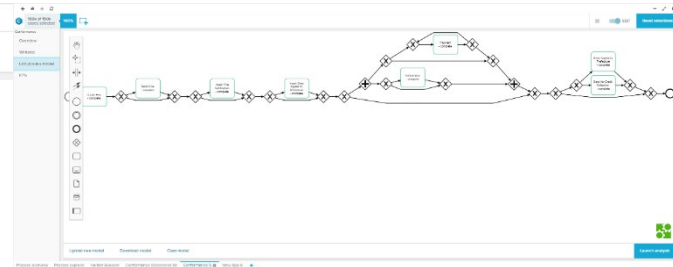
40+ tools

Software section highlights:

- ABBYY**: ABBYY Timeline by Abbyy. Academic license available. [Explore more](#)
- apromore**: Apromore by Apromore Pty Ltd. Academic license available. [Explore more](#)
- ARIS**: ARIS Process Mining by Software AG. Academic license available. [Explore more](#)
- Behfotab**: Behfotab by Behfotab. Academic license available. [Explore more](#)
- bupaR**: bupaR by the Business Informatics research group - RWTH Aachen University. Academic license available. [Explore more](#)
- businessOptix**: BusinessOptix by BusinessOptix. Academic license available. [Explore more](#)
- celonis**: Celonis by Celonis SE. Academic license available. [Explore more](#)
- datricks**: Datricks by Datricks. Academic license available. [Explore more](#)
- DCR**: DCR Portal by DCR Solutions. Academic license available. [Explore more](#)
- Disco**: Disco by Fluxicon. Academic license available. [Explore more](#)
- eBrowsers**: eBrowsers by Syntex Solutions. Academic license available. [Explore more](#)
- EVERFLOW**: EverFlow by Everflow. Academic license available. [Explore more](#)

www.processmining.org

Celonis was the first to focus on continuous process mining



From data scientists to process managers and from insights to actions.

Example: some of Celonis's customers

Technology



Financial Services & Insurance



Life Sciences & Chemicals



Consumer & Retail



Manufacturing



Telecommunications & Media



Energy & Utilities



Oil & Gas



Thousands of large organizations are using Celonis (approx. 50% of Fortune 500) and in some of these there are thousands of active users (e.g., Siemens, BMW, etc.)

**Much more than
process discovery
and conformance
checking.**

**Always related to
performance and
compliance.**

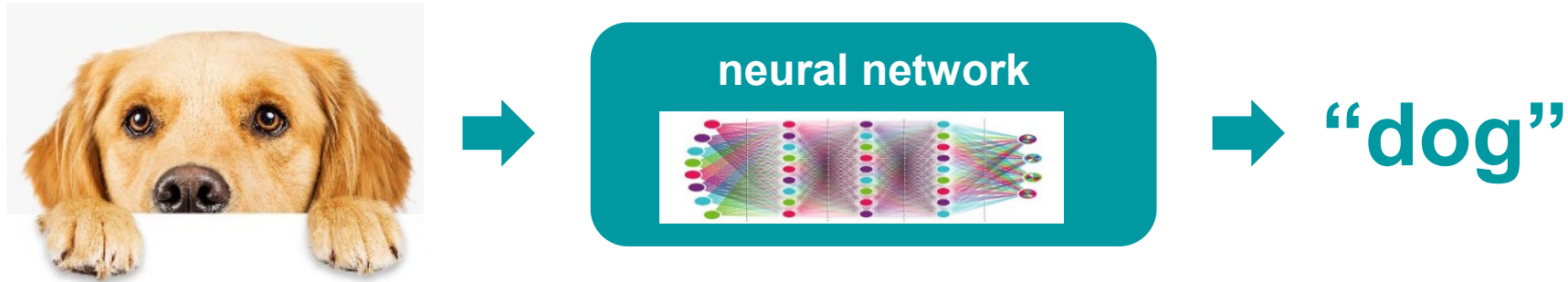


Relation to ML & AI

“Siri and Alexa cannot mine your processes”

Process mining is very different!

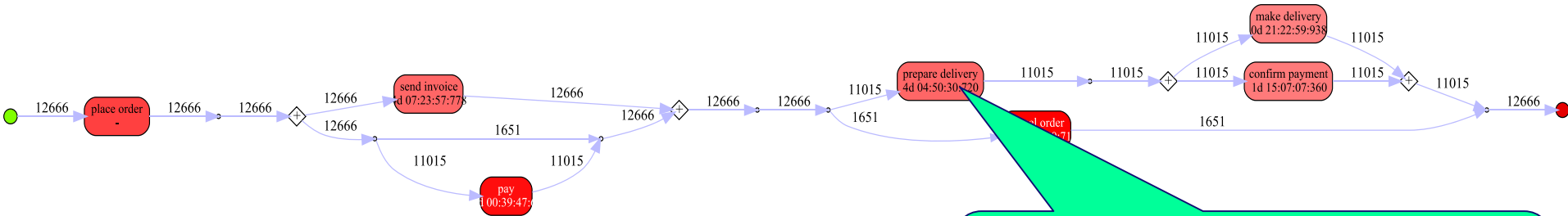
The core process mining techniques and tools do not use techniques from machine learning, artificial intelligence, data mining, etc.



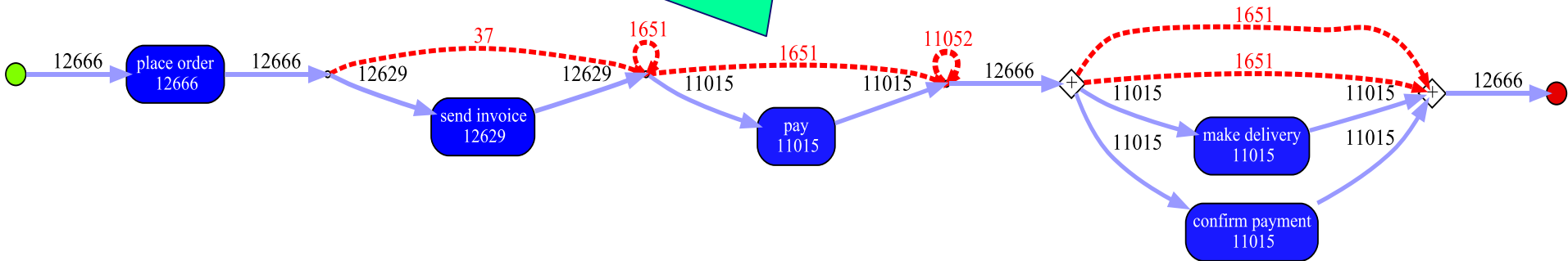
- The model needs to be visible and understandable by stakeholders.
- Process owners are not going to label training examples.

However, ...

PM can be used to generate ML problems



Why are payments skipped?
What do these cases have in common?
Can we predict such deviations?



Relation to Robotic Process Automation (RPA)

“enabling the poor man’s workflow management solution”

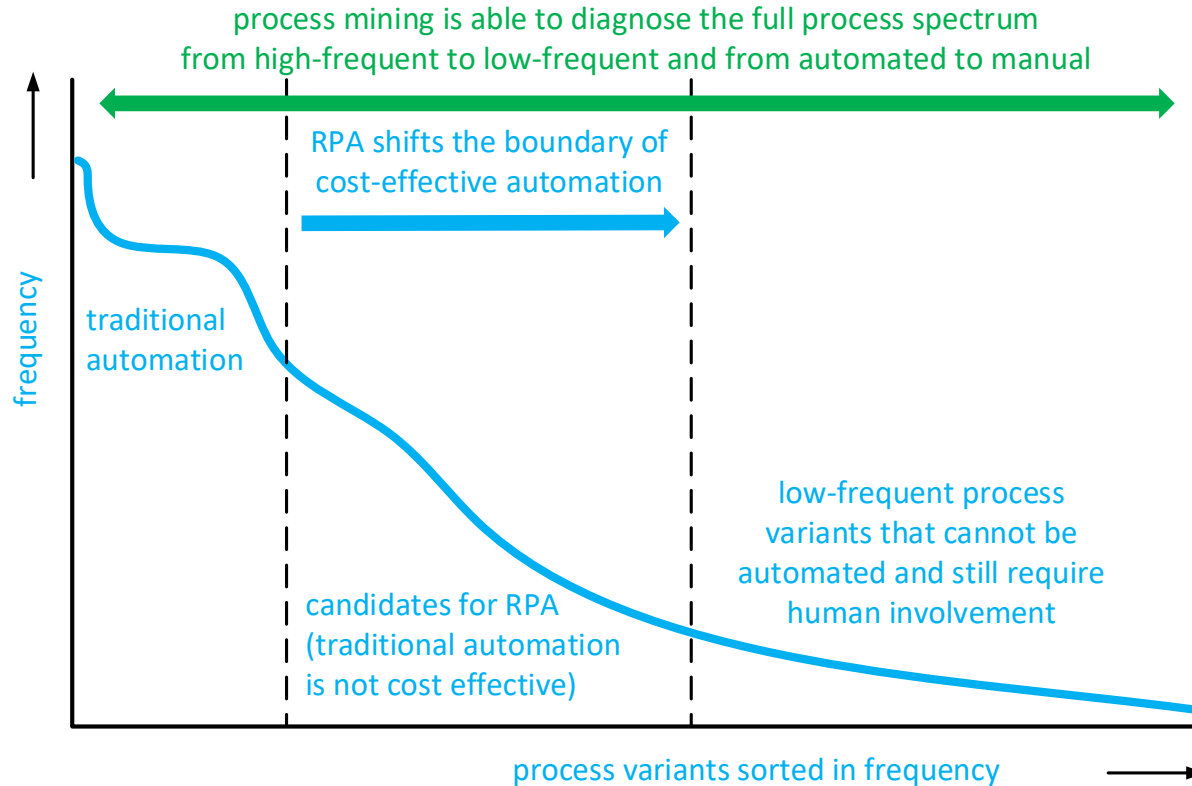


How to pick your
automation battles?



How to pick your automation battles?

The RPA connection



Object-Centric Event Logs

“Everything should be made as simple as possible - but no simpler!”

Let's make the following assumption

activity	time	orders	items	packages	customers	products	price	weight
pick item	2019-12-26 12:04:46	9911224}	{884803}	{}	{Wil van der Aalst}	{iPhone 8}	529.0	0.21
reorder item	2019-12-26 12:37:26	991271}	{885002}	{}	{Mohammadreza Fani Sani}	{Kindle Paperwhite}	129.0	0.495
place order	2019-12-26 12:44:23	991283}	{885038,885039}	{}	{Luis Santos}	{MacBook Air,iPad Pro,iPhone 11 Pro,iPhone 11 Pro Max}	2700.0	1.733
pick item	2019-12-26 14:01:16	991266}	{884983}	{}	{Marco Pegoraro}	{MacBook Air}	2200.0	1.25
create package	2019-12-26 14:01:16	991265}	{884975,884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}	{Fire Stick 4K,iPad Pro,iPad Pro,iPad Pro,Fire Stick,Kindle}	3506.97	2.412
send package	2019-12-26 14:16:11	991265}	{884975,884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}	{Fire Stick 4K,iPad Pro,iPad Pro,iPad Pro,Fire Stick,Kindle}	3506.97	2.412
pick item	2019-12-26 14:16:48	991279}	{885007}	{}	{Chunhua Wang}	{Kindle Paperwhite}	799.0	0.166
confirm order	2019-12-26 14:26:01	991283}	{885038,885039}	{}	{Luis Santos}	{MacBook Air,iPad}	2700.0	1.733
reorder item	2019-12-26 15:32:43	991251}	{884912}	{}	{Tobias Brockhoff}	{Fire Stick}	39.99	0.2
confirm order	2019-12-26 15:32:44	991282}	{885036,885037}	{}	{Lisa Mannel}	{Echo,Echo Dot}	134.98	1.16
pick item	2019-12-26 15:33:28	991278}	{885024}	{}	{Junxiong Gao}	{MacBook Pro}	2500.1	1.7
place order	2019-12-26 15:48:33	991284}	{885040,885038,885042,885043,885044}	{}	{Christine Dobbert}	{iPhone X,Fire Stick,MacBook Air,Echo Show 5,iPhone 11 Pro}	4222.98	2.79
failed delivery	2019-12-26 15:04:53	991240,991161}	{884913,884913,884913,884876,884938,884914,884939}	{660790}	{Tobias Brockhoff}	{iPad Air,Echo Studio,Echo Studio,Kindle,Kindle,Echo,iPhone mini,iPad Pro,iPad Pro}	5982.95	3.42
pick item	2019-12-26 15:12:00:05	991278}	{885005}	{}	{Julia Thom Gao}	{iPhone X}	699.0	0.72
confirm order	2019-12-26 15:25:00	991258}	{884938,884939,884940,884941,884942,884943}	{}	{Tobias Brockhoff}	{Echo,Fire Stick,iPad mini,iPad Pro,iPad Pro,iPad Air}	3267.99	2.666
send package	2019-12-26 15:26:49	991247,99125}	{884902,884922,884923,885004,885005,884901}	{660798}	{Mohammadreza Fani Sani}	{MacBook Air,iPad mini,iPad Pro,iPhone 11 Pro,iPad Pro,MacBook Pro}	8496.0	4.054
failed delivery	2019-12-26 15:36:16	991265}	{884975,884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}	{Fire Stick 4K,iPad Pro,iPad Pro,iPad Pro,Fire Stick,Kindle}	3506.97	2.412
confirm order	2019-12-26 15:40:51	991274}	{885008,885009,885010,885011}	{}	{Junxiong Gao}	{Kindle,iPhone X,Fire Stick,iPhone 8}	1352.98	1.065
failed delivery	2019-12-26 15:46:21	991128,99125}	{884424,884933,884933,884936,885005,885011,884903}	{660797}	{Junxiong Gao}	{Echo Show 8,Kindle Paperwhite,iPhone mini,Kindle,iPhone X,iPhone 8,Echo Show 5}	2145.99	1.7
payment reminder	2019-12-26 15:54:44	991169}	{884965,884966,884965,884966}	{}	{Gyungum Park}	{iPhone 8,Echo Plus,iPad Air,iPad mini}	1608.99	0.21
pick item	2019-12-26 15:55:38	991201}	{884917}	{}	{Eckhard Sa	{Echo Show 8}	129.99	0.18
pick item	2019-12-26 16:00:38	991251}	{884912}	{}	{Tobias Brockhoff}	{Fire Stick}	39.99	0.2
reorder item	2019-12-26 16:04:42	991265}	{884977}	{}	{Seran Uysal}	{Fire Stick 4K}	89.99	0.28
payment reminder	2019-12-26 16:11:39	991164}	{884542,884543,884544,884545,884546,884547}	{}	{Junxiong Gao}	{Kindle Paperwhite,iPad Air,iPhone 11,MacBook Air,iPad mini,Echo Dot}	4087.99	3.011
pick item	2019-12-26 16:22:04	991241}	{884882}	{}	{Lisa Mannel}	{iPhone 8}	529.0	0.21
create package	2019-12-26 16:22:04	991263,99126}	{884967,884964,884966}	{660799}	{Luis Santos}	{iPad Air,iPhone 8,iPad}	1500.0	1.133

event = activity + timestamp + objects + attributes



OCEL STANDARD

[Introduction](#)[Format](#)[Event Logs](#)[Tool Support](#)[Contact](#)[Download Document](#)

The purpose of the **OCEL** standard is to provide a general standard to interchange object-centric event data with multiple case notions. We set the following goals for the standard:

- **Interoperability:** with the provision of the OCEL standard and JSON/XML serializations of OCEL, we want to support a widespread collection of languages and systems.
- **Generalization:** the standard supports the storage of events, objects, and their attributes. Furthermore, the standard can be extended.
- **Provision of a collection of examples:** example logs, extracted from information systems supporting some widespread business processes, are provided for the OCEL standard.
- **Tool/Library Support:** to support the implementation of OCEL in custom applications, tool/library support shall be provided.



```
<events>
  <event>
    <string key="id" value="e1"/>
    <string key="activity" value="place_order"/>
    <date key="timestamp" value="2020-07-09T08:20:01.527+01:00"/>
    <list key="omap">
      <string key="object-id" value="i1"/>
      <string key="object-id" value="o1"/>
      <string key="object-id" value="i2"/>
    </list>
    <list key="vmap">
      <string key="resource" value="Alessandro"/>
      <float key="prepaid-amount" value="200.0"/>
    </list>
  </event>
  <event>
    <string key="id" value="e2"/>
    <string key="activity" value="check_availability"/>
    <date key="timestamp" value="2020-07-09T08:21:01.527+01:00"/>
  </event>
</events>
```

<http://ocel-standard.org/>



Chair of Process
and Data Science

RWTH AACHEN
UNIVERSITY

PROCESS AND DATA SCIENCE GROUP
RWTH AACHEN UNIVERSITY

OCEL Standard

Authors:

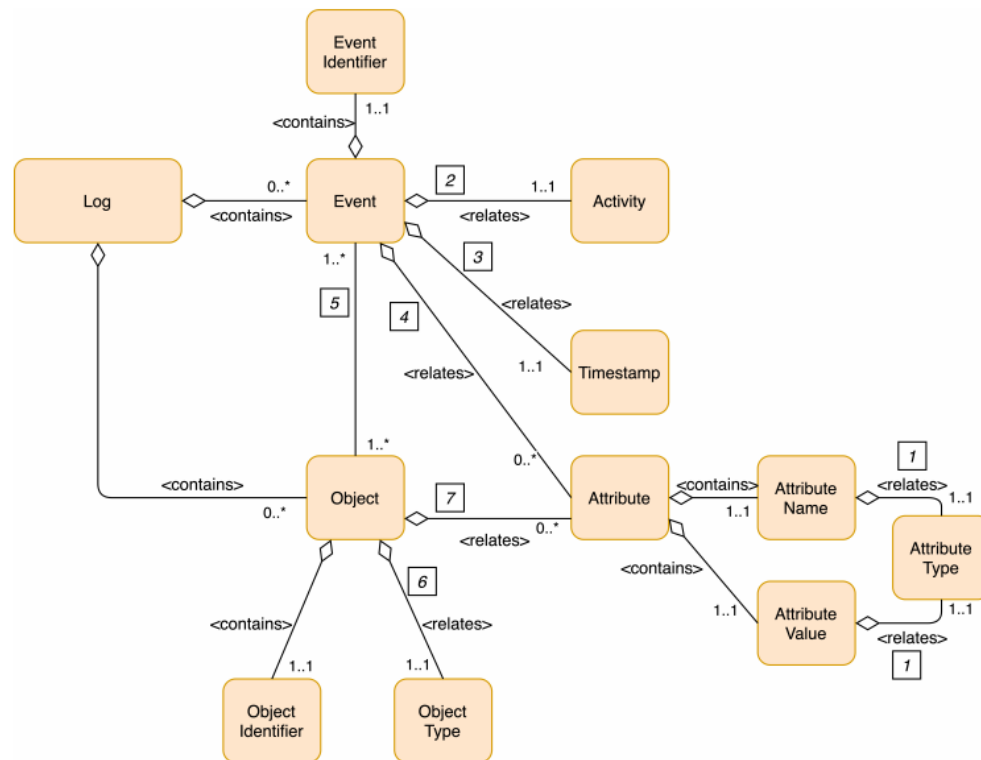
Anahita Farhang Ghahfarokhi

Gyunam Park

Alessandro Berti

Wil van der Aalst

January 8th, 2020



How to deal with this?



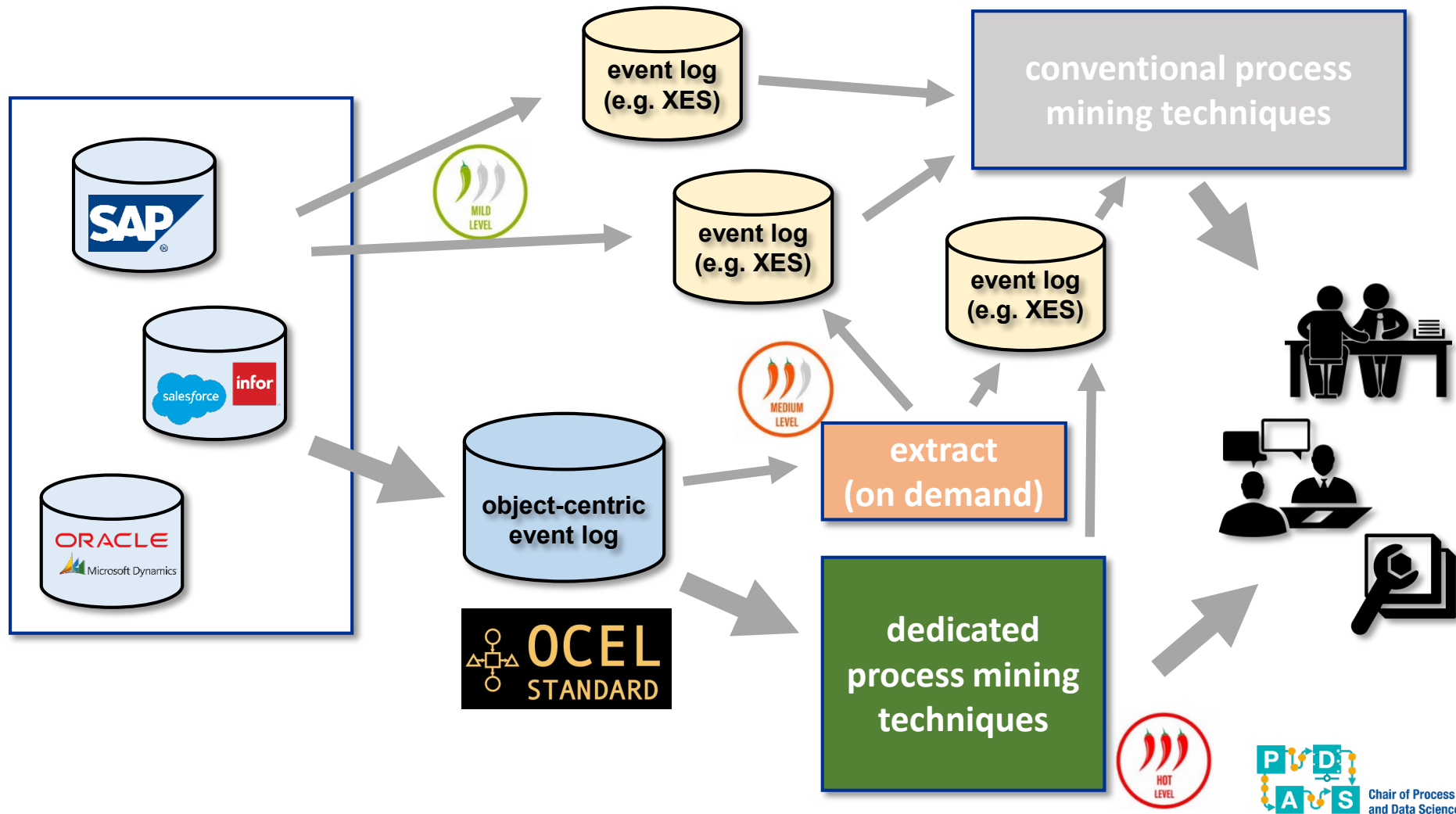
Directly extracting one or more conventional event logs (e.g. XES) realizing that there are may be convergence and divergence problems.



Extracting one object-centric event log and creating conventional event logs (e.g. XES) on demand.



Extracting one object-centric event log and using process mining techniques directly working on object-centric event logs.



Object-Centric Process Mining



extract from
data sources

Each row corresponds to an event which
refers to one activity and any number of
objects of (possibly many) different types.

automatically
discovered object-
centric Petri net

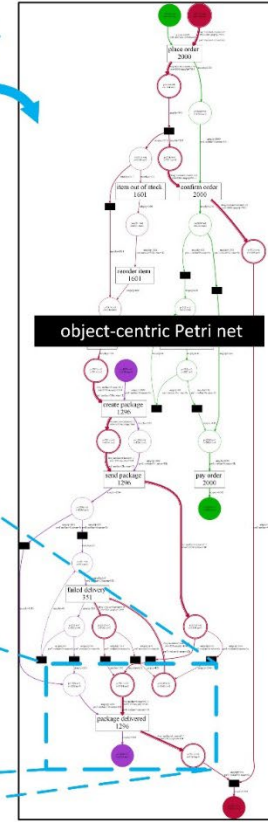
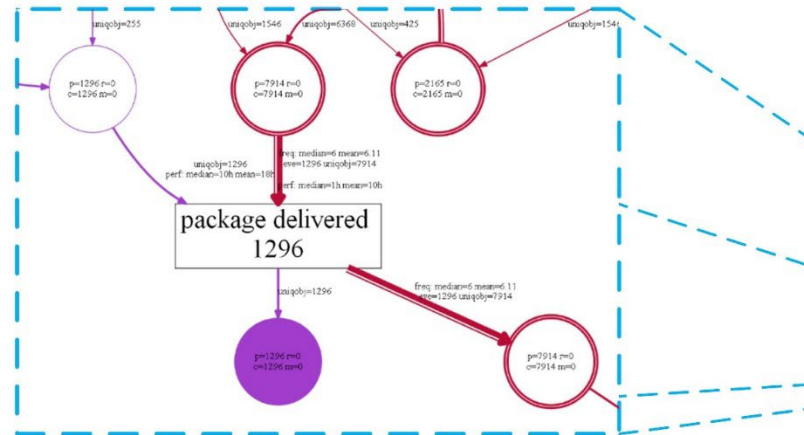
activity	time	orders	items	customers	products	price	weight
135001 place order	2019-12-24 17:27:18 (91269)		(884954,884905,884956,884957,884908)	(Wol van der Aalst)	(Fire Stick 4K,Echo Show 5,Echo Pad mini,Echo Show 8)	€ 913,06	3,70
135002 pick item	2019-12-24 17:33:55 (911261)		(884956)	(Christine Dobbels)	(Kindle Paperwhite)	€ 129,00	0,50
135003 pick item	2019-12-24 17:56			(Wol van der Aalst)	(Fire Stick 4K)	€ 89,99	0,28
135004 pick item	2019-12-24 17:58			(Kefang Ding)	(Fire Stick 4K)	€ 89,99	0,28
135005 any order	2019-12-24 18:05		(1048164)	(Mahsa Bahmani)	(Echo Studio,Kindle MacBook Pro,MacBook Air)	€ 4.984,98	4583,00
135006 pick item	2019-12-24 18:18 (911137)		(884823)	(Mahsa Bahmani)	(Kindle Paperwhite)	€ 129,00	0,50
135007 pick item	2019-12-24 18:19:53 (911182)		(884823)	(Christine Dobbels)	(Echo Show 8)	€ 129,99	0,98
135008 pick item	2019-12-24 18:19:59 (911198)		(1048164)	(Volker Brockhoff)	(Kindle Paperwhite)	€ 129,00	0,50
135009 any order	2019-12-24 18:47:07 (911263)		(884956,884957,884958,884959,884960,884961,884962)	(Christine Dobbels)	(Kindle Paperwhite,Pad,Kindle,Echo Show 5,Echo Studio,Kindle)	€ 1.348,95	5184,00
135010 place order	2019-12-24 19:34:23 (912170)		(884959)	(Jianrong Gao)	(Pad mini)	€ 454,00	0,28
135011 item out of stock	2019-12-24 20:39:02 (912124)		(884962)	(Wol van der Aalst)	(Phone 8)	€ 529,00	0,11
135012 package delivered	2019-12-24 20:46:47 (911119,911030,911209,911254,911254,911213,9)		(884188,884020,884749,884930,884926,884925,884706,88492)	(Kefang Ding)	(iPad Air,Echo Dot,MacBook Pro,iPad Air,Kindle Paperwhite,iPad Air,iPad)	€ 6.829,99	4719,00
135013 place order	2019-12-24 22:08:21 (912171)		(885000,885001,885002,885003)	(Muhammadreza Fani Sari)	(Fire Stick 4K,Kindle Paperwhite,Pad Air)	€ 1.348,99	1688,00
135014 pick item	2019-12-24 22:12:24 (911265)		(884814)	(Volker Brockhoff)	(Fire Stick)	€ 89,99	0,18
135015 place order	2019-12-25 08:08:00 (912172)		(885004,885005)	(Muhammadreza Fani Sari)	(iPhone 11 Pro,iPad Pro)	€ 2.253,00	0,87
135016 any order	2019-12-25 08:12:59 (912173)		(885004,885005)	(Muhammadreza Fani Sari)	(iPhone 11 Pro,iPad Pro)	€ 2.253,00	0,87
135017 any order	2019-12-25 09:13:08 (911248)		(884800,884801,884802)	(Jianrong Gao)	(Kindle Fire Stick 4K,Echo Pro)	€ 516,97	2043,00
135018 package delivered	2019-12-25 09:17:36 (911229,911130,911060,911226,911236)		(884824,884420,884173,884810,884805)	(Mahsa Bahmani)	(Kindle,Echo Show 8,iPad Pro,Echo MacBook Pro)	€ 9.908,27	4096,00

object-centric event log

object-centric Petri net

One of the 21887 events:

- **activity:** package delivered
- **time:** 2019-12-24 20:46:47
- **orders involved:** {991119,991030,991209,991254,991213,991206}
- **items involved:** {884386,884020,884749,884930,884926,884925,884766,884927,884736}
- **packages involved:** {660784}
- **customers involved:** {Kefang Ding}
- **products involved:** {iPad Air,Echo Dot,MacBook Pro,iPad Air,Kindle Paperwhite,iPad Air,iPad Pro,iPad,iPhone 11 Pro}
- **total price:** € 6.829,99
- **total weight:** 4,719 KG



W.M.P. van der Aalst and A. Berti.
Discovering Object-Centric Petri
Nets. Fundamenta Informaticae,
vol. 175, no. 1-4, pp. 1-40, 2020



Free Advice



Make process mining repeatable and actionable



*free
advice*

- ☐ Event logs provide views on reality.
- ☐ Process mining is not a project, but an ongoing activity.
 - ☐ The return-on-investment is typically low for one time extractions (proof-of-concepts are fine, but ...).
 - ☐ Results should be used on a daily basis.
- ☐ Not for one process, but for all processes you would like to improve.
 - ☐ Share efforts and expertise.
 - ☐ Use comparative process mining / benchmarking.

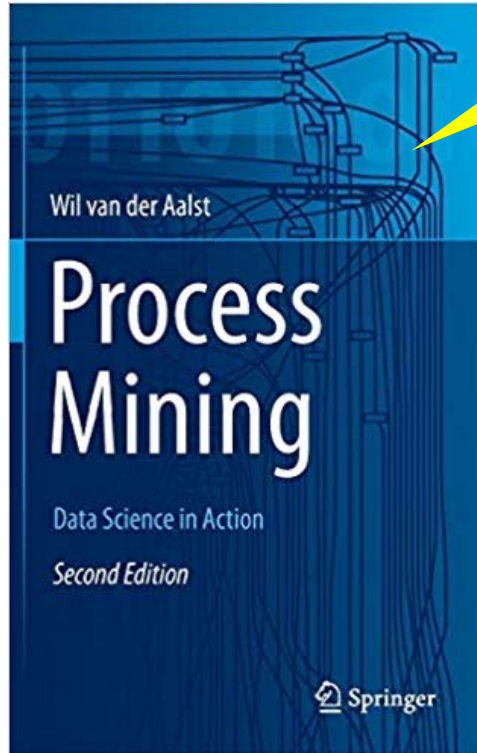
A close-up photograph of a person's hands being washed under a stream of water from a modern, chrome faucet. The water is clear and splashing slightly. The background is a blurred blue-grey color, suggesting a clean, professional environment like a hospital or office restroom. The overall tone is clean and hygienic.

Business process hygiene

Typical excuses:
privacy, data quality,
workload, etc.

Are you sure you need to
have a business case?

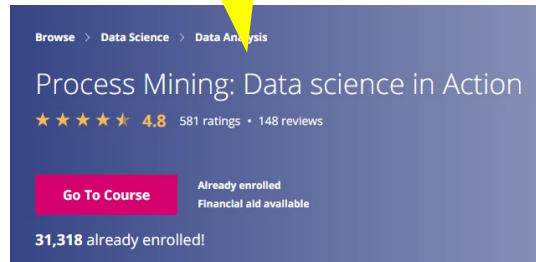
Learn more?



“PM Bible”

Over 135.000
participants

coursera



prof.dr.ir. Wil van der Aalst
RWTH Aachen University
W: vdaalst.com T:@wvdaalst

 **Fraunhofer**

FIT



Chair of Process
and Data Science

<https://www.coursera.org/learn/process-mining>



 **IEEE TASK FORCE ON PROCESS MINING**

www.tf-pm.org