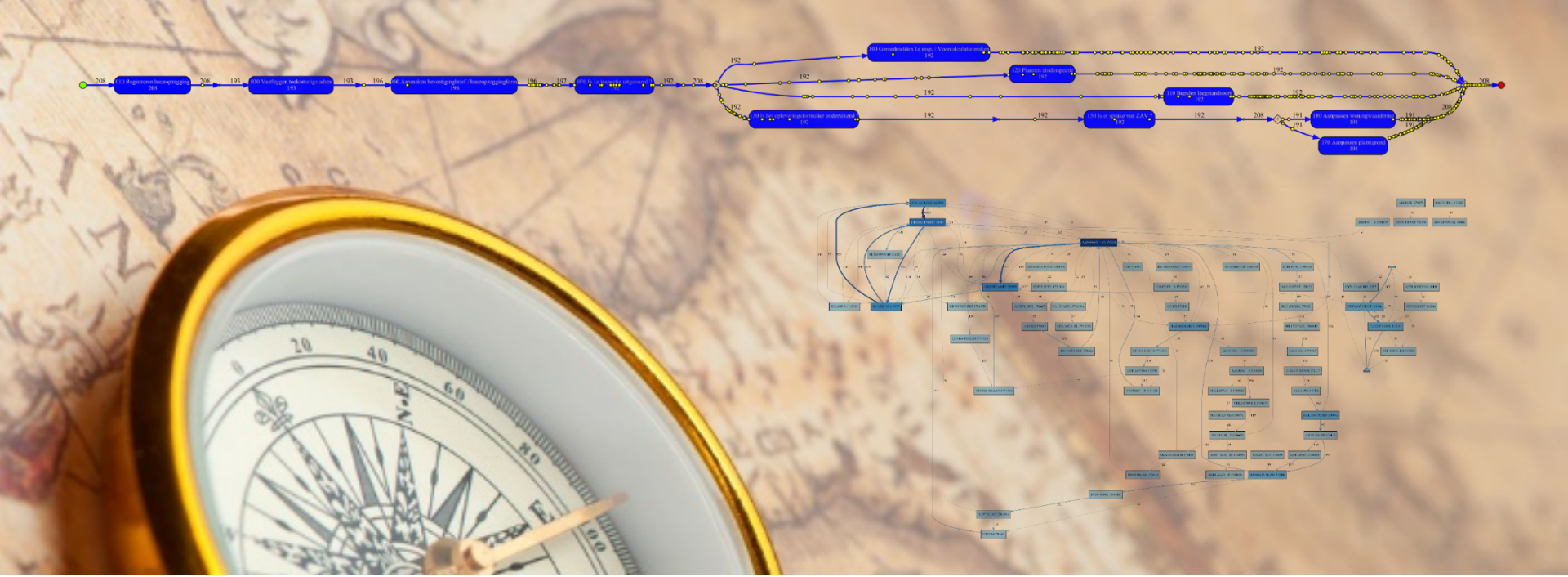


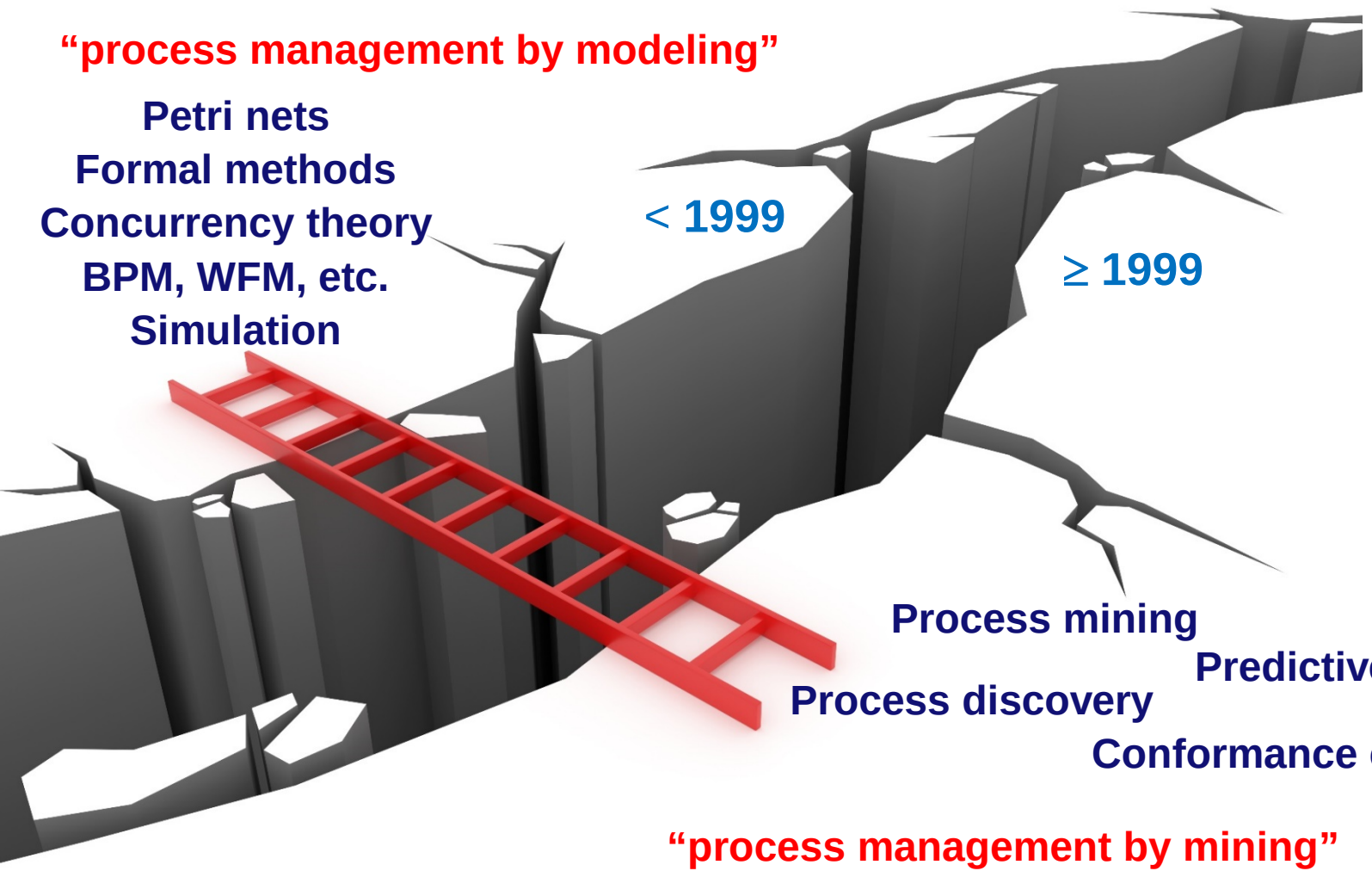
Adding Objects and Probabilities to Process Mining

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

ER Online Summer Seminar (EROSS-2020)
Wednesday, Sept. 2nd 2020

Fraunhofer FIT
Alexander von Humboldt Stiftung/Foundation
Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

© Wil van der Aalst (use only with permission & acknowledgements)



“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”

Chair of Process and Data Science

© Wil van der Aalst (use only with permission & acknowledgements)

Milestones

- 1999 start of process mining research at TU/e
- 2000-2002 Alpha and Heuristic miner
- 2004 first version of ProM
- 2004-2006 token-based conformance checking, organization mining, decision mining, etc.
- 2007 first process mining company (Futura PI)
- 2010 alignment-based conformance checking
- 2011 founding of Celonis
- 2011 first process mining book
- 2014 Coursera process mining MOOC
- 2016 “Process mining data science in action” book
- 2018 Market Guide for Process Mining by Gartner
- 2018 30+ process mining companies
- 2018 Celonis becomes a Unicorn
- 2019 ICPM 2019: First PM conf.

20 years of process mining

Over 30 process mining vendors today

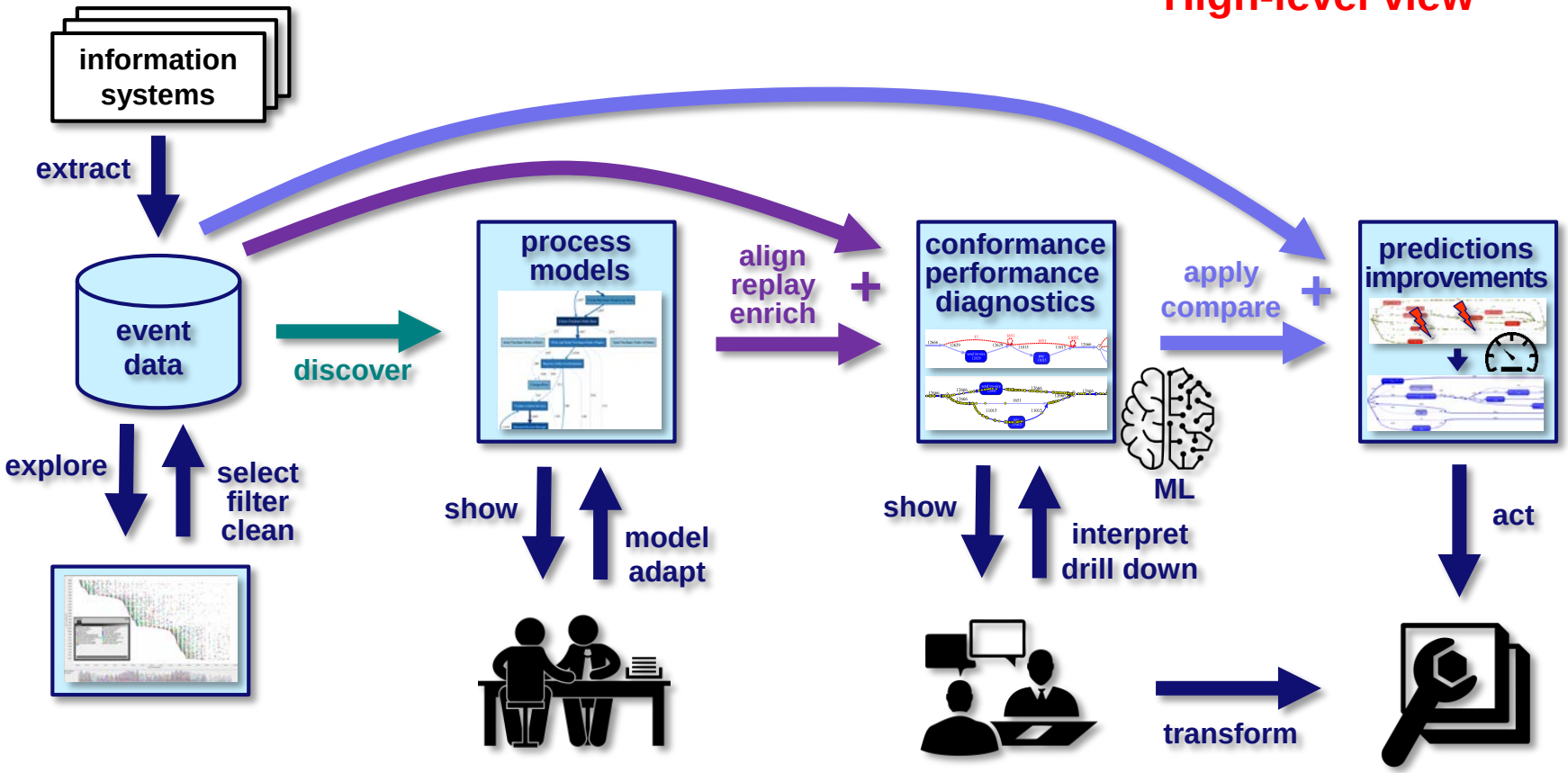
© Wil van der Aalst (use only with permission & acknowledgments)

IEEE TASK FORCE ON PROCESS MINING

www.tf-pm.org

© Wil van der Aalst (use only with permission & acknowledgments)

High-level view



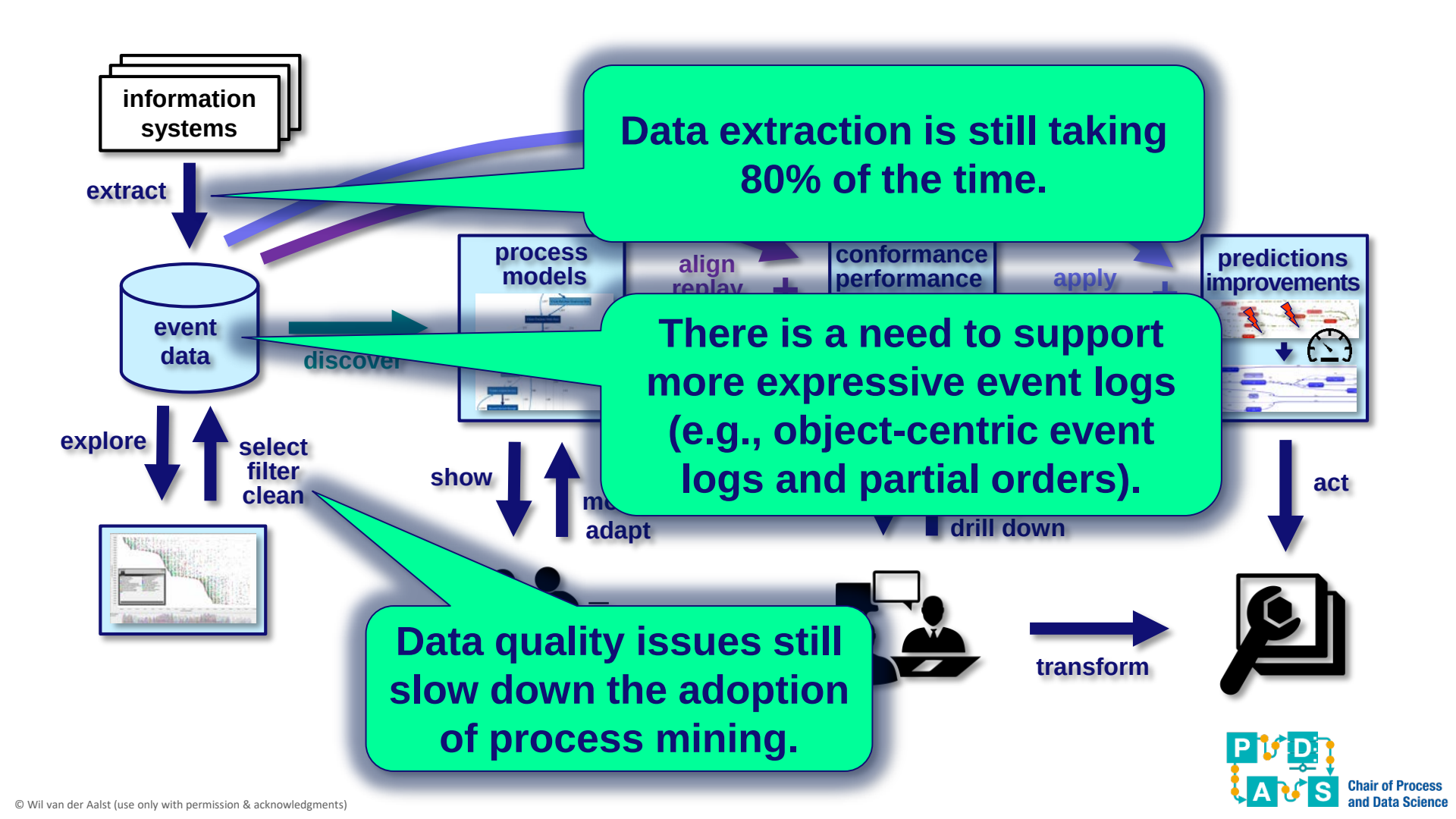
© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

Chair of Process and Data Science

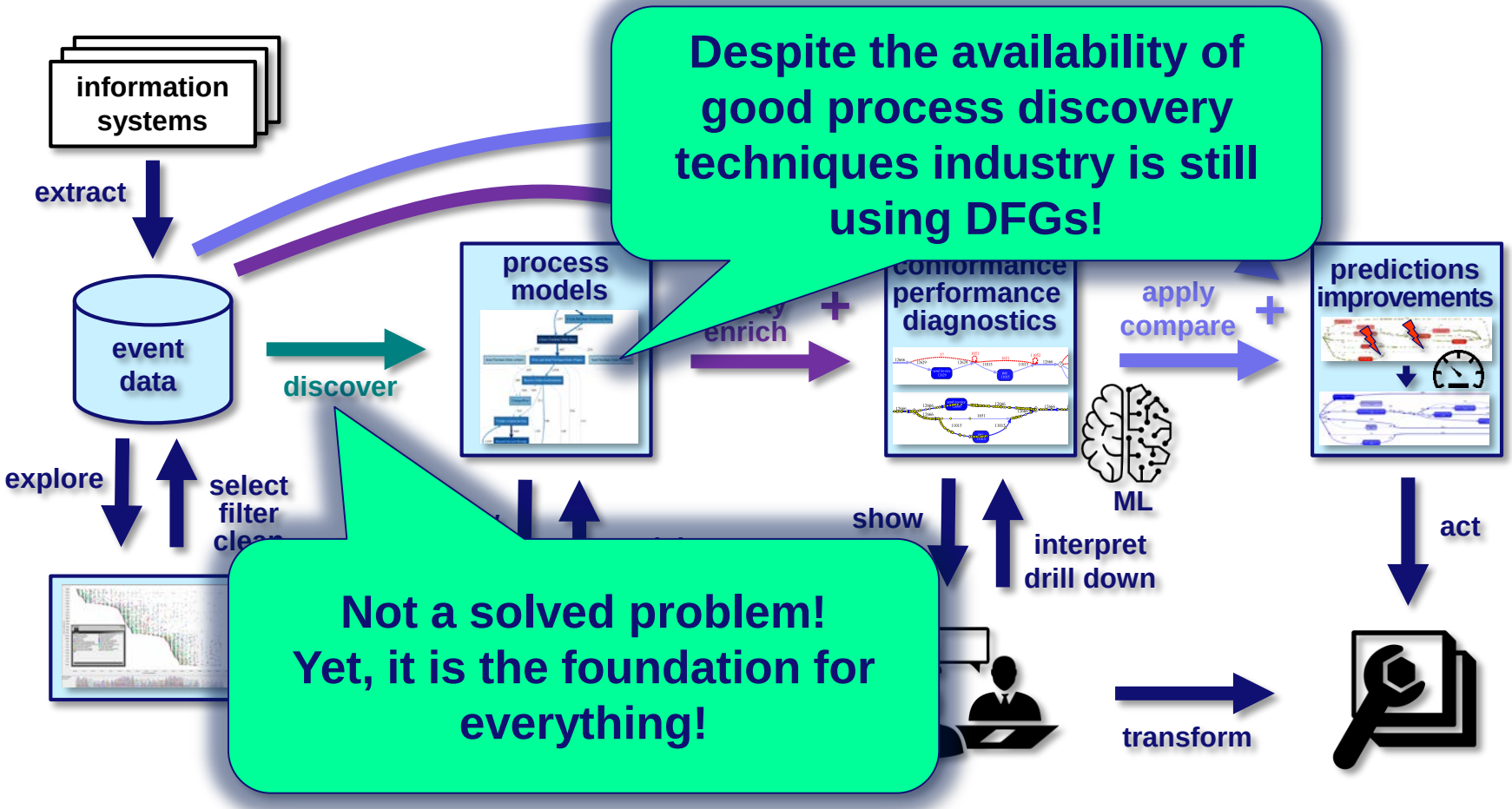
RWTH AACHEN UNIVERSITY

© Wil van der Aalst (use only with permission & acknowledgements)



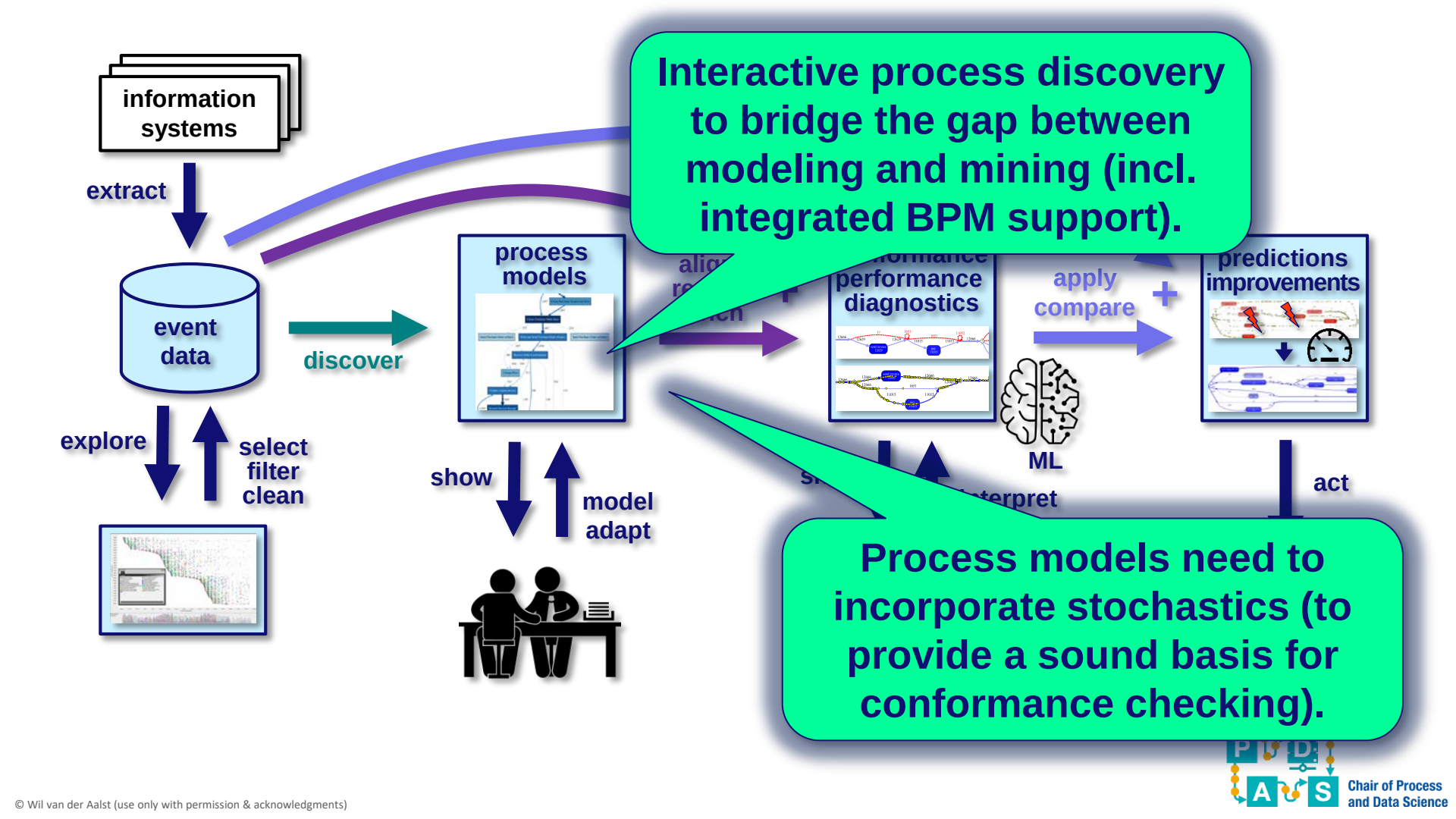
© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science



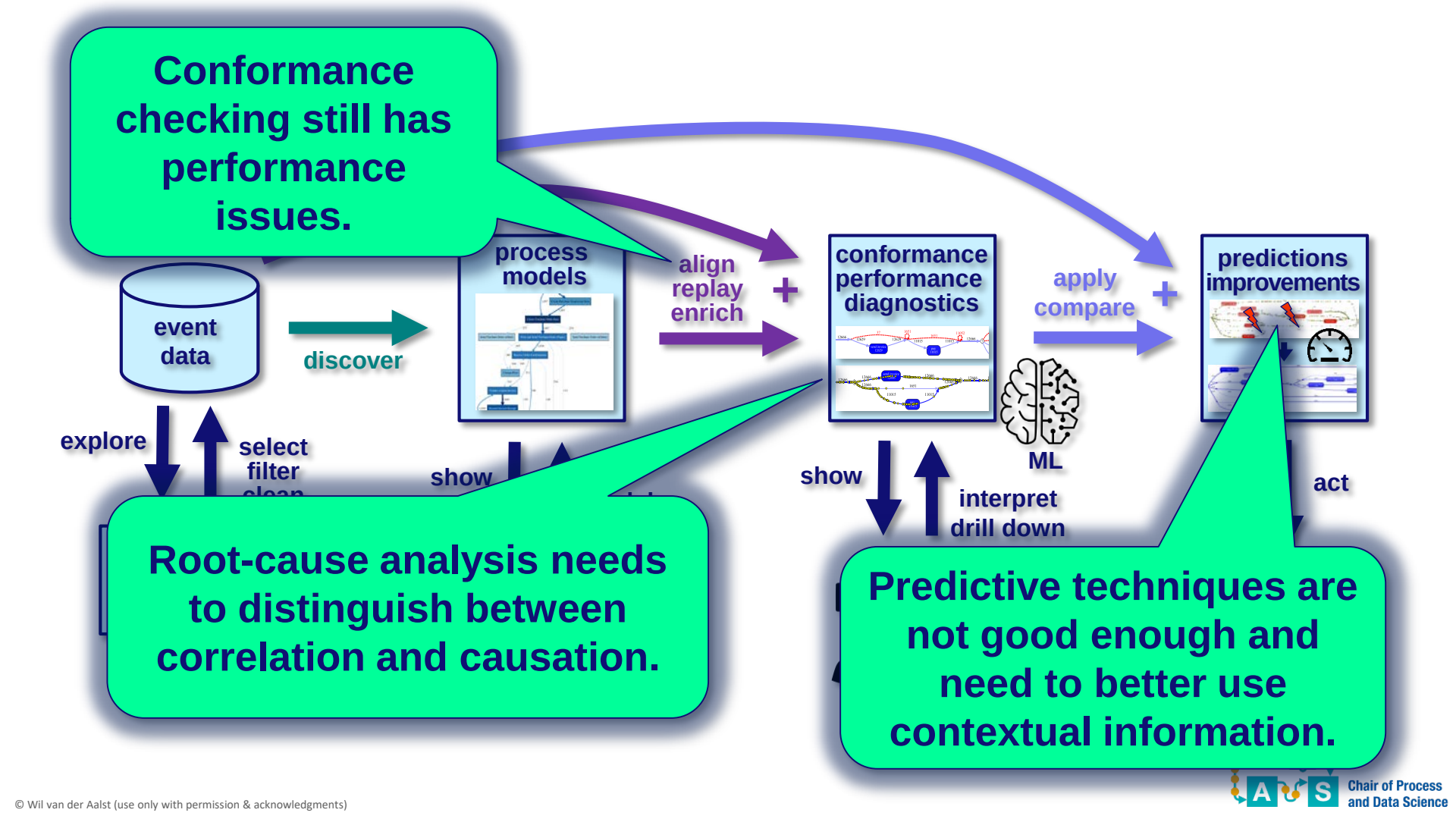
© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

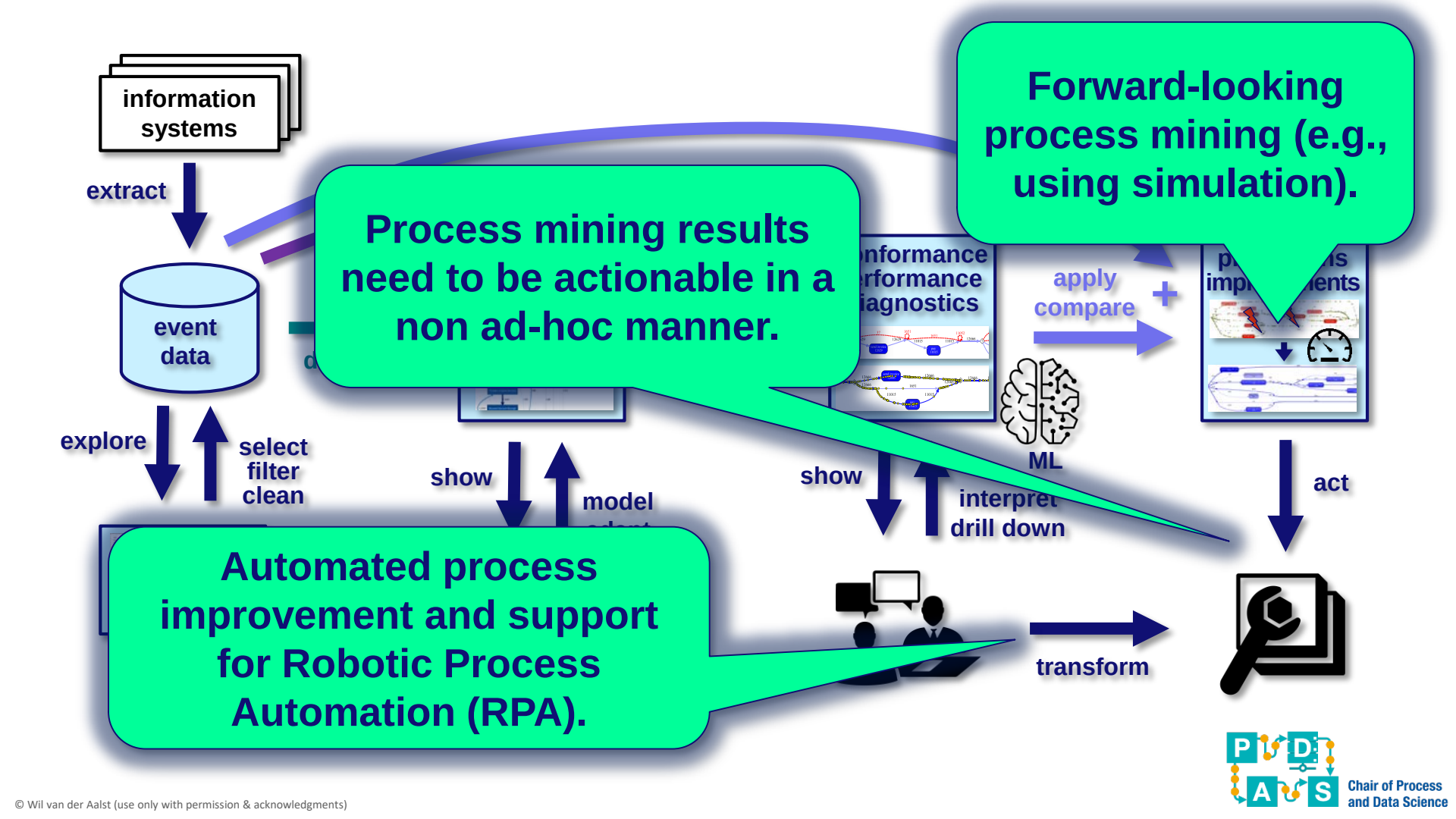


© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

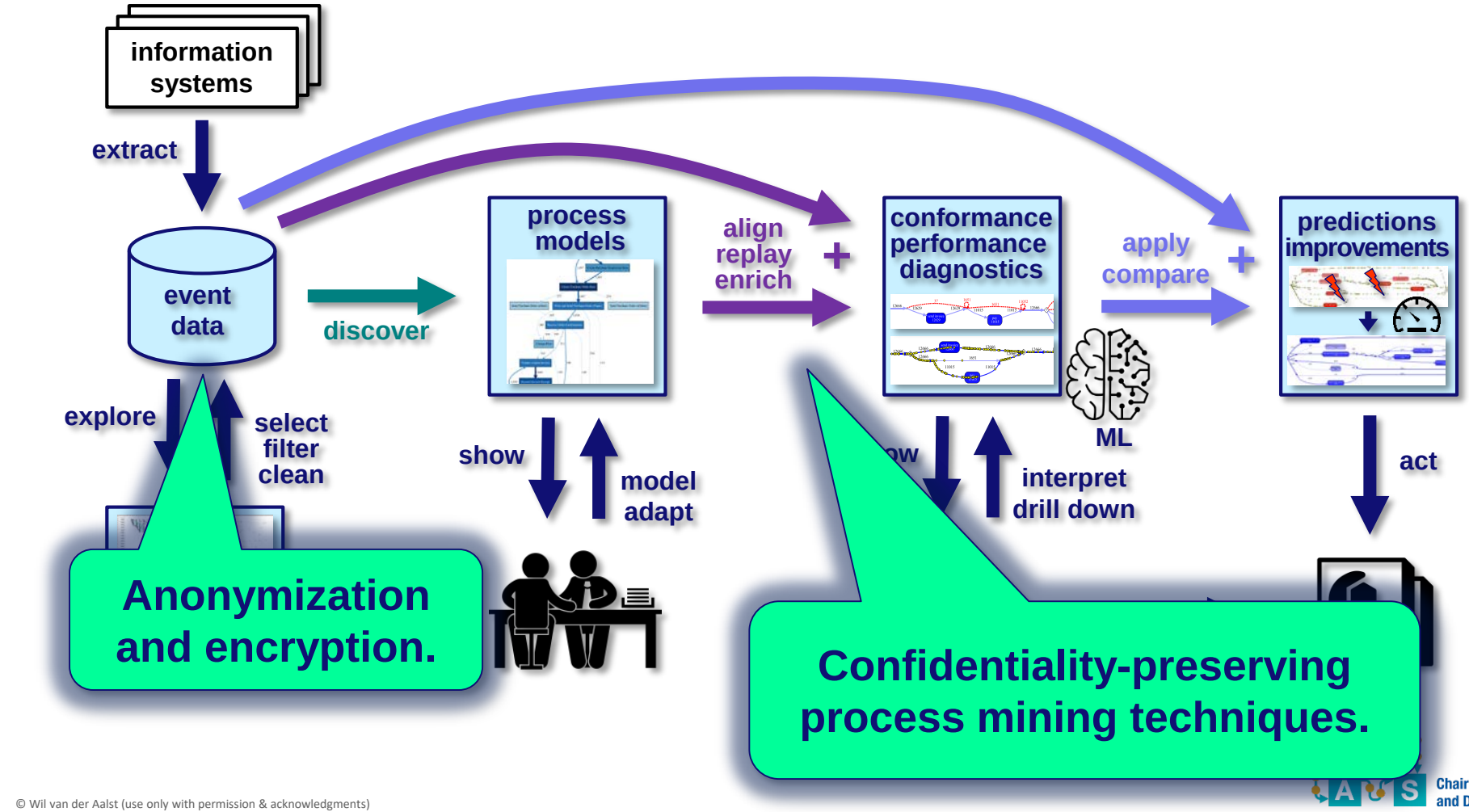


© Wil van der Aalst (use only with permission & acknowledgments)



© Wil van der Aalst (use only with permission & acknowledgments)

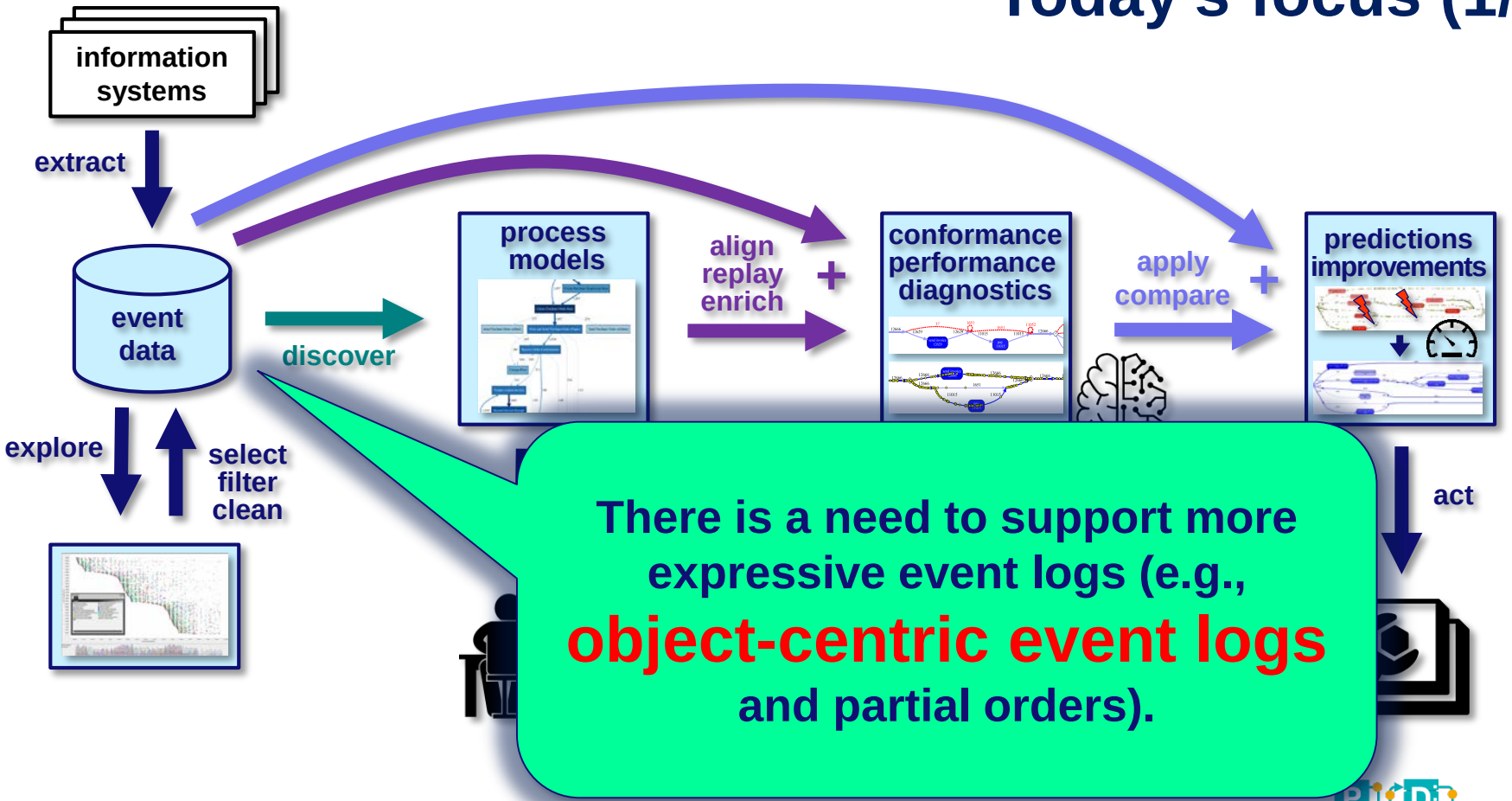
Chair of Process and Data Science



© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

Today's focus (1/2)



© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

© Wil van der Aalst (use only with permission & acknowledgements)

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
…	…	….	…	…	…	…	…

event

71,043 events
12,666 cases
7 activities

© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

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 S6 32 GB	543.99 €	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 S6 32 GB	543.99 €	1	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Galaxy S6 32 GB	543.99 €	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
…	…	….	…	…	…	…	…

event = case + activity + timestamp + …

© Wil van der Aalst (use only with permission & acknowledgments)

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

© Wil van der Aalst (use only with permission & acknowledgments)

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

© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

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

Order 6351

© Wil van der Aalst (use only with permission & acknowledgments)

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

Order 6351

Order 6352

© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

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

Order 6351

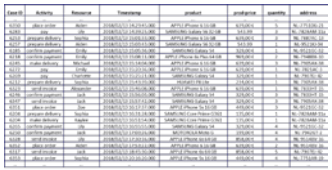
Order 6352

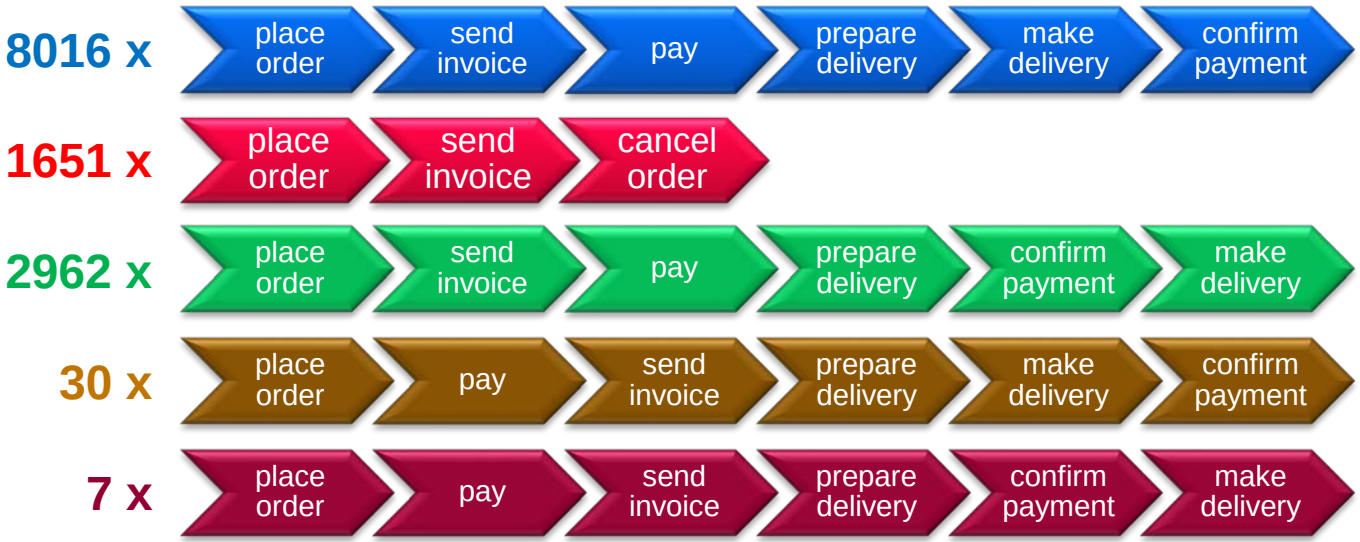
© Wil van der Aalst (use only with permission & acknowledgments)

Chair of Process and Data Science

Let’s look at the whole event log again

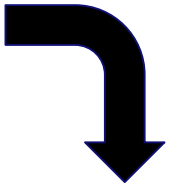
71,043 events
12,666 cases
7 activities

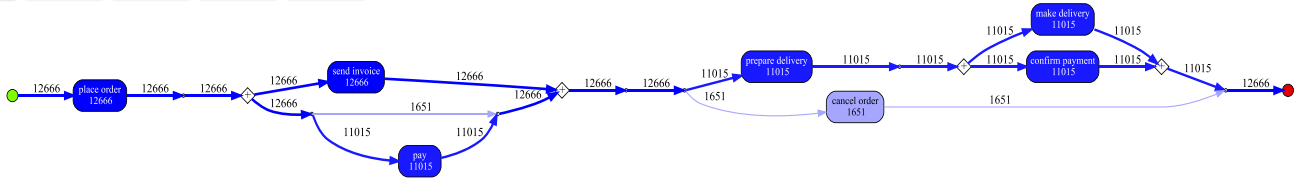


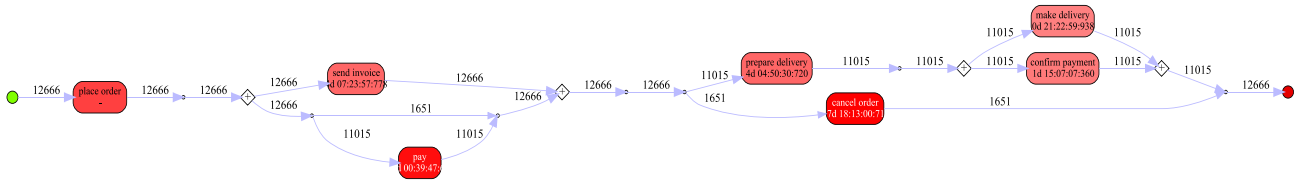


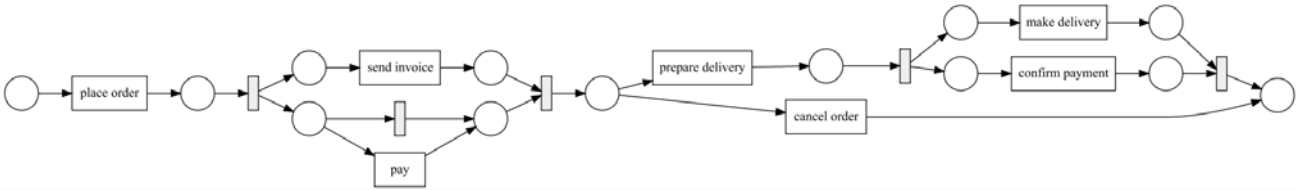
© Wil van der Aalst (use only with permission & acknowledgments)

Using the whole event log



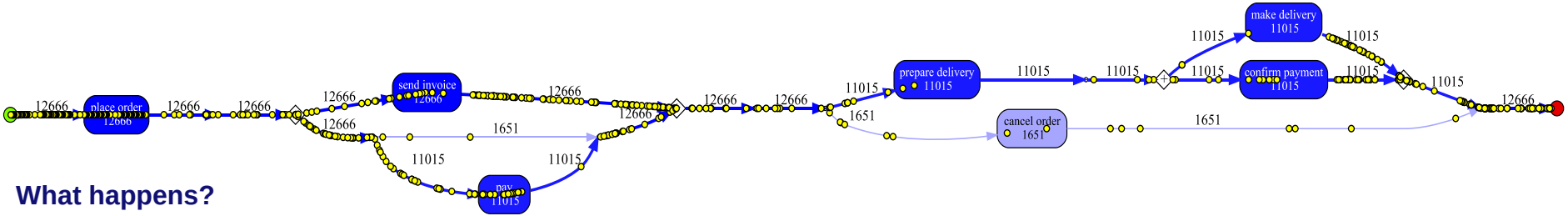




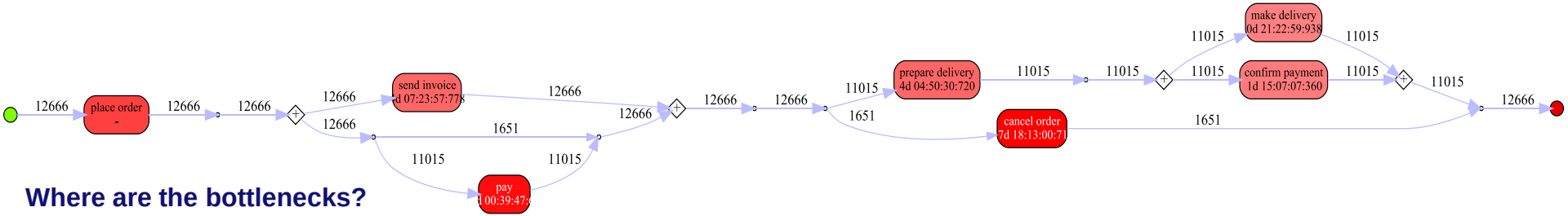


© Wil van der Aalst (use only with permission & acknowledgments)

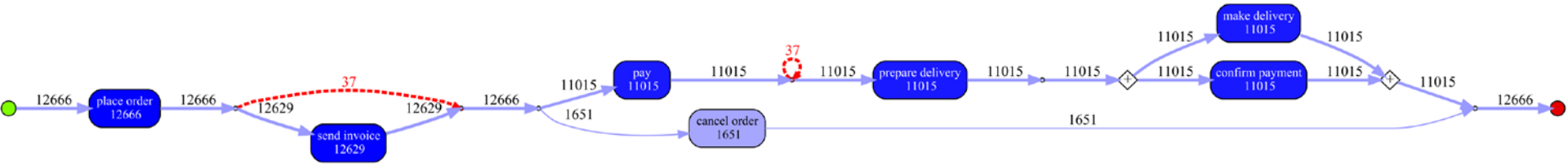
Performance and Compliance



What happens?



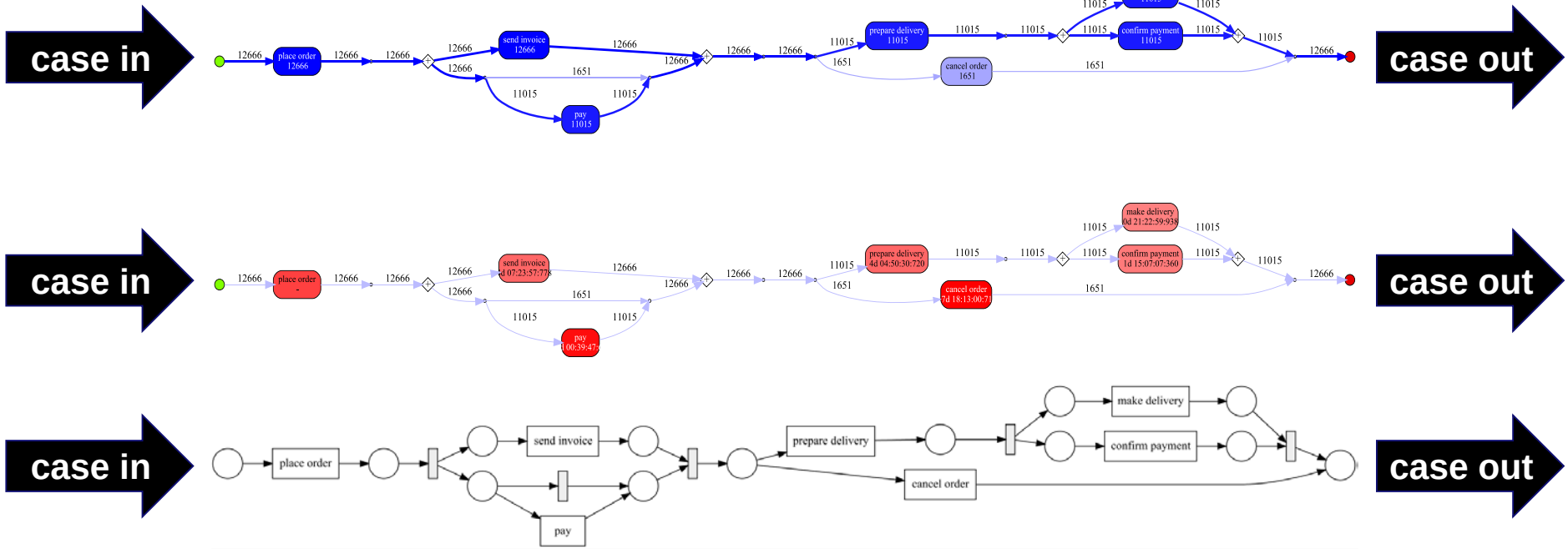
Where are the bottlenecks?



Where do we deviate from the happy path?

© Wil van der Aalst (use only with permission & acknowledgments)

Traditional Process Models Assume a Single Case Notion!



Also BPMN, UML ADs, EPCs, etc.

Chair of Process and Data Science

Object-Centric Process Mining

"event data and processes are not flat"

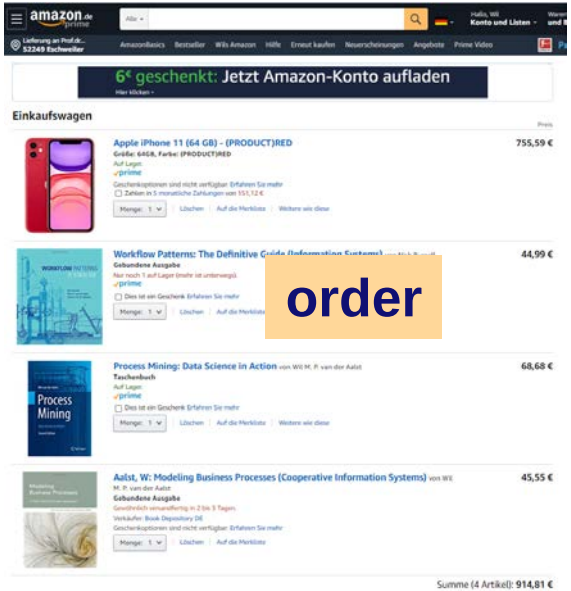
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

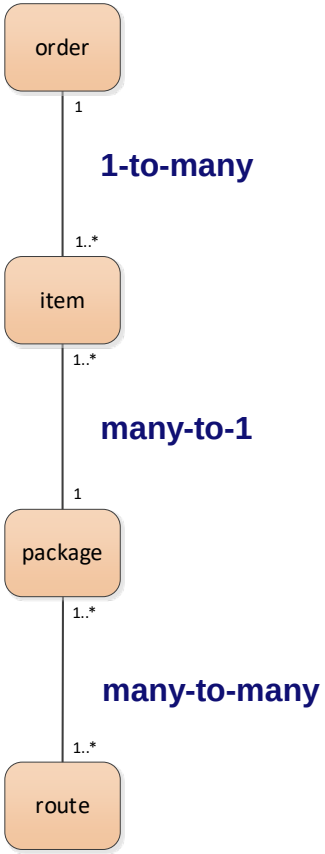
© Wil van der Aalst (use only with permission & acknowledgements)

Example illustrating object-centric PM

item



order

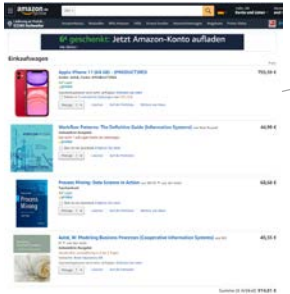


package

route

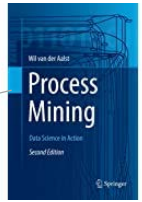
Example illustrating object-centric PM

(No activities, just describing the relationships among objects)



order1 order2 order3

item1 item2 item3 item4 item5 item6 item7

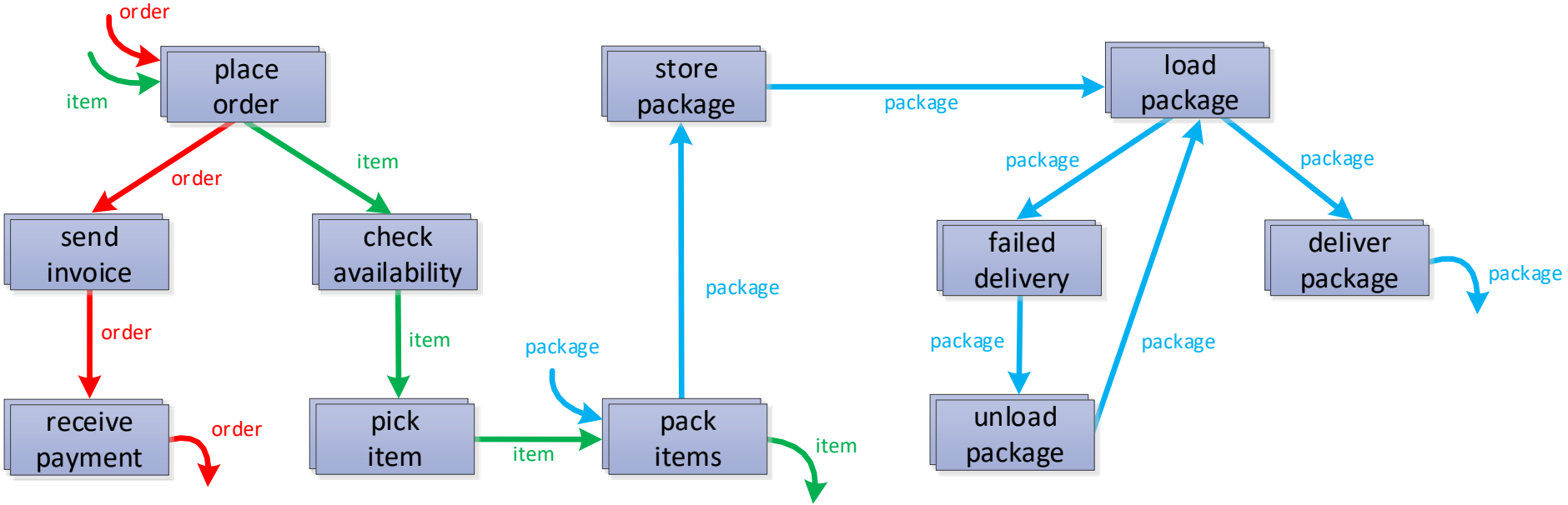


package1 package2 package3 package4

route1 route2 route3

© Wil van der Aalst (use only with permission & acknowledgments)

Goal: One model showing multiple object types



© Wil van der Aalst (use only with permission & acknowledgments)

See Wil van der Aalst: Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. SEFM 2019, 3-25 https://doi.org/10.1007/978-3-030-30446-1_1

Chair of Process and Data Science

Let's make the following assumption

activity	time	orders	items	packages	customers	products	price	weight
pick item	2019-12-26 12:04:46	991224}	{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,iPa	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}	{884977}	{}	{Junxiong Gao}	{iPhone X,1}	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 14:32:43	991251}	{884912}	{}	{Tobias Brockhoff}	{Fire Stick}	39.99	0.2
confirm order	2019-12-26 14:32:44	991282}	{885036,885037}	{}	{Lisa Mannel}	{Echo,Echo Dot}	134.98	1.16
pick item	2019-12-26 14:33:28	991278}	{885024}	{}	{Junxiong Gao}	{MacBook Pro}	2500.0	1.37
place order	2019-12-26 14:48:33	991284}	{885040,885041,885042,885043,885044}	{}	{Christine Dobbert}	{iPhone X,Fire Stick,MacBook Air,Echo Show 8,iPad Pro,1 Pro}	4222.98	2.79
failed delivery	2019-12-26 15:04:53	991240,99116	{884902,884922,884923,884913,884876,884938,884914,884949}	{660780}	{Tobias Brockhoff}	{iPad Air,Echo Studio,Echo Studio,Kindle,Kindle,Echo Show 8,iPad mini,iPad Pro,iPad P	5982.95	7.742
pick item	2019-12-26 15:20:05	991278}	{885024}	{}	{Junxiong Gao}	{iPhone X}	699.0	0.272
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}	3267.98	1.666
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	{884424,884933,884934,884935,884936,884937,884903}	{660797}	{Junxiong Gao}	{Echo Show 8,Kindle Paperwhite,iPad mini,Kindle,iPhone X,iPhone 8,Echo Sho	2145.97	1.4
payment reminder	2019-12-26 15:54:44	991169}	{884955,884956,884957,884958,884959}	{}	{Gyuchan Park}	{iPhone 8,Echo Plus,iPad Air,iPad mini}	1608.99	2.11
pick item	2019-12-26 15:55:38	991201}	{884917}	{}	{Lisa Mannel}	{Echo Show 8}	129.99	0.8
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

activity

timestamp

customers

tests

packages

machines

objects

orders

products

parts

patients

attributes

event = activity + timestamp + objects + attributes

© Wil van der Aalst (use only with permission & acknowledgments)

See Wil van der Aalst: Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. SEFM 2019, 3-25 https://doi.org/10.1007/978-3-030-30446-1_1

Chair of Process and Data Science

e the following assumption

a “place order” event may refer to multiple items

a “failed delivery” event refers to one package, one or more items, one or more orders, etc.

a “pick item” event refers to one item

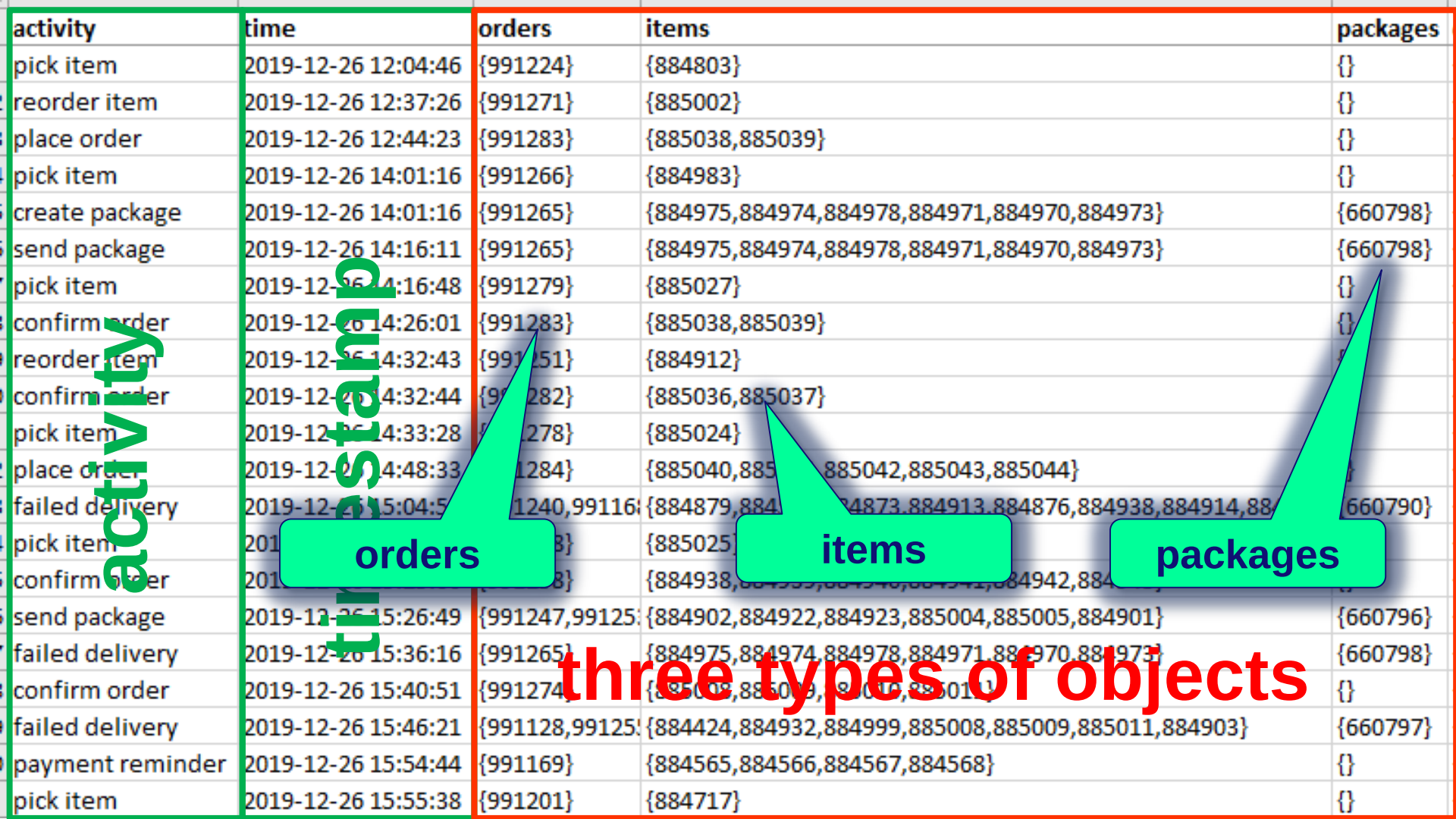
activity	time	orders	items	packages	customers		price	weight
pick item	2019-12-26 12:04:46	{991224}	{884803}	{}	{Wil van der Aalst}		529.0	0.21
reorder item	2019-12-26 12:37:26	{991271}	{885002}	{}	{Mohammadreza Fani Sani}		129.0	0.495
place order	2019-12-26 12:44:23	{991283}	{885038,885039}	{}	{Luis Santos}		2700.0	1.733
pick item	2019-12-26 14:01:16	{991266}	{884983}	{}	{Marco Pegoraro}		2200.0	1.25
create package	2019-12-26 14:01:16	{991265}	{884975,884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}		3506.97	[illegible]12
send package	2019-12-26 14:16:11	{991265}	{884975,884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}		3506.97	[illegible]12
pick item	2019-12-26 14:16:48	{991279}	{885027}	{}	{Claudia Graf}		799.0	0.166
confirm order	2019-12-26 14:26:01	{991[illegible]	{885038,885039}	{}	{Luis Santos}		2700.0	[illegible]33
reorder item	2019-12-26 14:32:43	[illegible]	{884912}	{}	{Tobias Brockhoff}	{Fire [illegible]	39.99	[illegible]
confirm order	2019-12-26 [illegible]	[illegible]	{885036,885037}	{}	{Lisa Mannel}	[illegible]	134.98	1.16
[illegible]	[illegible]	[illegible]	[illegible]	{}	{Junxiong Gao}	[illegible]k Pro}	2500.0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]885041,885042,885043,885044}	{}	{Christine Dobbert}	[illegible]hone X,Fire Stick,MacBook Air,Echo Show 8,iPhone 11 Pro}	4222.98	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]884561,884873,884913,884876,884938,884914,884941,	{660790}	{Tobias Brockhoff}	{iPad Air,Echo Studio,Echo Studio,Kindle,Kindle,Echo,iPad mini,iPad Pro,iPad P	5982.95	[illegible]42
[illegible]	[illegible]	[illegible]	[illegible]	{}	{Junxiong Gao}	{iPhone X}	699.0	[illegible]72
[illegible]	[illegible]	[illegible]	[illegible]884939,884940,884941,884942,884943}	{}	{Tobias Br[illegible]}	{Echo,Fire Stick,iPad mini,iPad Pro,iPad Pro,iPad Air}	3267.9[illegible]	[illegible]66
[illegible]	[illegible]	[illegible]	[illegible]884922,884923,885004,885005,884901}	{660796}	{Moha[illegible]adreza Fani Sani}	{MacBook Air,iPad mini,iPad Pro,iPhone 11 Pro,iPad Pro,MacBook Pro}	8496.0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]884974,884978,884971,884970,884973}	{660798}	{Seran Uysal}	{Fire Stick 4K,iPa[illegible]	[illegible]	[illegible]
confirm order	2019-12-26 15:40:51	{991274}	{885008,885009,885010,885011}	{}	{Junxiong Gao}	{Kindle,iPhone [illegible]	[illegible]	[illegible]
failed delivery	2019-12-26 15:46:21	{991128,99125[illegible]	{884424,884932,884999,885008,885009,885011,884903}	{660797}	{Junxiong Gao}	{Echo Show 8,Ki[illegible]	[illegible]	[illegible]
payment reminder	2019-12-26 15:54:44	{991169}	{884565,884566,884567,884568}	{}	{Gyunam Park}	{iPhone 8,Echo [illegible]	[illegible]	[illegible]
pick item	2019-12-26 15:55:38	{991201}	{884717}	{}	{Seran Uysal}	{Echo Show 8}	[illegible]	[illegible]
pick item	2019-[illegible]:00:38	{991251}	{884912}	{}	{Tobias Brockhoff}	{Fire Stick}	39.99	[illegible]
reorder item	2019-12-26 16:04:42	{991265}	{884977}	{}	{Seran Uysal}	{Fire Stick 4K}	89.99	[illegible]28
payment reminder	2019-12-26 16:11:39	{991164}	{884542,884543,884544,884545,884546,884547}	{}	{Ju[illegible]}	{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}	{}	{Li[illegible]}	{iPhone 8}	529.0	0.21
create package	2019-12-26 16:22:04	{991263,99126[illegible]	{884967,884964,884966}	{660799}	{Luis Santos}	{iPad Air,iPhone 8,iPad}	1500.0	1.133

objects

let’s simplify to focus on the essence

Objects are typed and events may have any number of objects.

© Wil van der Aalst (use only with permission & acknowledgments)



activity	time	orders	items	packages
pick item	2019-12-26 12:04:46	{991224}	{884803}	{}
reorder item	2019-12-26 12:37:26	{991271}	{885002}	{}
place order	2019-12-26 12:44:23	{991283}	{885038,885039}	{}
pick item	2019-12-26 14:01:16	{991266}	{884983}	{}
create package	2019-12-26 14:01:16	{991265}	{884975,884974,884978,884971,884970,884973}	{660798}
send package	2019-12-26 14:16:11	{991265}	{884975,884974,884978,884971,884970,884973}	{660798}
pick item	2019-12-26 14:16:48	{991279}	{885027}	{}
confirm order	2019-12-26 14:26:01	{991283}	{885038,885039}	{}
reorder item	2019-12-26 14:32:43	{991251}	{884912}	
confirm order	2019-12-26 14:32:44	{991282}	{885036,885037}	
pick item	2019-12-26 14:33:28	{991278}	{885024}	
place order	2019-12-26 14:48:33	{991284}	{885040,885042,885043,885044}	
failed delivery	2019-12-26 15:04:5	1240,99116	{884879,884873,884913,884876,884938,884914,88	660790}
pick item	201	3}	{885025}	
confirm order	201	8}	{884938,884942,884	
send package	2019-12-26 15:26:49	{991247,99125	{884902,884922,884923,885004,885005,884901}	{660796}
failed delivery	2019-12-26 15:36:16	{991265}	{884975,884974,884978,884971,884970,884973}	{660798}
confirm order	2019-12-26 15:40:51	{991274}	{885038,885039,885040,885041}	{}
failed delivery	2019-12-26 15:46:21	{991128,99125	{884424,884932,884999,885008,885009,885011,884903}	{660797}
payment reminder	2019-12-26 15:54:44	{991169}	{884565,884566,884567,884568}	{}
pick item	2019-12-26 15:55:38	{991201}	{884717}	{}

Disco - New project

products

column is used

Case

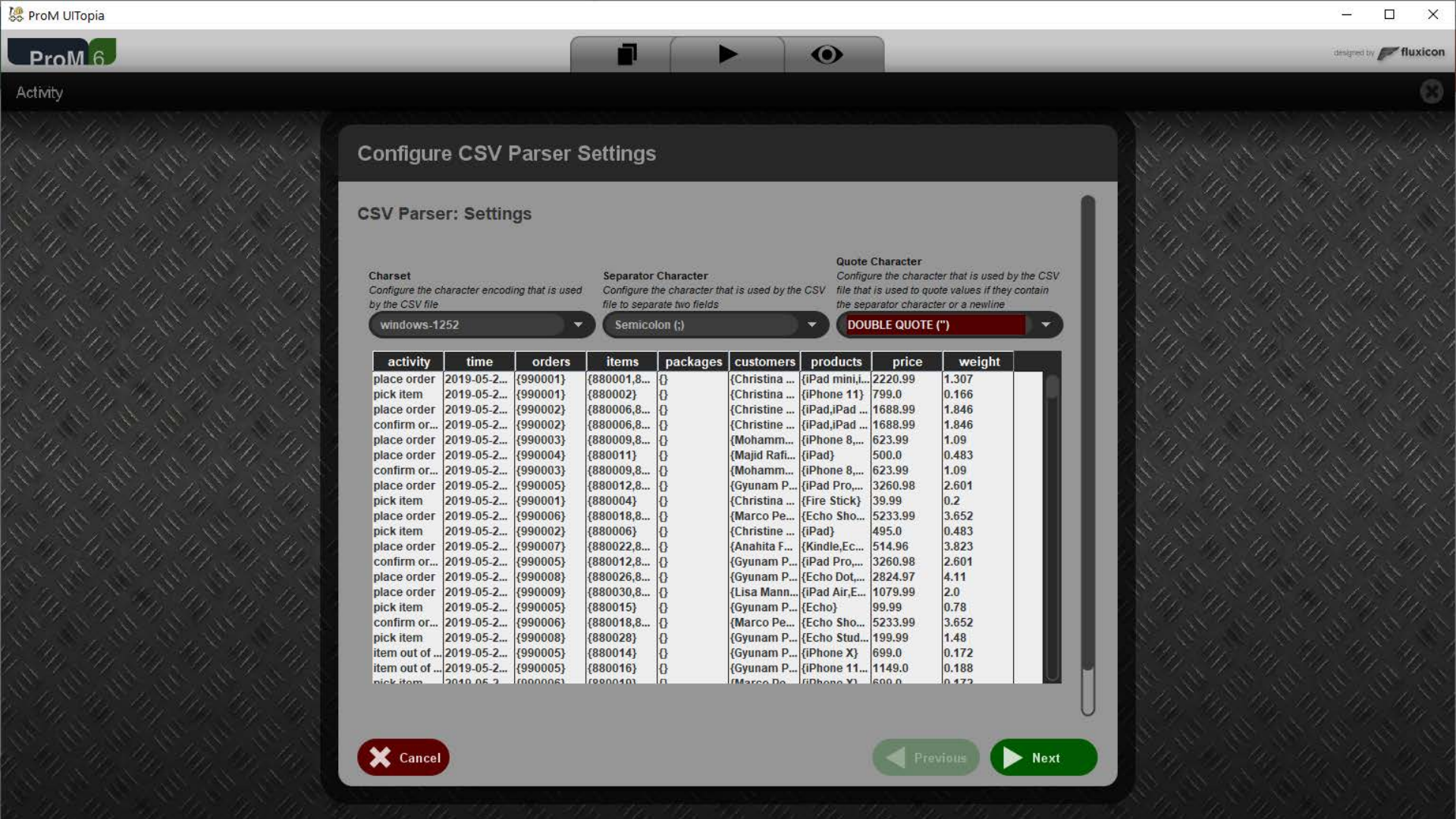
	activity	time	orders	items	packages
255	pick item	2019-05-27 15:04:22	{990039}	{880154}	{}
256	pick item	2019-05-27 15:04:51	{990033}	{880121}	{}
257	confirm order	2019-05-27 15:14:58	{990036}	{880131,880132,880133,880134}	{}
258	pick item	2019-05-27 15:17:22	{990038}	{880147}	{}
259	create package	2019-05-27 15:17:22	{990007}	{880023,880025,880024,880022}	{660010}
260	reorder item	2019-05-27 15:43:45	{990038}	{880146}	{}
261	send package	2019-05-27 15:49:53	{990007}	{880023,880025,880024,880022}	{660010}
262	pick item	2019-05-27 16:07:59	{990005}	{880014}	{}
263	pick item	2019-05-27 16:08:21	{990030}	{880107}	{}
264	place order	2019-05-27 16:12:18	{990046}	{880173,880174,880175,880176}	{}
265	pick item	2019-05-27 16:14:56	{990013}	{880047}	{}
266	item out of stock	2019-05-27 16:17:26	{990041}	{51}	{}
267	pick item	2019-05-27 16:18:58	{990033}	{4}	{}
268	create package	2019-05-27 16:18:58	{990018,990039}	{6,880070,880069,880067,880154,880071}	{660011}
269	send package	2019-05-27 16:24:47	{990016,990014,990031}	0,880051,880050,880056,880059,880049,880058,880057,880117,880...	{660007}
270	pick item	2019-05-27 16:45:33	{990030}	0113}	{}
271	reorder item	2019-05-27 16:50:50	{990043}	{880168}	{}
272	pick item	2019-05-27 16:51:02	{990021}	{880080}	{}
273	pay order	2019-05-27 16:56:23	{990023}	{880089,880090}	{}
274	item out of stock	2019-05-27 17:10:55	{990037}	{880143}	{}
275	pay order	2019-05-27 17:14:22	{990034}	{880125,880126,880127,880128}	{}
276	confirm order	2019-05-27 17:14:43	{990039}	{880150,880151,880152,880153,880154,880155,880156}	{}
277	item out of stock	2019-05-27 17:19:33	{990039}	{880155}	{}
278	confirm order	2019-05-27 17:21:37	{990043}	{880165,880166,880167,880168,880169}	{}
279	pick item	2019-05-27 17:21:57	{990029}	{880105}	{}
280	send package	2019-05-27 17:32:16	{990009,990015,990024}	{880032,880053,880055,880031,880092,880091,880030}	{660005}
281	pick item	2019-05-27 17:40:09	{990037}	{880139}	{}
282	place order	2019-05-27 17:44:23	{990047}	{880177,880178,880179,880180}	{}
283	reorder item	2019-05-27 17:58:16	{990032}	{880118}	{}

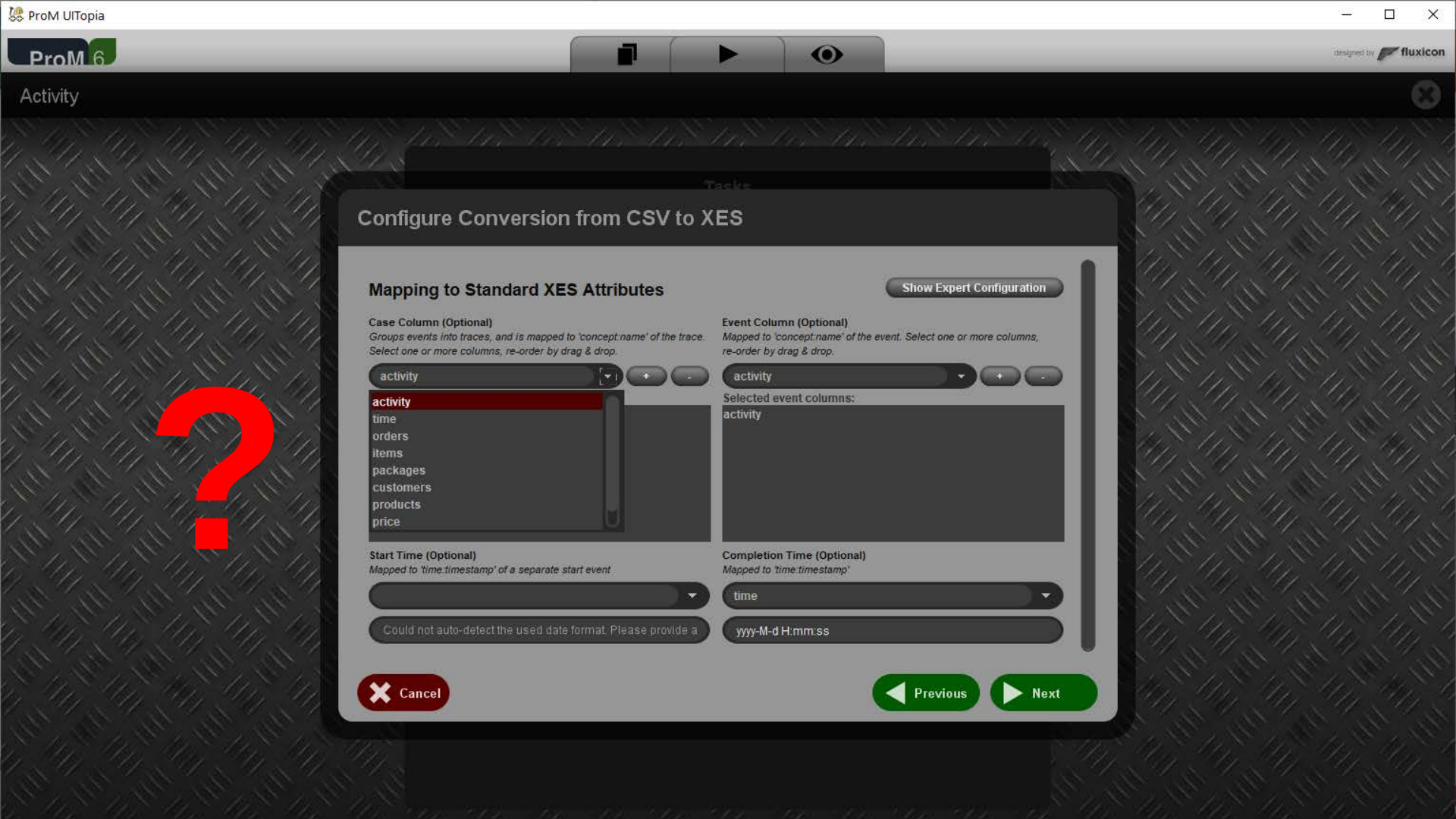
?

Cancel

File encoding: UTF-8

Use quotes


ProM UITopia
ProM 6
designed by fluxicon
Activity
Configure CSV Parser Settings
CSV Parser: Settings
Charset
Configure the character encoding that is used by the CSV file
windows-1252
Separator Character
Configure the character that is used by the CSV file to separate two fields
Semicolon (;)
Quote Character
Configure the character that is used by the CSV file that is used to quote values if they contain the separator character or a newline
DOUBLE QUOTE (")
activity time orders items packages customers products price weight
place order 2019-05-2... {990001} {880001,8... {} {Christina ... {iPad mini,i... 2220.99 1.307
pick item 2019-05-2... {990001} {880002} {} {Christina ... {iPhone 11} 799.0 0.166
place order 2019-05-2... {990002} {880006,8... {} {Christine ... {iPad,iPad ... 1688.99 1.846
confirm or... 2019-05-2... {990002} {880006,8... {} {Christine ... {iPad,iPad ... 1688.99 1.846
place order 2019-05-2... {990003} {880009,8... {} {Mohamm... {iPhone 8,... 623.99 1.09
place order 2019-05-2... {990004} {880011} {} {Majid Rafi... {iPad} 500.0 0.483
confirm or... 2019-05-2... {990003} {880009,8... {} {Mohamm... {iPhone 8,... 623.99 1.09
place order 2019-05-2... {990005} {880012,8... {} {Gyunam P... {iPad Pro,... 3260.98 2.601
pick item 2019-05-2... {990001} {880004} {} {Christina ... {Fire Stick} 39.99 0.2
place order 2019-05-2... {990006} {880018,8... {} {Marco Pe... {Echo Sho... 5233.99 3.652
pick item 2019-05-2... {990002} {880006} {} {Christine ... {iPad} 495.0 0.483
place order 2019-05-2... {990007} {880022,8... {} {Anahita F... {Kindle,Ec... 514.96 3.823
confirm or... 2019-05-2... {990005} {880012,8... {} {Gyunam P... {iPad Pro,... 3260.98 2.601
place order 2019-05-2... {990008} {880026,8... {} {Gyunam P... {Echo Dot,... 2824.97 4.11
place order 2019-05-2... {990009} {880030,8... {} {Lisa Mann... {iPad Air,E... 1079.99 2.0
pick item 2019-05-2... {990005} {880015} {} {Gyunam P... {Echo} 99.99 0.78
confirm or... 2019-05-2... {990006} {880018,8... {} {Marco Pe... {Echo Sho... 5233.99 3.652
pick item 2019-05-2... {990008} {880028} {} {Gyunam P... {Echo Stud... 199.99 1.48
item out of ... 2019-05-2... {990005} {880014} {} {Gyunam P... {iPhone X} 699.0 0.172
item out of ... 2019-05-2... {990005} {880016} {} {Gyunam P... {iPhone 11... 1149.0 0.188
Cancel
Previous
Next


ProM UITopia
ProM 6
designed by fluxicon
Activity
Tasks
Configure Conversion from CSV to XES
Mapping to Standard XES Attributes
Show Expert Configuration
Case Column (Optional)
Groups events into traces, and is mapped to 'concept:name' of the trace. Select one or more columns, re-order by drag & drop.
activity
+
-
activity
time
orders
items
packages
customers
products
price
Event Column (Optional)
Mapped to 'concept:name' of the event. Select one or more columns, re-order by drag & drop.
activity
+
-
Selected event columns:
activity
Start Time (Optional)
Mapped to 'time:timestamp' of a separate start event
Completion Time (Optional)
Mapped to 'time:timestamp'
time
Could not auto-detect the used date format. Please provide a
yyyy-M-d H:mm:ss
Cancel
Previous
Next
?

Need to flatten the event data when using a conventional process mining technique

- **Pick an object type as the case notion.**
- **Replicate each event for each object of the corresponding type.**

activity	time	orders	items	packages
…	…	…	…	…
place order	2019-12-26	{991283}	{885038,885039}	{ }
…	…	…	…	…

one event if order is used as a case notion

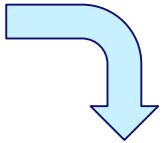
two events if item is used as a case notion

no event if package is used as a case notion

© Wil van der Aalst (use only with permission & acknowledgments)

Order as a case notion

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	{88001, 88002}	{ }
pick item	2020-6-22	{99001}	{88001}	{ }
pick item	2020-6-23	{99001}	{88002}	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	{88002, 88003, 88004}	{66001}
…	…	…	…	…



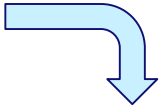
activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	99001	{88001, 88002}	{ }
pick item	2020-6-22	99001	{88001}	{ }
pick item	2020-6-23	99001	{88002}	{ }
…	…	…	…	…
send package	2020-6-25	99001	{88002, 88003, 88004}	{66001}
send package	2020-6-25	99002	{88002, 88003, 88004}	{66001}
…	…	…	…	…

Events may be duplicated

© Wil van der Aalst (use only with permission & acknowledgments)

Item as a case notion

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	{88001, 88002}	{ }
pick item	2020-6-22	{99001}	{88001}	{ }
pick item	2020-6-23	{99001}	{88002}	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	{88002, 88003, 88004}	{66001}
…	…	…	…	…



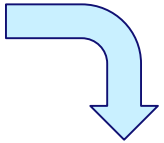
activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	88001	{ }
place order	2020-6-20	{99001}	88002	{ }
pick item	2020-6-22	{99001}	88001	{ }
pick item	2020-6-23	{99001}	88002	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	88002	{66001}
send package	2020-6-25	{99001, 99002}	88003	{66001}
send package	2020-6-25	{99001, 99002}	88004	{66001}

Events may be duplicated

© Wil van der Aalst (use only with permission & acknowledgments)

Package as a case notion

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	{88001, 88002}	{ }
pick item	2020-6-22	{99001}	{88001}	{ }
pick item	2020-6-23	{99001}	{88002}	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	{88002, 88003, 88004}	{66001}
…	…	…	…	…



activity	time	orders	items	packages
…	…	…	…	…
send package	2020-6-25	99002	{88002, 88003, 88004}	66001
…	…	…	…	…

Events may disappear

© Wil van der Aalst (use only with permission & acknowledgments)

Possible problems

- **Deficiency**: Events in the original event log that have no corresponding events in the flattened event log may unintentionally disappear from the data set.
- **Convergence**: Events referring to multiple objects of the selected type are replicated, possibly leading to unintentional duplication.
- **Divergence**: Events referring to different objects of a type not selected as the case notion are considered to be causally related.

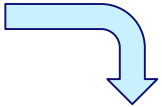
© Wil van der Aalst (use only with permission & acknowledgments)

See Wil van der Aalst: Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. SEFM 2019, 3-25 https://doi.org/10.1007/978-3-030-30446-1_1

Convergence

Events referring to multiple objects of the selected type are replicated, possibly leading to unintentional duplication

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	{88001, 88002}	{ }
pick item	2020-6-22	{99001}	{88001}	{ }
pick item	2020-6-23	{99001}	{88002}	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	{88002, 88003, 88004}	{66001}
…	…	…	…	…



activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	88001	{ }
place order	2020-6-20	{99001}	88002	{ }
pick item	2020-6-22	{99001}	88001	{ }
pick item	2020-6-23	{99001}	88002	{ }
…	…	…	…	…
send package	2020-6-25	{99001, 99002}	88002	{66001}
send package	2020-6-25	{99001, 99002}	88003	{66001}
send package	2020-6-25	{99001, 99002}	88004	{66001}

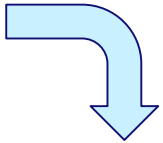
How to compute costs, times, frequencies, etc. when events are replicated?

© Wil van der Aalst (use only with permission & acknowledgments)

Divergence

Events referring to different objects of a type not selected as the case notion are considered to be causally related

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	{99001}	{88001, 88002, 88003}	{ }
pick item	2020-6-22	{99001}	{88001}	{ }
pick item	2020-6-23	{99001}	{88002}	{ }
pack item	2020-6-22	{99001}	{88002}	{ }
pack item	2020-6-23	{99001}	{88001}	{ }
pick item	2020-6-22	{99001}	{88003}	{ }
pack item	2020-6-23	{99001}	{88003}	{ }
…	…	…	…	…

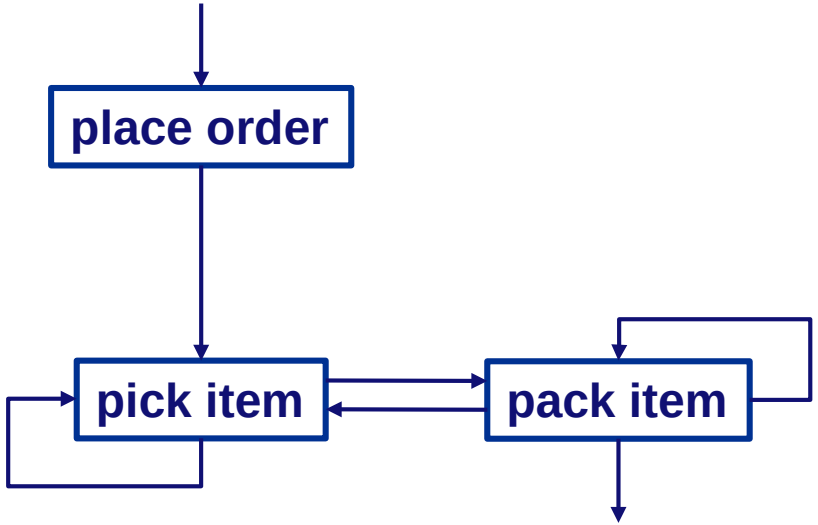


activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	99001	{88001, 88002, 88003}	{ }
pick item	2020-6-22	99001	{88001}	{ }
pick item	2020-6-23	99001	{88002}	{ }
pack item	2020-6-22	99001	{88002}	{ }
pack item	2020-6-23	99001	{88001}	{ }
pick item	2020-6-22	99001	{88003}	{ }
pack item	2020-6-23	99001	{88003}	{ }
…	…	…	…	…

© Wil van der Aalst (use only with permission & acknowledgments)

Divergence

Events referring to different objects of a type not selected as the case notion are considered to be causally related



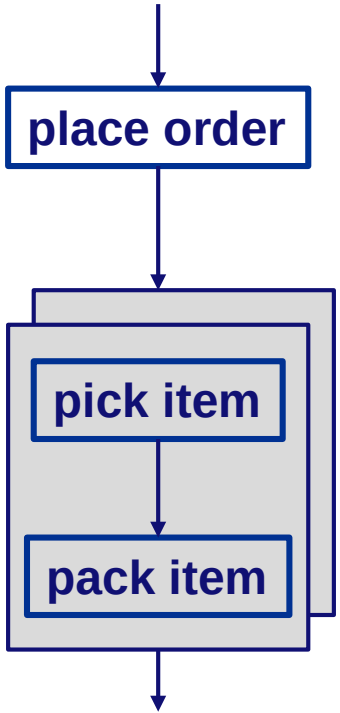
Things happen in a fixed order but this is not visible!

activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	99001	{88001, 88002, 88003}	{ }
pick item	2020-6-22	99001	{88001}	{ }
pick item	2020-6-23	99001	{88002}	{ }
pack item	2020-6-22	99001	{88002}	{ }
pack item	2020-6-23	99001	{88001}	{ }
pick item	2020-6-22	99001	{88003}	{ }
pack item	2020-6-23	99001	{88003}	{ }
…	…	…	…	…

© Wil van der Aalst (use only with permission & acknowledgments)

Divergence

Events referring to different objects of a type not selected as the case notion are considered to be causally related



activity	time	orders	items	packages
…	…	…	…	…
place order	2020-6-20	99001	{88001, 88002, 88003}	{ }
pick item	2020-6-22	99001	{88001}	{ }
pick item	2020-6-23	99001	{88002}	{ }
pack item	2020-6-22	99001	{88002}	{ }
pack item	2020-6-23	99001	{88001}	{ }
pick item	2020-6-22	99001	{88003}	{ }
pack item	2020-6-23	99001	{88003}	{ }
…	…	…	…	…

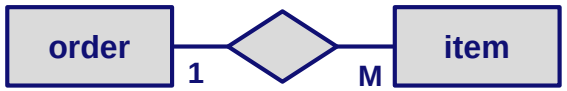
Concurrency and causality!

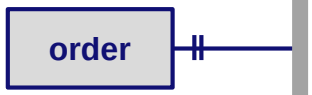
© Wil van der Aalst (use only with permission & acknowledgments)

Relevance? See any d

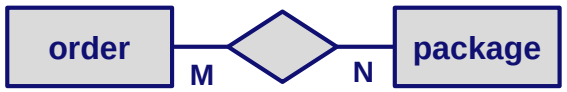
(Chen notation, Crow's foot notation, UML class diagrams)

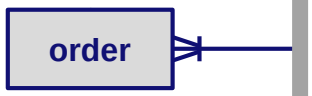
one-to-many



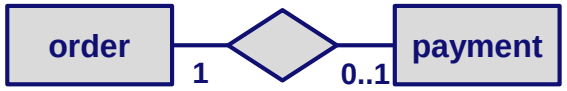


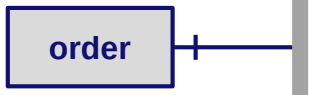
many-to-many

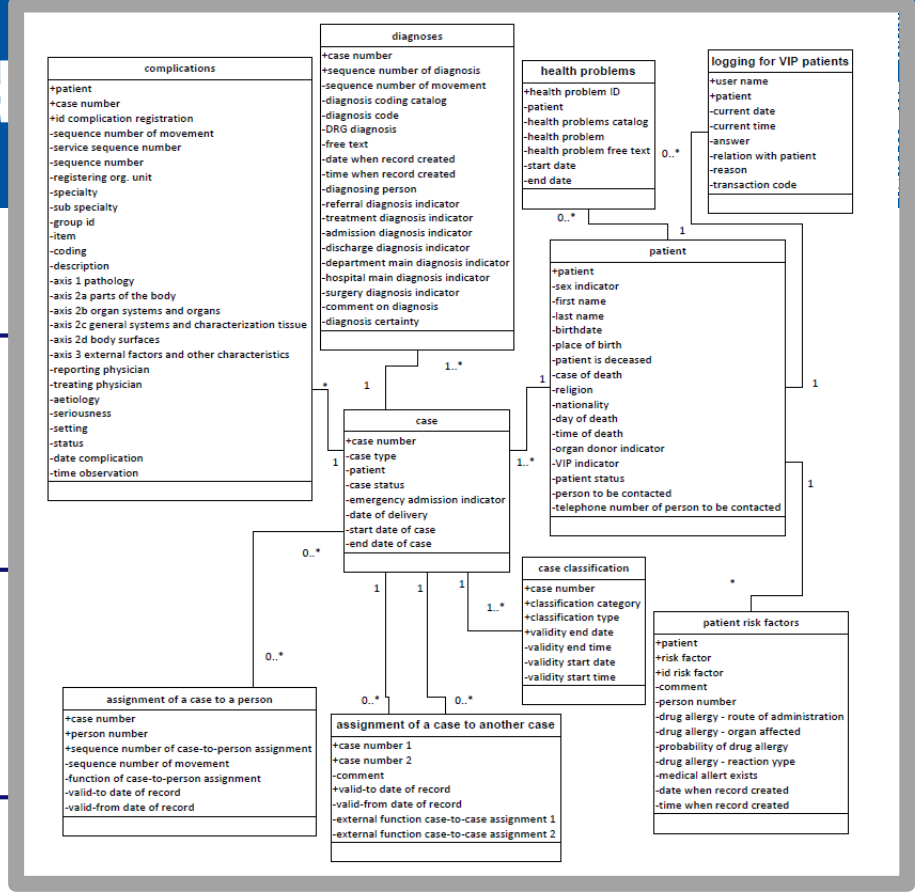




one-to-one







Very few relations are one-to-one!

Hence, a single case notion is not enough!

© Wil van der Aalst (use only with permission & acknowledgments)

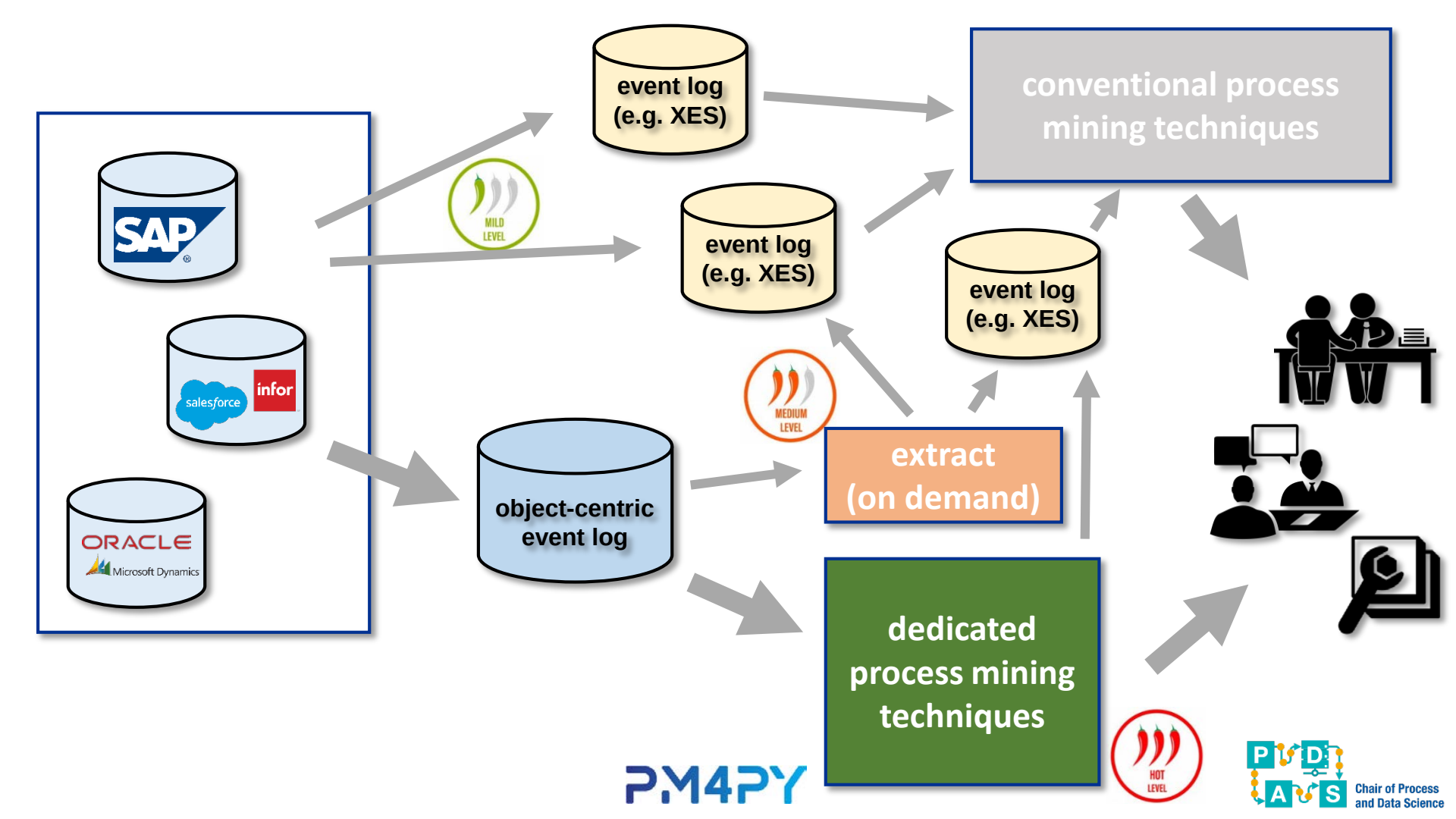
How to deal with this?

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.

© Wil van der Aalst (use only with permission & acknowledgments)


event log
(e.g. XES)
conventional process
mining techniques
MILD
LEVEL
event log
(e.g. XES)
event log
(e.g. XES)
SAP
salesforce
infor
MEDIUM
LEVEL
extract
(on demand)
object-centric
event log
ORACLE
Microsoft Dynamics
dedicated
process mining
techniques
HOT
LEVEL
PM4PY
Chair of Process
and Data Science

Object-Centric Process Mining

extract from data sources

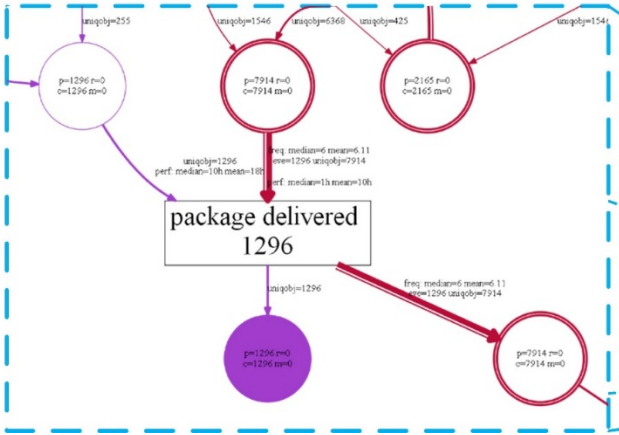
Each row corresponds to an event which refers to one activity and any number of objects of (possibly many) different types.

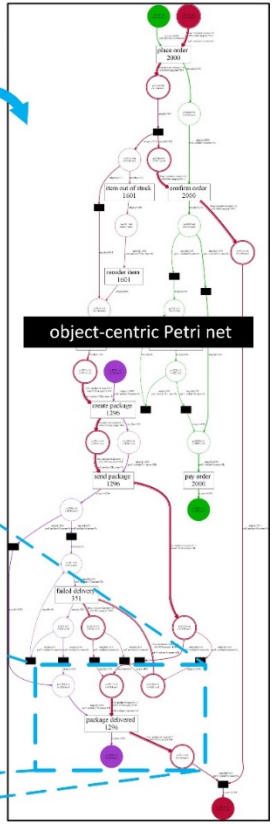
automatically discovered object-centric Petri net

object-centric event log

	activity	time	orders	items	packages	customers	products	price	weight
13501	place order	2019-12-24 17:27:18 (991269)		(884994,884995,884996,884997,884998)	[]	(Wil van der Aalst)	(Fire Stick 4K,Echo Show 5,Echo Plus,iPad mini,Echo Show 8)	€ 913,96	3,70
13502	pick item	2019-12-24 17:33:55 (991261)		(884954)	[]	(Christine Dobbert)	(Kindle Paperwhite)	€ 129,00	0,50
13503	pick item	2019-12-24 17:50:			[]	(Wil van der Aalst)	(Fire Stick 4K)	€ 89,00	0,28
13504	pick item	2019-12-24 17:58:			[]	(Kefang Ding)	(Fire Stick 4K)	€ 89,00	0,28
13505	pay order	2019-12-24 18:05:		850)	[]	(Mahsa Bafrani)	(Echo Studio,Kindle,MacBook Pro,MacBook Air)	€ 4.984,98	4583,00
13506	pick item	2019-12-24 18:18:30 (991237)		(884864)	[]	(Mahnaz Qafari)	(Kindle Paperwhite)	€ 129,00	0,50
13507	pick item	2019-12-24 18:19:53 (991182)		(884823)	[]	(Christine Dobbert)	(Echo Show 8)	€ 129,99	0,98
13508	pick item	2019-12-24 18:33:59 (991258)		(884940)	[]	(Tobias Brockhoff)	(iPad mini)	€ 449,00	0,28
13509	pay order	2019-12-24 18:47:07 (991261)		(884956,884957,884958,884959,884960,884961,884962)	[]	(Christine Dobbert)	(Kindle Paperwhite,iPad,Kindle,Echo Show 5,Echo Show 5,Echo Studio,Ki	€ 1.168,95	5184,00
13510	place order	2019-12-24 19:14:23 (991270)		(884999)	[]	(Junxiong Gao)	(iPad mini)	€ 454,00	0,28
13511	item out of stock	2019-12-24 20:39:01 (991224)		(884803)	[]	(Wil van der Aalst)	(iPhone 8)	€ 529,00	0,21
13512	package delivered	2019-12-24 20:46:47 (991119,991030,991209,991254,991213,9		(884886,884020,884749,884930,884926,884925,884766,884927	(660784)	(Kefang Ding)	(iPad Air,Echo Dot,MacBook Pro,iPad Air,Kindle Paperwhite,iPad Air,iPad	€ 6.829,99	4719,00
13513	place order	2019-12-24 22:08:21 (991271)		(885000,885001,885002,885003)	[]	(Mohammadreza Fani Sani)	(Fire Stick 4K,iPad,Kindle Paperwhite,iPad Air)	€ 1.194,99	1698,00
13514	pick item	2019-12-24 22:12:24 (991239)		(884874)	[]	(Tobias Brockhoff)	(Fire Stick)	€ 39,99	0,20
13515	place order	2019-12-25 08:08:00 (991272)		(885004,885005)	[]	(Mohammadreza Fani Sani)	(iPhone 11 Pro,iPad Pro)	€ 2.253,00	0,67
13516	confirm order	2019-12-25 08:12:59 (991272)		(885004,885005)	[]	(Mohammadreza Fani Sani)	(iPhone 11 Pro,iPad Pro)	€ 2.253,00	0,67
13517	pay order	2019-12-25 09:13:08 (991244)		(884890,884891,884892)	[]	(Junxiong Gao)	(Kindle,Fire Stick 4K,Echo Plus)	€ 324,97	2043,00
13518	package delivered	2019-12-25 09:17:36 (991229,991130,991060,991226,991236)		(884824,884429,884173,884810,884859)	(660788)	(Mahsa Bafrani)	(Kindle,Echo Show 8,iPad Pro,Echo,MacBook Pro)	€ 3.908,97	4096,00

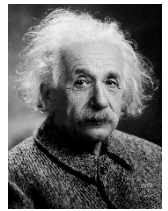
One of the 21887 events:
- **activity**: package delivered
- **time**: 2019-12-24 20:46:47
- **orders involved**: {991119,991030,991209,991254, 991213,991206}
- **items involved**: {884386,884020,884749,884930, 884926,884925,884766,884927,884736}
- **packages involved**: {660784}
- **customers involved**: {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





object-centric Petri net

Everything should be made as simple as possible, but no simpler.



PM4PY

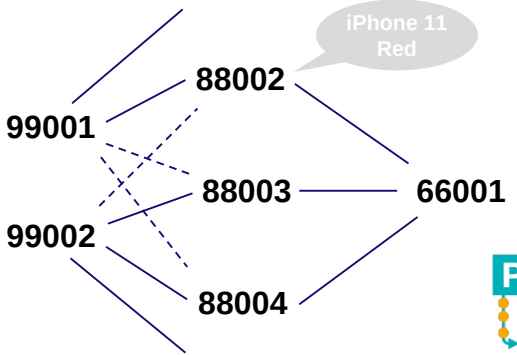
Chair of Process and Data Science

© Wil van der Aalst (use only with permission & acknowledgments)

The small print

- **Many design choices when creating object-centric process models: types of nodes/arcs, concurrency, etc.**
- **When to add objects to events? (indirect/changing/partial relations).**
- **Objects are more than identifiers (object attributes).**

activity	time	orders	items	packages
send package	2020-6-25	{99001, 99002}	{88002, 88003, 88004}	{66001}



© Wil van der Aalst (use only with permission & acknowledgments)

Some pointers

Alessandro Berti

Anahita Farhang Ghahfarokhi

- W.M.P. van der Aalst. Object-Centric Process Mining: Dealing With Divergence and Convergence in Event Data. In Software Engineering and Formal Methods (SEFM 2019), volume 11724 of Lecture Notes in Computer Science, pages 3–25. Springer-Verlag, Berlin, 2019.
- W.M.P. van der Aalst and A. Berti. Discovering Object-Centric Petri Nets. Fundamenta Informaticae, 2020.
- A. Berti and W.M.P. van der Aalst. Discovering Multiple Viewpoint Models from Relational Databases. In International Symposium on Data-driven Process Discovery and Analysis, volume 379 of Lecture Notes in Business Information Processing, pages 24–51. Springer-Verlag, Berlin, 2020.

- W.M.P. van der Aalst, A. Artale, M. Montali, and S. Tritini. Object-Centric Behavioral Constraints: Integrating Data and Declarative Process Modelling. In Proceedings of the 30th International Workshop on Description Logics (DL 2017), volume 1879 of CEUR Workshop Proceedings. CEUR-WS.org, 2017.
- D. Fahland. Describing Behavior of Processes with Many-to-Many Interactions. In S. Donatelli and S. Haar, editors, Applications and Theory of Petri Nets 2019, volume 11522 of Lecture Notes in Computer Science, pages 3–24. Springer-Verlag, Berlin, 2019.
- D. Fahland, M. De Leoni, B. van Dongen, and W.M.P. van der Aalst. Many-to-Many: Some Observations on Interactions in Artifact Choreographies. In Proceedings of the 3rd Central-European Workshop on Services and their Composition (ZEUS 2011), pages 9–15. CEUR-WS.org, 2011.
- G. Li, R. Medeiros de Carvalho, and W.M.P. van der Aalst. Automatic Discovery of Object-Centric Behavioral Constraint Models. In Business Information Systems (BIS 2017), LNBIP 288, 2017.

Chair of Process and Data Science

© Wil van der Aalst (use only with permission & acknowledgments)

Incorporating stochastics

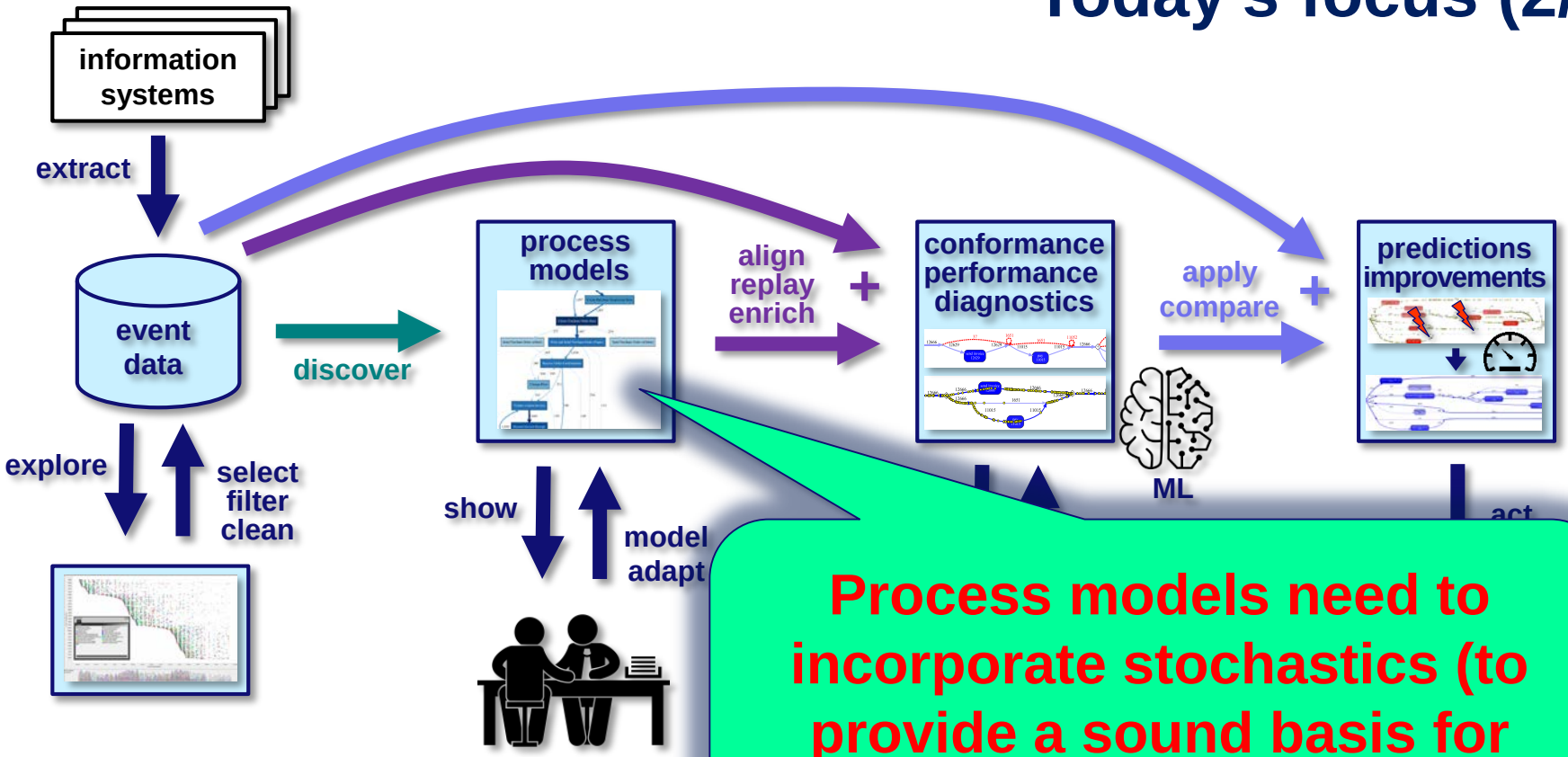
“precision is ill-defined as long as models have no probablities”

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

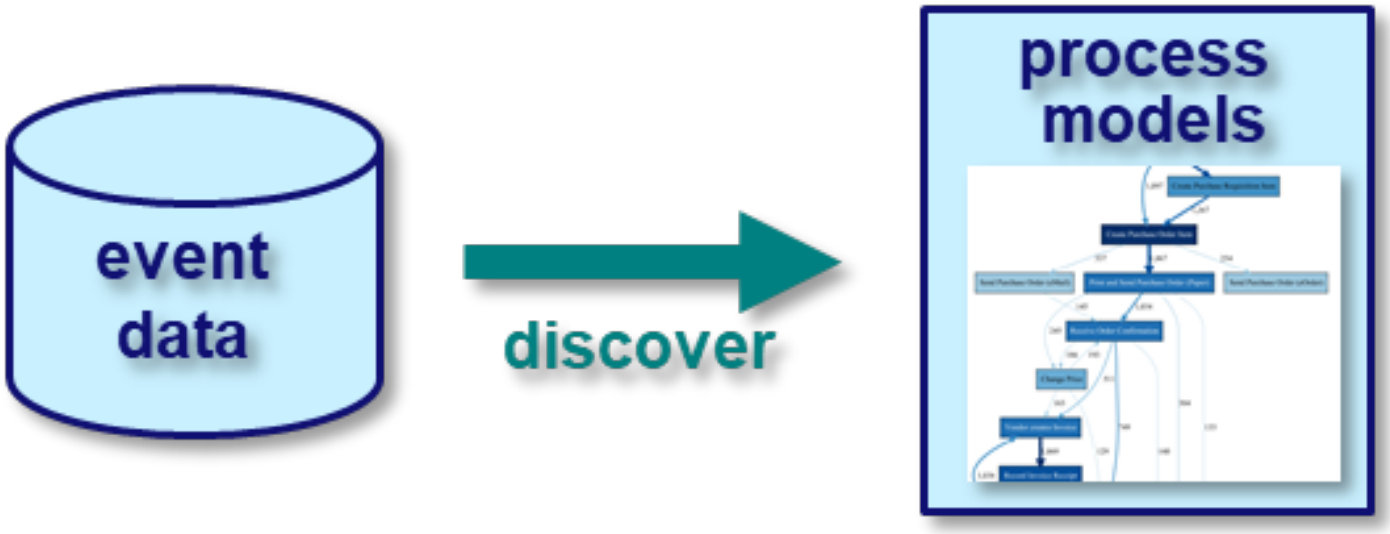
© Wil van der Aalst (use only with permission & acknowledgements)

Today's focus (2/2)



© Wil van der Aalst (use only with permission & acknowledgments)

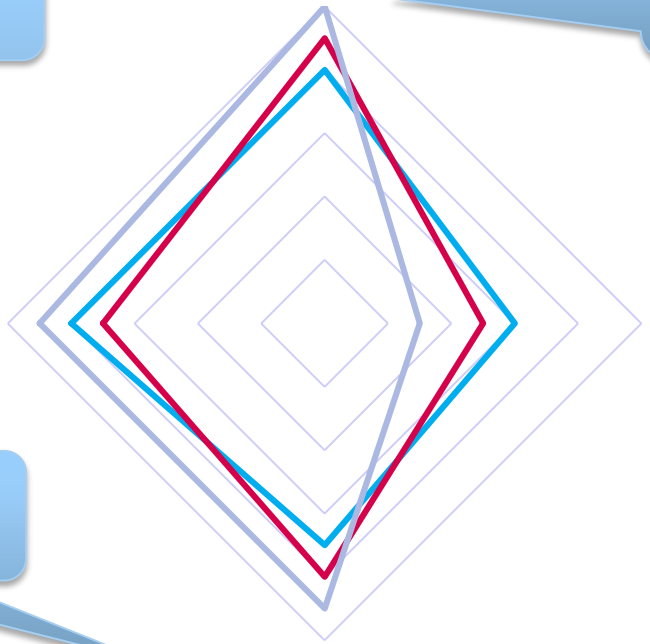
Fundamental question



How good is it?

© Wil van der Aalst (use only with permission & acknowledgments)

Traditional four quality dimensions



Does the event log show behavior not possible in the model ?

unrelated to behavior

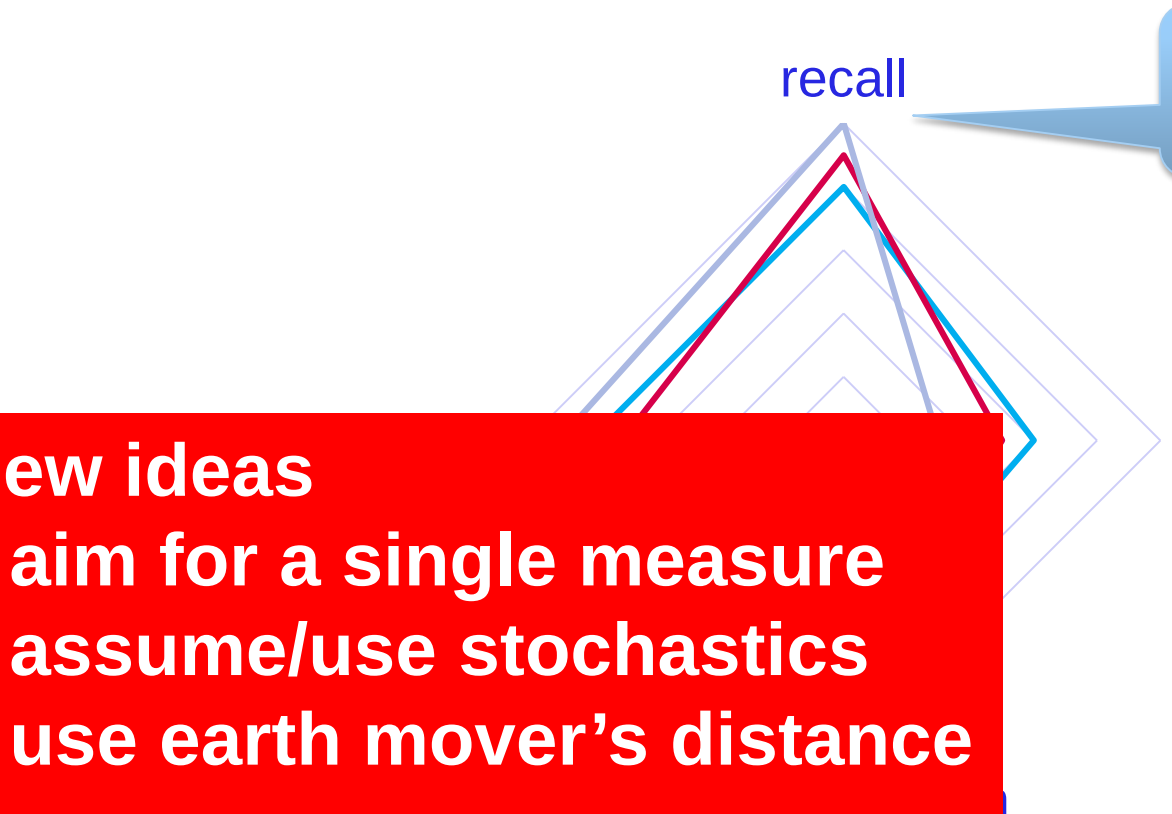
Does the model allow for behavior not in the event log ?

focus on future observations

© Wil van der Aalst (use only with permission & acknowledgements)

Chair of Process and Data Science

Focus on log – model behavior



Does the event log show behavior not possible in the model ?

Does the model allow for behavior not in the event log ?

New ideas

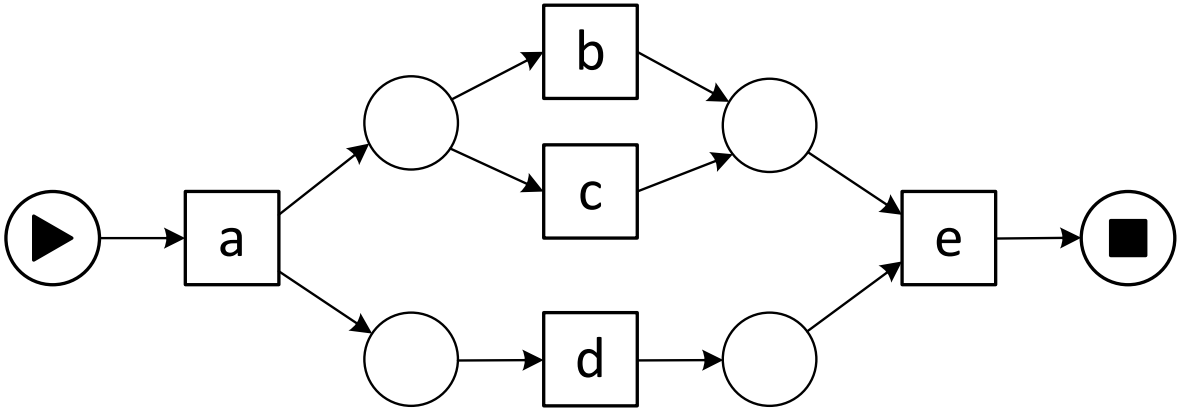
- aim for a single measure
- assume/use stochastics
- use earth mover's distance

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

© Wil van der Aalst (use only with permission & acknowledgements)

Rate conformance



$$L_1 = [\langle a, b, d, e\rangle^{490}, \langle a, d, b, e\rangle^{490}, \langle a, c, d, e\rangle^{10}, \langle a, d, c, e\rangle^{10}]$$

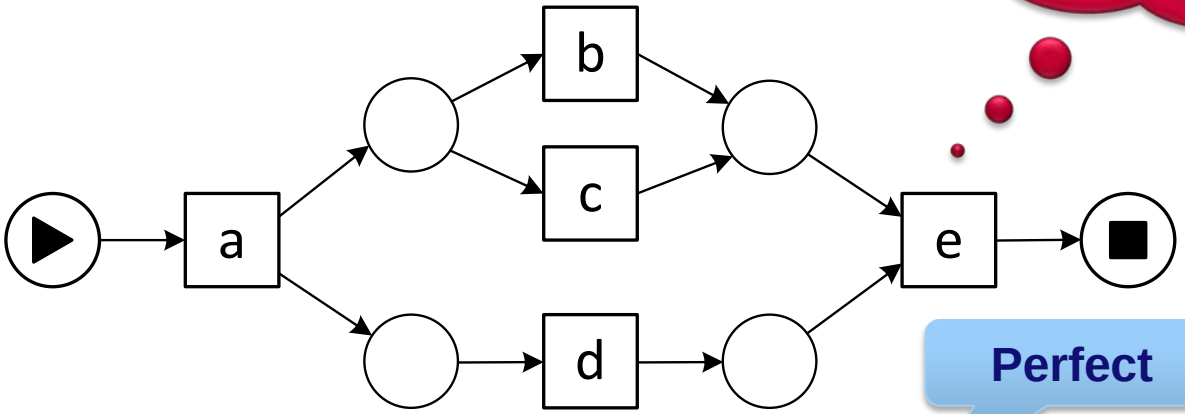
$$L_2 = [\langle a, b, d, e\rangle^{245}, \langle a, d, b, e\rangle^{245}, \langle a, c, d, e\rangle^{5}, \langle a, d, c, e\rangle^{5}, \langle a, b, e\rangle^{500}]$$

$$L_3 = [\langle a, b, d, e\rangle^{489}, \langle a, d, b, e\rangle^{489}, \langle a, c, d, e\rangle^{10}, \langle a, d, c, e\rangle^{10}, \langle a, b, e\rangle^{2}]$$

© Wil van der Aalst (use only with permission & acknowledgements)

Rate conformance

Frequencies matter, why not probabilities?



Perfect

$$L_1 = [\langle a,b,d,e\rangle^{490}, \langle a,d,b,e\rangle^{490}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}]$$

Terrible

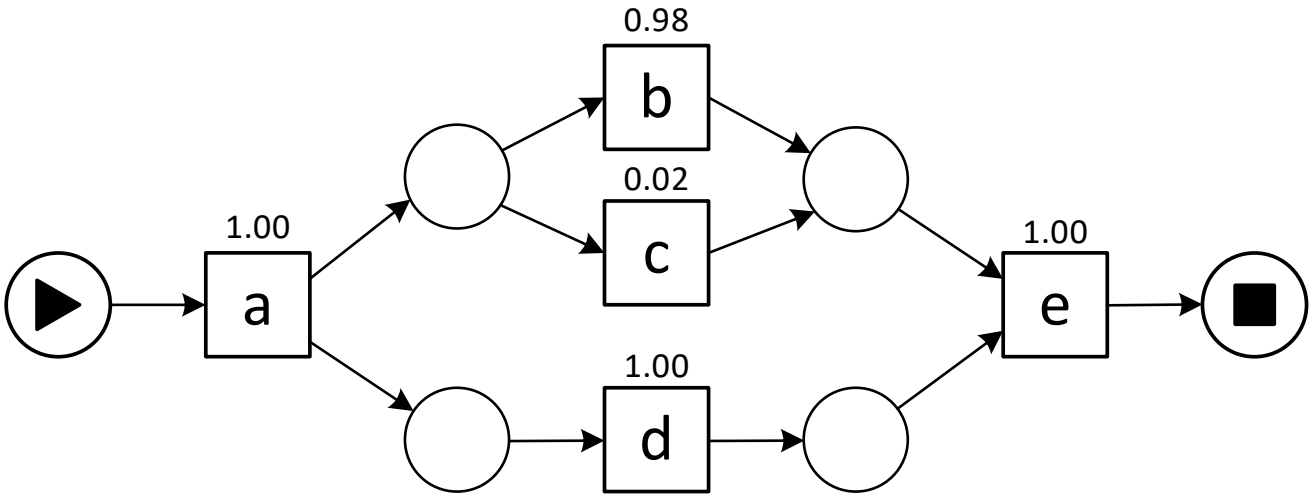
$$L_2 = [\langle a,b,d,e\rangle^{245}, \langle a,d,b,e\rangle^{245}, \langle a,c,d,e\rangle^{5}, \langle a,d,c,e\rangle^{5}, \langle a,b,e\rangle^{500}]$$

Almost perfect

$$L_3 = [\langle a,b,d,e\rangle^{489}, \langle a,d,b,e\rangle^{489}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}, \langle a,b,e\rangle^{2}]$$

© Wil van der Aalst (use only with permission & acknowledgements)

Assume the model describes a stochastic language



$$M = \left[\langle a, b, d, e\rangle^{0.49}, \langle a, d, b, e\rangle^{0.49}, \langle a, c, d, e\rangle^{0.01}, \langle a, d, c, e\rangle^{0.01}\right]$$

© Wil van der Aalst (use only with permission & acknowledgements)

Same idea can be applied to event logs

$$L_1 = [\langle a,b,d,e\rangle^{490}, \langle a,d,b,e\rangle^{490}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}]$$

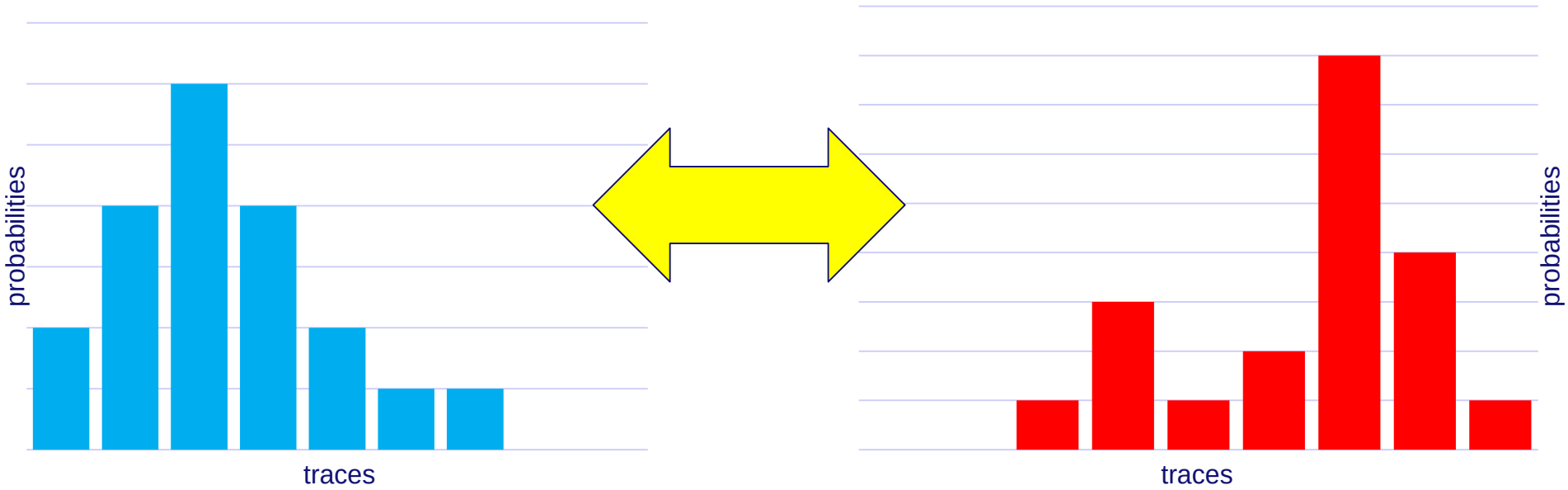
$$L_1 = [\langle a,b,d,e\rangle^{0.49}, \langle a,d,b,e\rangle^{0.49}, \langle a,c,d,e\rangle^{0.01}, \langle a,d,c,e\rangle^{0.01}]$$

$$L_2 = [\langle a,b,d,e\rangle^{245}, \langle a,d,b,e\rangle^{245}, \langle a,c,d,e\rangle^{5}, \langle a,d,c,e\rangle^{5}, \langle a,b,e\rangle^{500}]$$

$$L_2 = [\langle a,b,d,e\rangle^{0.245}, \langle a,d,b,e\rangle^{0.245}, \langle a,c,d,e\rangle^{0.005}, \langle a,d,c,e\rangle^{0.005}, \langle a,b,e\rangle^{0.5}]$$

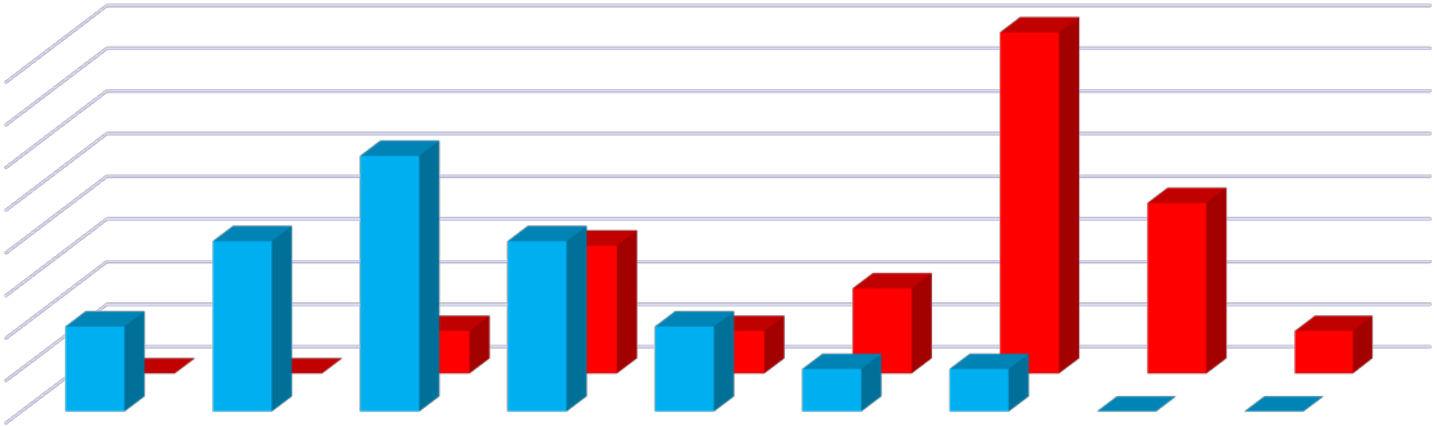
© Wil van der Aalst (use only with permission & acknowledgements)

Problem is now “simple”: Comparing stochastic languages

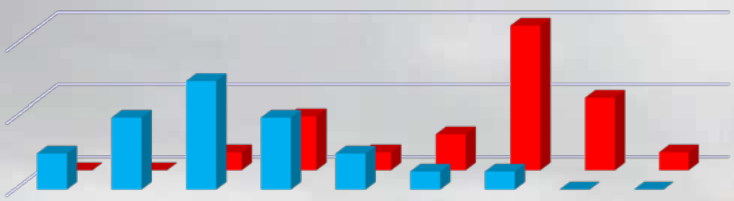


© Wil van der Aalst (use only with permission & acknowledgements)

Earth Mover's Distance (EMD)

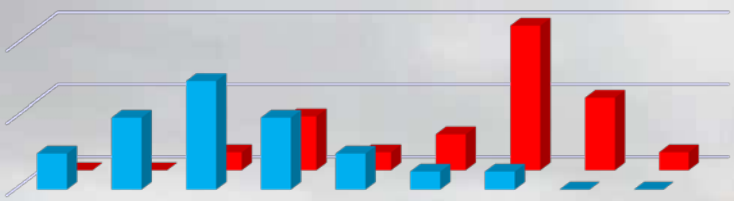


© Wil van der Aalst (use only with permission & acknowledgements)



$$M = \left[\langle a, b, d, e\rangle^{0.49}, \langle a, d, b, e\rangle^{0.49}, \langle a, c, d, e\rangle^{0.01}, \langle a, d, c, e\rangle^{0.01}\right]$$

$$L_2 = \left[\langle a, b, d, e\rangle^{0.245}, \langle a, d, b, e\rangle^{0.245}, \langle a, c, d, e\rangle^{0.005}, \langle a, d, c, e\rangle^{0.005}, \langle a, b, e\rangle^{0.5}\right]$$



Different distance notions are considered. Here we assume the normalized edit distance.

How much do I need to move ?

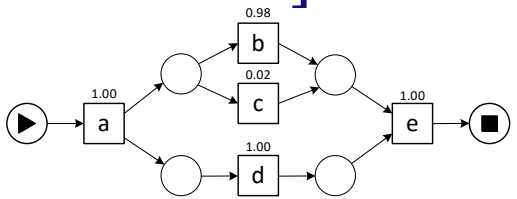
How far do I need to move it ?

Minimize

$$cost(\sigma, L, M) = \sum_{\sigma \in \tilde{L}} \sum_{\sigma' \in \tilde{M}} reallocated(\sigma, \sigma')\, distance(\sigma, \sigma')$$

Example: One model and five event logs

$$M = \left[\langle a,b,d,e\rangle^{0.49}, \langle a,d,b,e\rangle^{0.49}, \langle a,c,d,e\rangle^{0.01}, \langle a,d,c,e\rangle^{0.01}\right]$$



$$L_1 = \left[\langle a,b,d,e\rangle^{490}, \langle a,d,b,e\rangle^{490}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}\right]$$

$$L_2 = \left[\langle a,b,d,e\rangle^{245}, \langle a,d,b,e\rangle^{245}, \langle a,c,d,e\rangle^{5}, \langle a,d,c,e\rangle^{5}, \langle a,b,e\rangle^{500}\right]$$

$$L_3 = \left[\langle a,b,d,e\rangle^{489}, \langle a,d,b,e\rangle^{489}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}, \langle a,b,e\rangle^{2}\right]$$

$$L_4 = \left[\langle a,b,d,e\rangle^{500}, \langle a,d,b,e\rangle^{500}\right]$$

$$L_5 = \left[\langle a,c,d,e\rangle^{500}, \langle a,d,c,e\rangle^{500}\right]$$

© Wil van der Aalst (use only with permission & acknowledgements)

Consider the last event log $L_5 = [\langle a,c,d,e\rangle^{500}, \langle a,d,c,e\rangle^{500}]$

$$L_5 = [\langle a,c,d,e\rangle^{0.5}, \langle a,d,c,e\rangle^{0.5}]$$

0.49 0.01 0.49 0.01

$$M = [\langle a,b,d,e\rangle^{0.49}, \langle a,d,b,e\rangle^{0.49}, \langle a,c,d,e\rangle^{0.01}, \langle a,d,c,e\rangle^{0.01}]$$

$cost(\sigma, L_5, M)$	$\langle a,b,d,e\rangle^{0.49}$	$\langle a,d,b,e\rangle^{0.49}$	$\langle a,c,d,e\rangle^{0.01}$	$\langle a,d,c,e\rangle^{0.01}$
$\langle a,c,d,e\rangle^{0.5}$	$0.49 \cdot {}^{1}/_{4}$	$0 \cdot {}^{2}/_{4}$	$0.01 \cdot {}^{0}/_{4}$	$0 \cdot {}^{2}/_{4}$
$\langle a,d,c,e\rangle^{0.5}$	$0 \cdot {}^{2}/_{4}$	$0.49 \cdot {}^{1}/_{4}$	$0 \cdot {}^{2}/_{4}$	$0.01 \cdot {}^{0}/_{4}$

$$cost(\sigma, L_5, M) = {}^{980}/_{1000}\ {}^{1}/_{4} = {}^{49}/_{200} = 0.245$$

© Wil van der Aalst (use only with permission & acknowledgements)

Example: One model and five event logs

$$M = [\langle a,b,d,e\rangle^{0.49}, \langle a,d,b,e\rangle^{0.49}, \langle a,c,d,e\rangle^{0.01}, \langle a,d,c,e\rangle^{0.01}]$$

recall: b is approx. 50x more likely than c

$$L_1 = [\langle a,b,d,e\rangle^{490}, \langle a,d,b,e\rangle^{490}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}]$$

$$L_2 = [\langle a,b,d,e\rangle^{245}, \langle a,d,b,e\rangle^{245}, \langle a,c,d,e\rangle^{5}, \langle a,d,c,e\rangle^{5}, \langle a,b,e\rangle^{500}]$$

$$L_3 = [\langle a,b,d,e\rangle^{489}, \langle a,d,b,e\rangle^{489}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}, \langle a,b,e\rangle^{2}]$$

$$L_4 = [\langle a,b,d,e\rangle^{500}, \langle a,d,b,e\rangle^{500}]$$

$$L_5 = [\langle a,c,d,e\rangle^{500}, \langle a,d,c,e\rangle^{500}]$$

$$cost(\sigma, L_1, M) = 0$$

$$cost(\sigma, L_2, M) = {}^{500}/_{1000}\ {}^{1}/_{4} = {}^{1}/_{8} = 0.125$$

$$cost(\sigma, L_3, M) = {}^{2}/_{1000}\ {}^{1}/_{4} = {}^{1}/_{2000} = 0.0005$$

$$cost(\sigma, L_4, M) = {}^{20}/_{1000}\ {}^{1}/_{4} = {}^{1}/_{200} = 0.005$$

$$cost(\sigma, L_5, M) = {}^{980}/_{1000}\ {}^{1}/_{4} = {}^{49}/_{200} = 0.245$$

© Wil van der Aalst (use only with permission & acknowledgements)

Ordering the logs from good to bad

$$M = [\langle a,b,d,e\rangle^{0.49}, \langle a,d,b,e\rangle^{0.49}, \langle a,c,d,e\rangle^{0.01}, \langle a,d,c,e\rangle^{0.01}]$$

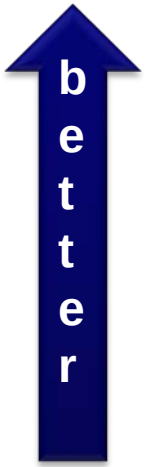
$$L_1 = [\langle a,b,d,e\rangle^{490}, \langle a,d,b,e\rangle^{490}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}]$$
$cost(\sigma, L_1, M) = 0$

$$L_3 = [\langle a,b,d,e\rangle^{489}, \langle a,d,b,e\rangle^{489}, \langle a,c,d,e\rangle^{10}, \langle a,d,c,e\rangle^{10}, \langle a,b,e\rangle^{2}]$$
very small recall problem
$cost(\sigma, L_3, M) = 0.0005$

$$L_4 = [\langle a,b,d,e\rangle^{500}, \langle a,d,b,e\rangle^{500}]$$ $cost(\sigma, L_4, M) = 0.005$
minor precision problem: model allows for two very unlikely paths that were not observed

$$L_2 = [\langle a,b,d,e\rangle^{245}, \langle a,d,b,e\rangle^{245}, \langle a,c,d,e\rangle^{5}, \langle a,d,c,e\rangle^{5}, \langle a,b,e\rangle^{500}]$$
major recall problem: half of the traces are not fitting the model
$cost(\sigma, L_2, M) = 0.125$

$$L_5 = [\langle a,c,d,e\rangle^{500}, \langle a,d,c,e\rangle^{500}]$$ $cost(\sigma, L_5, M) = 0.245$
major precision problem: model allows for two very likely paths that were not observed



© Wil van der Aalst (use only with permission & acknowledgements)

Next steps (joint work with Sander Leemans, Tobias Brockhoff, Artem Polyvyanyy, and Seran Uysal)

- Approach solves many of the known problems related to conformance checking (e.g., precision).
- Why consider frequencies in logs and not probabilities in models?
- Challenges currently being addressed:
  - Mapping existing modeling onto our representation for stochastic languages (improved path vs trace)
  - Providing diagnostics.
  - Concept drift
  - Comparative process mining

© Wil van der Aalst (use only with permission & acknowledgements)

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

© Wil van der Aalst (use only with permission & acknowledgements)

How about ontologies?

- Diego Calvanese, Tahir Emre Kalayci, Marco Montali, Ario Santoso, Wil M. P. van der Aalst: Conceptual Schema Transformation in Ontology-Based Data Access. EKAW 2018: 50-67
- Diego Calvanese, Marco Montali, Alifah Syamsiyah, Wil M. P. van der Aalst: Ontology-Driven Extraction of Event Logs from Relational Databases. Business Process Management Workshops 2015: 140-153
- Ana Karla Alves de Medeiros, Wil M. P. van der Aalst: Process Mining Towards Semantics. Advances in Web Semantics I 2009: 35-80
- Ana Karla Alves de Medeiros, Wil M. P. van der Aalst, Carlos Pedrinaci: Semantic Process Mining Tools: Core Building Blocks. ECIS 2008: 1953-1964
- Ana Karla Alves de Medeiros, Carlos Pedrinaci, Wil M. P. van der Aalst, John Domingue, Minseok Song, Anne Rozinat, Barry Norton, Liliana Cabral: An Outlook on Semantic Business Process Mining and Monitoring. OTM Workshops (2) 2007: 1244-1255

When it is there we use it!

ontop

Chair of Process and Data Science

© Wil van der Aalst (use only with permission & acknowledgments)

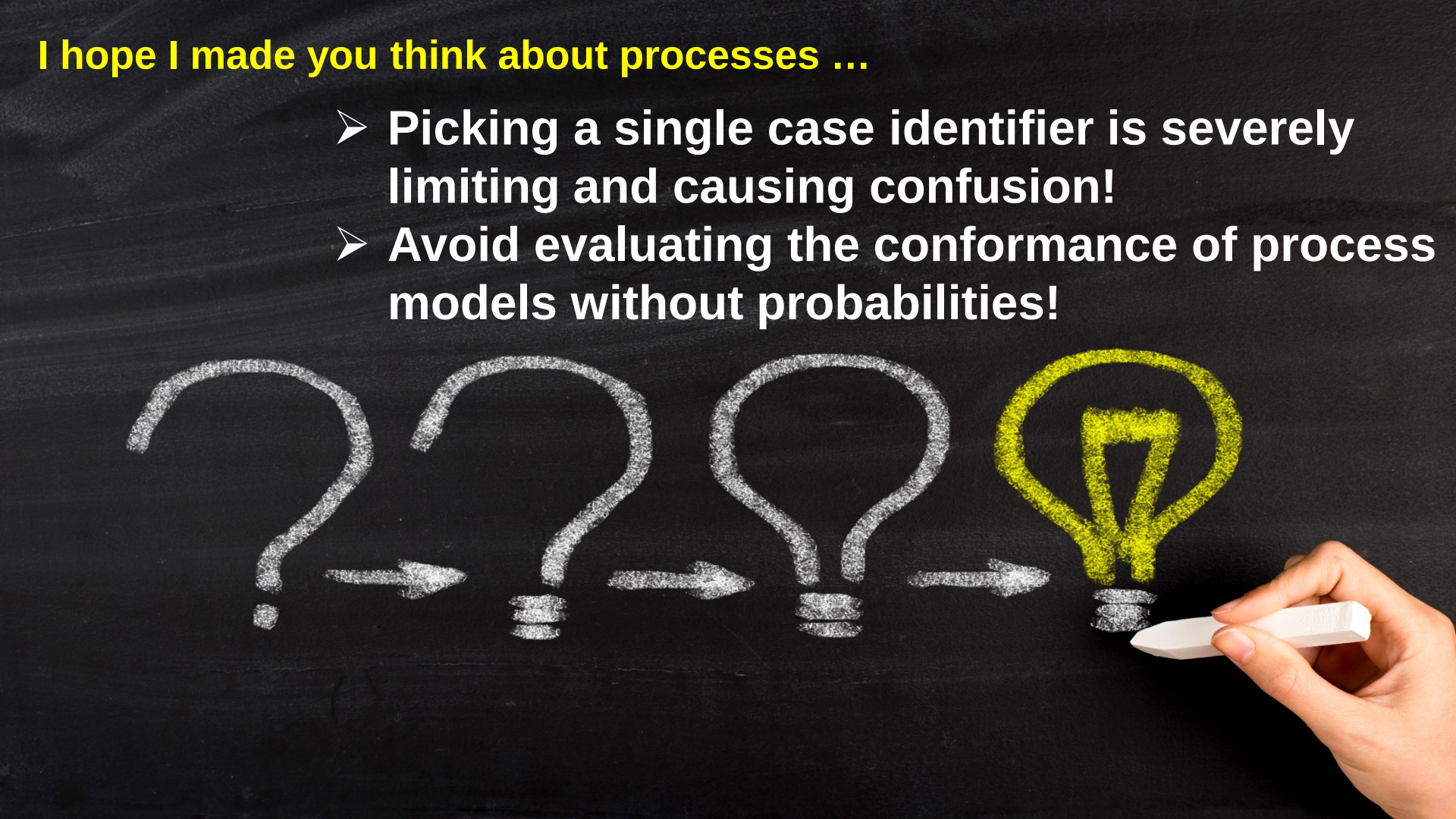
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

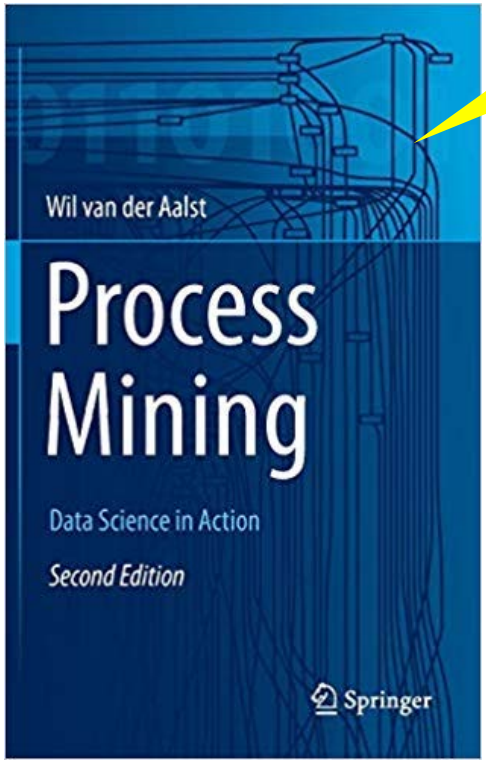
© Wil van der Aalst (use only with permission & acknowledgements)

I hope I made you think about processes …

- Picking a single case identifier is severely limiting and causing confusion!
- Avoid evaluating the conformance of process models without probabilities!



Learn more?

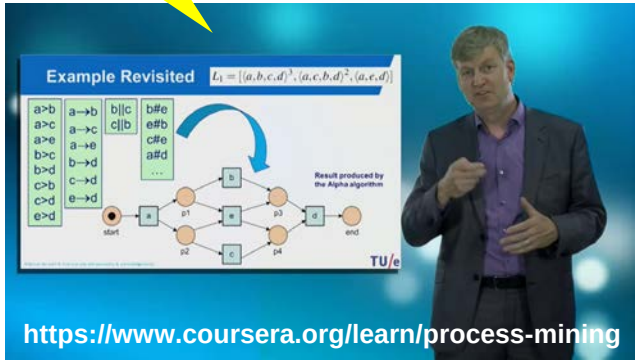


“PM Bible”

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

Over 125.000 participants

coursera



Fraunhofer FIT

Chair of Process and Data Science

© Wil van der Aalst (use only with permission & acknowledgments)