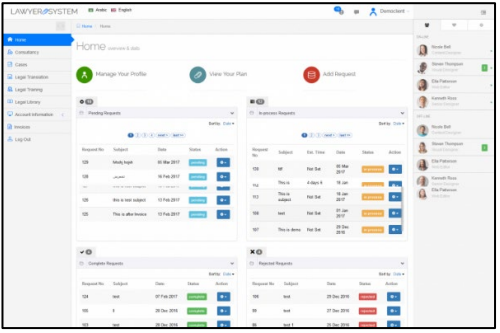
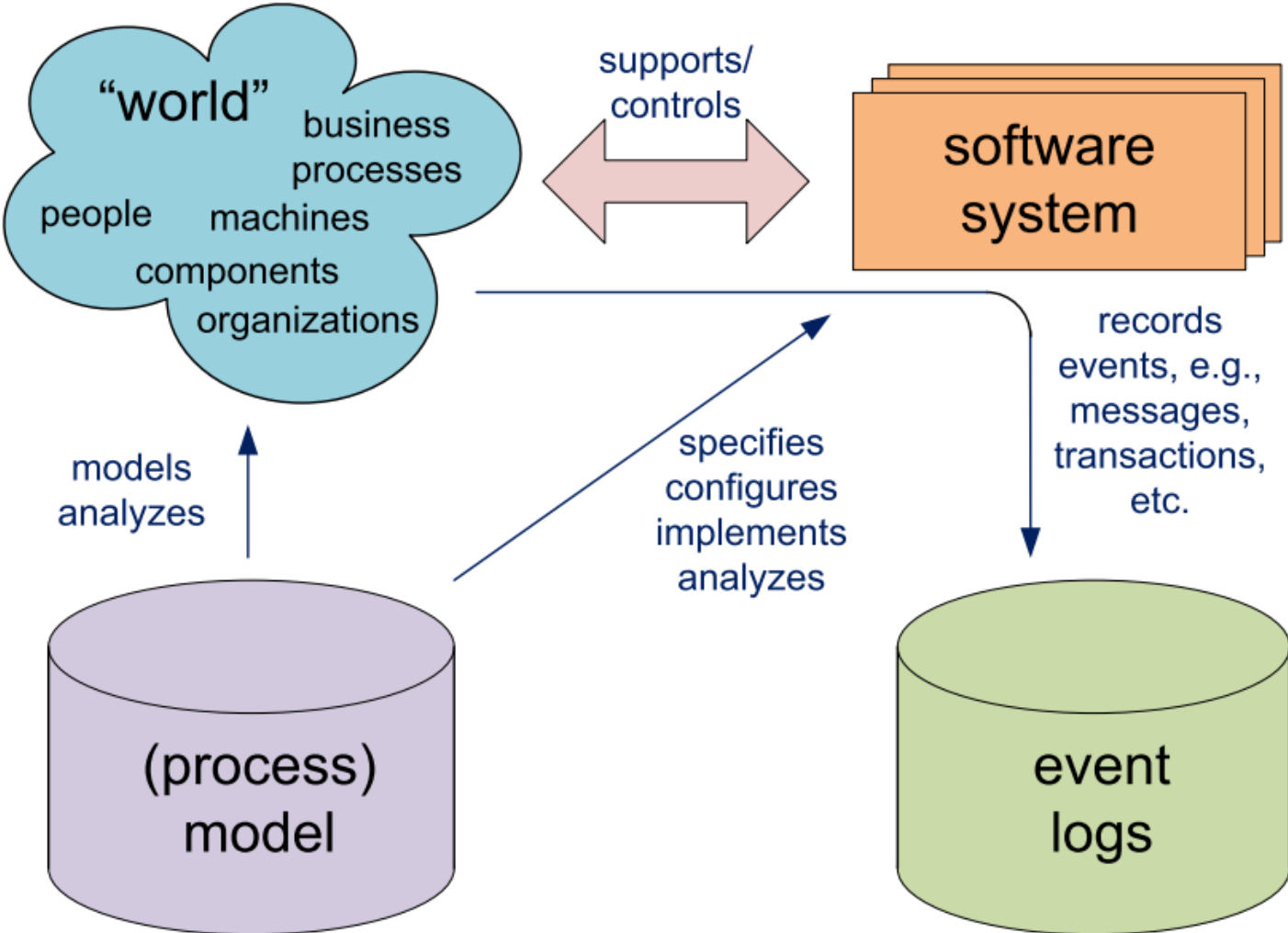


# Streaming Process Mining: State of the Art and Research Directions

Andrea Burattin

[andbur@dtu.dk](mailto:andbur@dtu.dk)

# Process mining overview

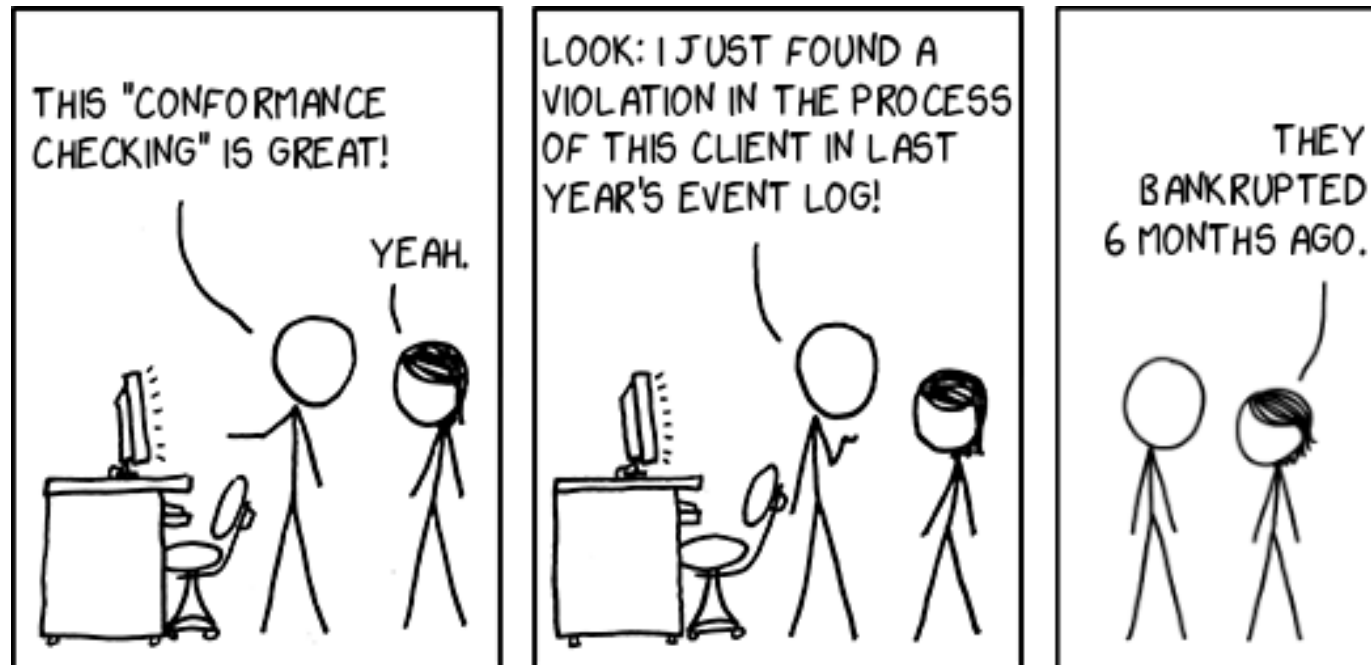


Date	Project	Client	Project Name	Project Manager	Status	Value	Region	Category
25-Apr-2018	MC	Parsons	Parsons Taylor	Strategy for Parsons Global Research	Completed	\$1,000,000	North America	IT
25-Apr-2018	PS	Goodyear	Goodyear	As a Goodyear product line promotion	Phase 1/2	\$1,000,000	NA	Marketing
25-Apr-2018	PS	Goodyear	Goodyear	Product Label 10 year anniversary promotion	Phase 1/2	\$1,000,000	NA	Marketing
25-Apr-2018	PS	Goodyear	Goodyear	Walmart Label 10 year anniversary promotion	Phase 1/2	\$1,000,000	NA	Marketing
25-Apr-2018	PS	Goodyear	Goodyear	State Street City anniversary promotion	Phase 1/2	\$1,000,000	NA	Marketing
25-Apr-2018	PS	Goodyear	Goodyear	Municipal website migration project	Phase 1/2	\$1,000,000	NA	IT
25-Apr-2018	PS	Goodyear	Goodyear	Agway 10th anniversary	Phase 1/2	\$1,000,000	NA	Marketing
25-Apr-2018	SDG	Manoff	Gustafson, K	National Women's Centre	Completed	\$1,200,000	CA	IT
25-Apr-2018	ACSA	Bank	Patrick, G	Bank 100th Anniversary 100th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	F01	Routledge	Routledge	Funding for 10th Anniversary and 10th anniversary project	Completed	\$1,000,000	NA	Marketing
25-Apr-2018	EBD	Bank	Digital	Launch of 10th Anniversary 10th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	EBD	Duclos	Vanquish	Alaska Housing project - Resurrection Lane	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	INF	Changhong	Changhong	Agway 10th anniversary - Bank	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	INF	Changhong	Changhong	Agway 10th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	FEDCOR	Bank	Lafayette	Central Ontario 100th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	FEDCOR	Bank	Lafayette	Central Ontario 100th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	MC	Parsons	Parsons Taylor	Parsons Taylor	Completed	\$1,000,000	NA	IT
25-Apr-2018	PRGSC	Quintech	Quintech	Advisory Services - Phase 1/2 project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	ADT	Lumen	Lumen	New Chief Executive Office	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	EBD	Duclos	Vanquish	Alaska Housing project - Resurrection Lane	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	SDG	Manoff	Gustafson	Central Ontario 100th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	AMC	Illinois	Illinois	Illinois Housing Authority - Con design	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	ACSA	Bank	Patrick, G	Bank 100th Anniversary 100th anniversary project	Marketing	\$1,000,000	NA	Marketing
25-Apr-2018	TC	Germany	Germany	Bank 100th Anniversary 100th anniversary project	Marketing	\$1,000,000	NA	Marketing

Image source: Wil van der Aalst, “Process Mining - Data Science in Action”, Springer

# Motivation for streaming process mining

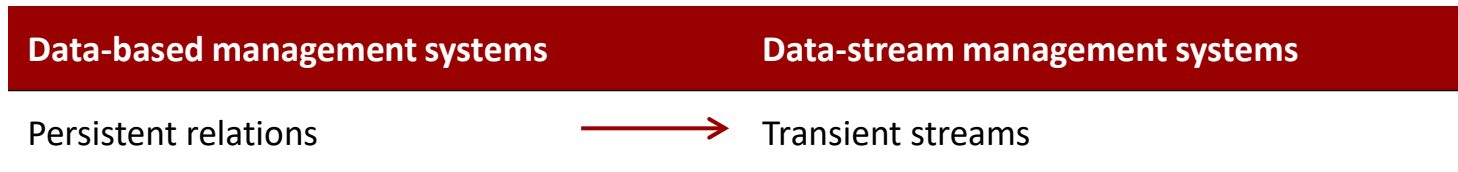
- Process mining on historical (i.e., past) data is called *post-mortem* or *offline*
- Post-mortem process mining is good for *forensic* reasons
- Post-mortem process mining produces results that are *outdated* by design



Not a real xkcd comic.

# Streaming data/process mining

- The input is a *stream*



Gama, J. (2010). *Knowledge Discovery from Data Streams*. Chapman and Hall/CRC.

- Peculiarities of the stream mining problem
  - Impossible to **store** the entire stream (*approximation*)
  - Unbounded **backtracking not feasible** over streams (algorithms required to make one pass on datas)
  - Quickly adapt the model to cope with **unusual data values** (*concept drifts*)

# Event log and event stream

Event #	Activity	Originator	Time	
<b>Case id: C1</b>				
1	A	U1	2021-08-01	...
2	B	U1	2021-08-02	...
3	C	U2	2021-08-03	...
4	E	U2	2021-08-04	...
<b>Case id: C2</b>				
1	A	U1	2021-08-02	...
2	B	U1	2021-08-03	...
3	D	U3	2021-08-04	...
4	E	U3	2021-08-05	...

Post-mortem scenario

Time	Case id	Activity	Originator	
2021-08-01	C1	A	U1	...
2021-08-02	C2	A	U1	...
2021-08-02	C1	B	U1	...
2021-08-03	C1	C	U2	...
2021-08-03	C2	B	U1	...
2021-08-04	C1	E	U2	...
2021-08-04	C2	D	U3	...
2021-08-05	C2	E	U3	...
...				

Streaming scenario

# How traces can appear in a streaming setting

- In streaming setting we have an incomplete view of the data
  - Based on our observation window + data



# Taxonomy for streaming process mining approaches

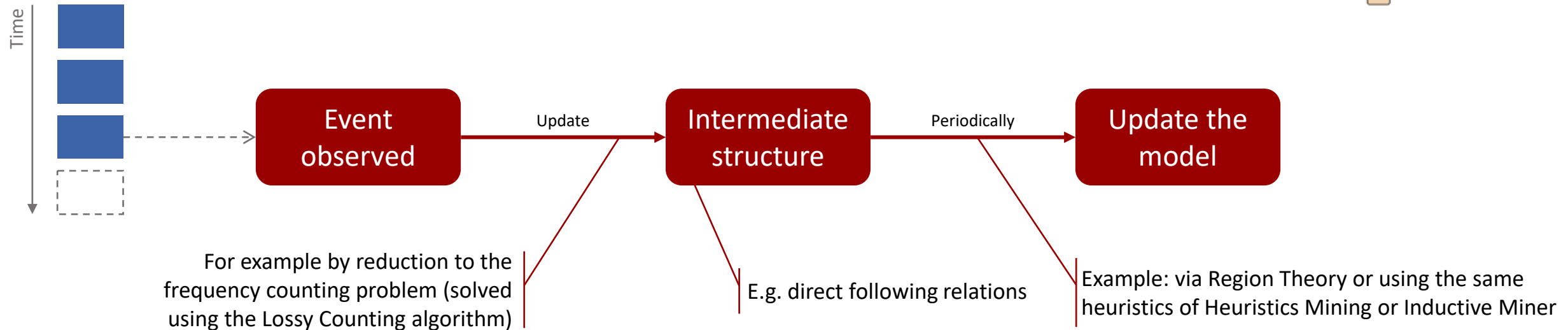
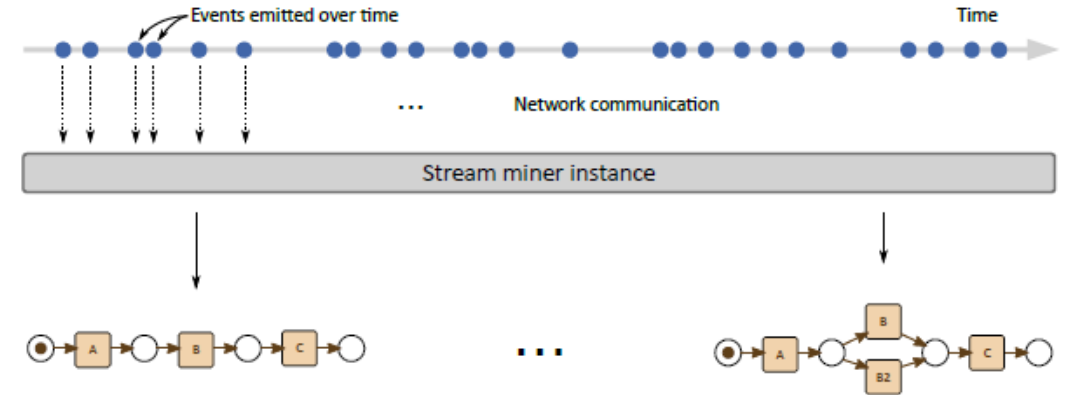
Stream process mining approaches taxonomy

Window-based models
<ol style="list-style-type: none"><li>1. Store a window of recent event</li><li>2. Apply any offline approach for any process mining problem</li></ol>

Burattin A. (2018) Streaming Process Discovery and Conformance Checking. In: Encyclopedia of Big Data Technologies. Springer

# Streaming process discovery

- Most approaches for discovery belong to “problem reduction” category
- Fairly good results (subj. to the domain)



A. Burattin, A. Sperduti, and W. M.P. van der Aalst. Control-flow Discovery from Event Streams. In proc. of the IEEE CEC, 2014.


S. J. van Zelst, B. van Dongen, and W. M.P. van der Aalst. Event stream-based process discovery using abstract representations. Knowledge and Information Systems, 2018.



# Streaming conformance checking

- On declarative models, under the name *operational support*, several works
  - Via replay, associate each constraint to: permanently or temporarily violated or fulfilled
  - Main focus on real timeness, not on data incompleteness

 F.M. Maggi, M. Montali, and W. van der Aalst. An operational decision support framework for monitoring business constraints. In proc. of FASE, 2012.

 F.M. Maggi, M. Montali, M. Westergaard, and W. M.P. van der Aalst. Monitoring Business Constraints with Linear Temporal Logic : An Approach Based on Colored Automata. In proc. of BPM, 2011.

- On imperative models

- Optimal alignments (i.e., exact solutions) can be calculated for *prefix traces*

 D. Schuster and S. J. van Zelst. Online Process Monitoring Using Incremental State-Space Expansion: An Exact Algorithm. In proc. of BPM, 2020.

 S. J. van Zelst, A. Bolt, M. Hassani, B. van Dongen, and W. M.P. van der Aalst. Online conformance checking: relating event streams to process models using prefix-alignments. Int. J. Data Sci. Anal., 2017.

Here we are not really in a streaming setting

# Streaming conformance checking (cont.)

- On imperative models

- On general streaming data some solutions are available

- 📄 A. Burattin and J. Carmona. A Framework for Online Conformance Checking. In proc. of BPI, 2017.

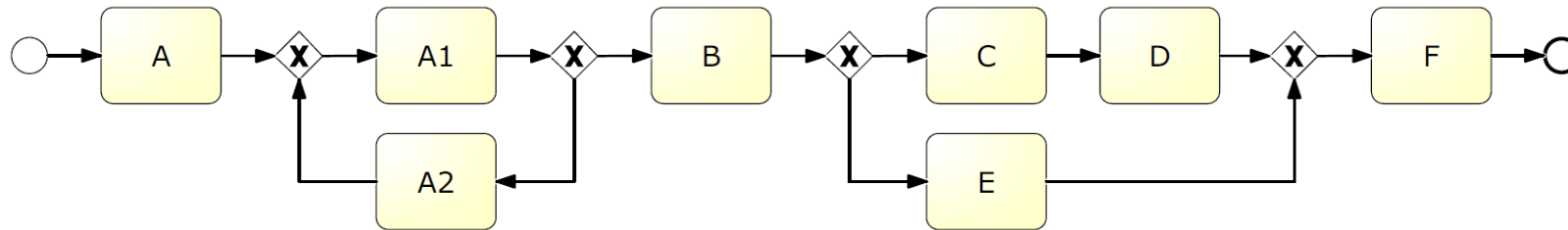
- 📄 A. Burattin, S. J. van Zelst, A. Armas-Cervantes, B. van Dongen, and J. Carmona. Online Conformance Checking Using Behavioural Patterns. In proc. of BPM, 2018.

- 📄 W.L.J. Lee, A. Burattin, J. Munoz-Gama, and M. Sepulveda. Orientation and conformance: A HMM-based approach to online conformance checking. Information Systems, 2020.

- Solutions based on “offline computation”: precompute structures from the reference model and then use them against the stream (to meet all computational requirements)
  - The problem is not solved though: it requires approximation, potentially domain dependent



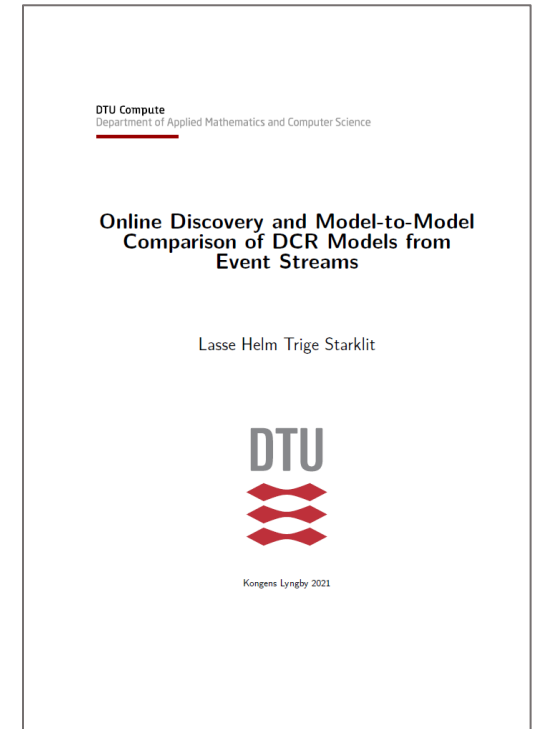
# Streaming conformance checking – Running example



Observation	Offline conformance
$\langle A, A1, B, E, F \rangle$	1.00
$\langle A, A1, A2, A1, B \rangle$	0.80
$\langle B, C, D, F \rangle$	0.78
$\langle B, C, D \rangle$	0.62

# Streaming process discovery - Ongoing research

- In the context of EcoKnow, together with Lasse Starklit (DTU) and Hugo Andrés López (KU)
- Algorithm for streaming discovery of DCR models
  1. Stream converted into augmented DFG
  2. DCR patterns identified on the augmented DFG
- Result of the mining compared against family of models
- Ultimate goal
  - Identify which “version” of a process (i.e., a law) is currently being executed



# Conclusions

- Streaming process mining allows the understanding of currently running processes
  - Opposed to post-mortem process mining, which is good for forensic/historical purposes
- Streaming process discovery is fairly well established
  - Several algorithms based on similar assumptions
  - Ongoing work on stream mining of DCR models
- Streaming conformance checking is challenging
  - Current solutions can cope well with the problem but just under specific circumstances (i.e., prefixes)
  - Streaming and exact solutions are theoretically impossible, effort on finding good balance