Christian-Albrechts-Universität zu Kiel

Process Mining for Sciences

Prof. Dr. Agnes Koschmider

Group Process Analytics, Kiel University, Germany

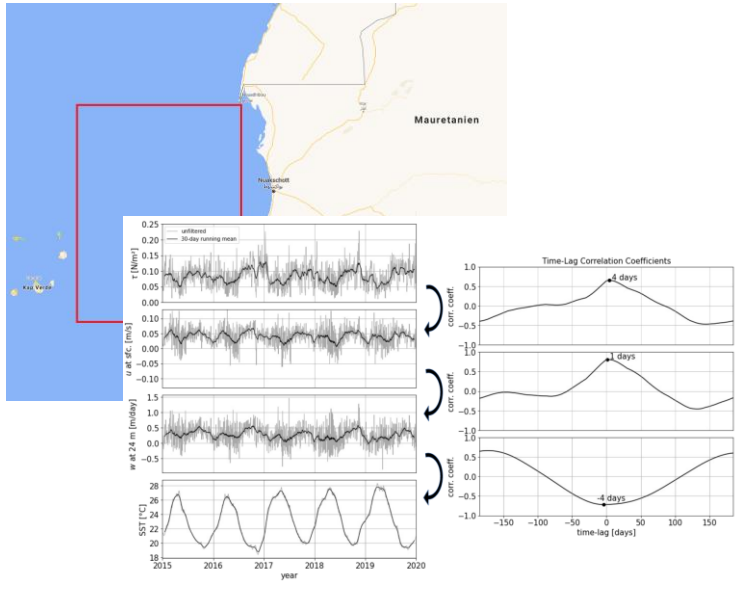
Prof. Dr. Agnes Koschmider

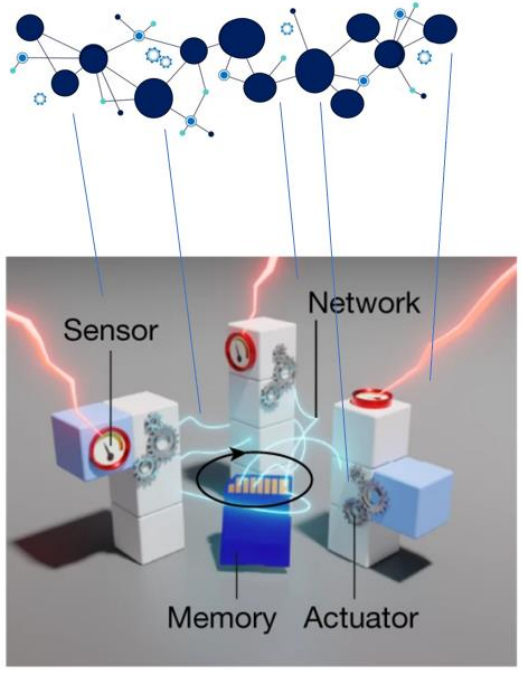
Since 05/2019:
Professor of Business Informatics (Process Analytics)
Computer Science Department
Kiel University
https://www.pa.informatik.uni-kiel.de/en

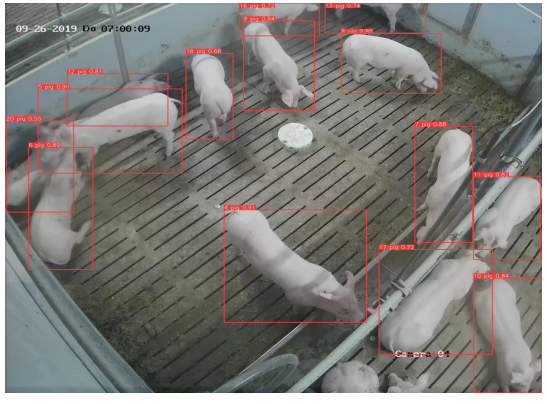
Education:
Habilitation, Applied Informatics, KIT
Promotion, Applied Informatics, KIT

TU/e
UNIVERSITAT POLITÈCNICA DE VALÈNCIA
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ
Almaden Institute
IBM

Research Questions



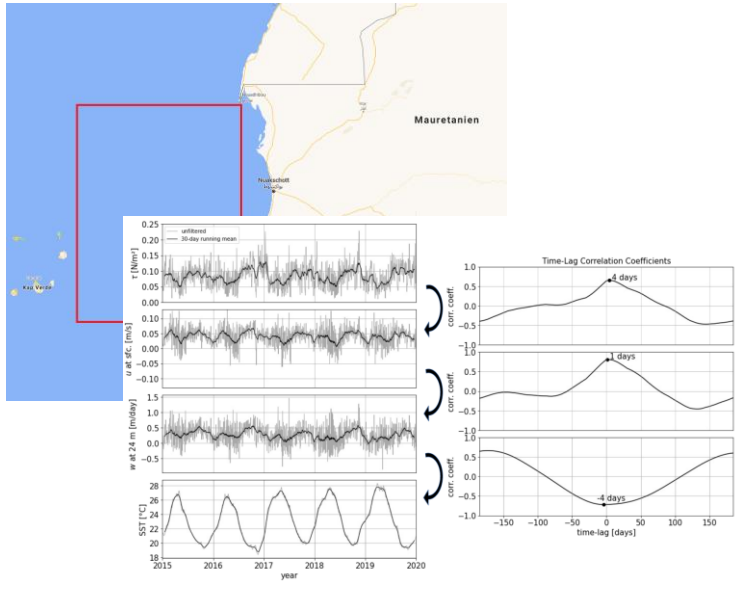




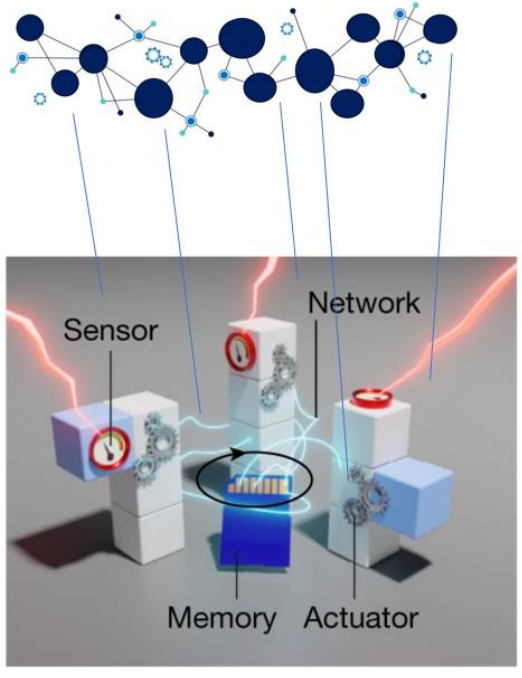
- what unknown processes are acting (i.e., did we found all processes that exist)?
- whether the found processes actually work as thought

Research Domains

Christian-Albrechts-Universität zu Kiel



time series data



sensor event data

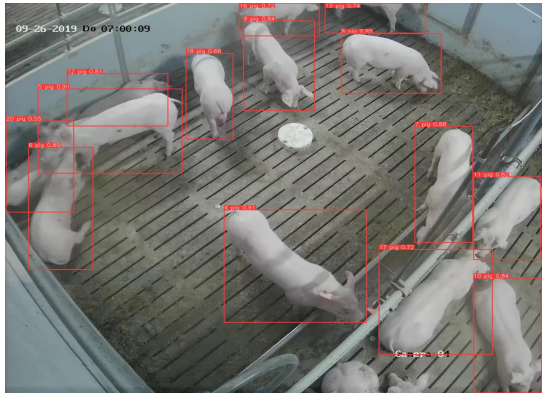
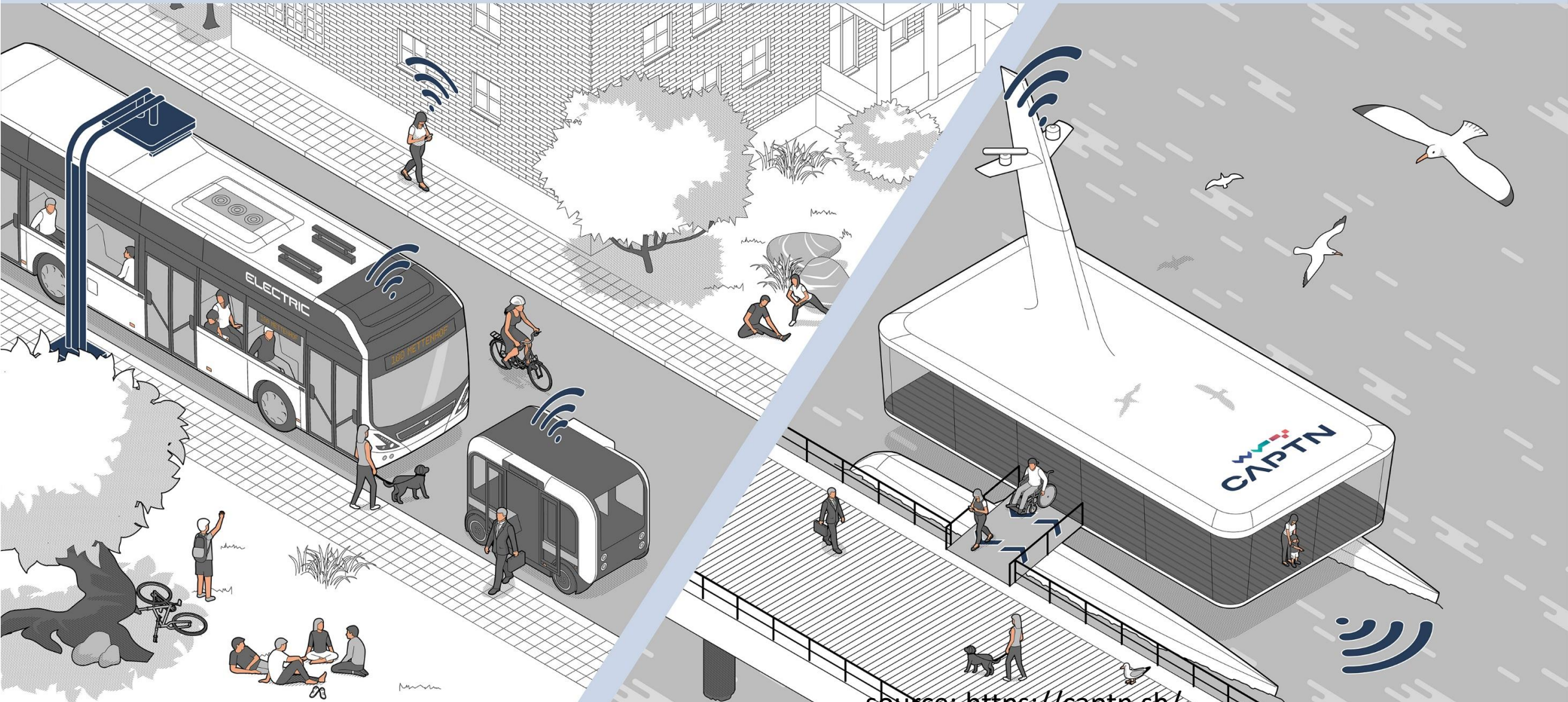


image data

Smart Applications: Integration of Techniques

Christian-Albrechts-Universität zu Kiel

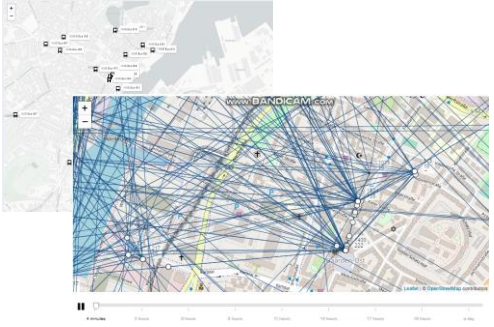


source: https://captn.ch/

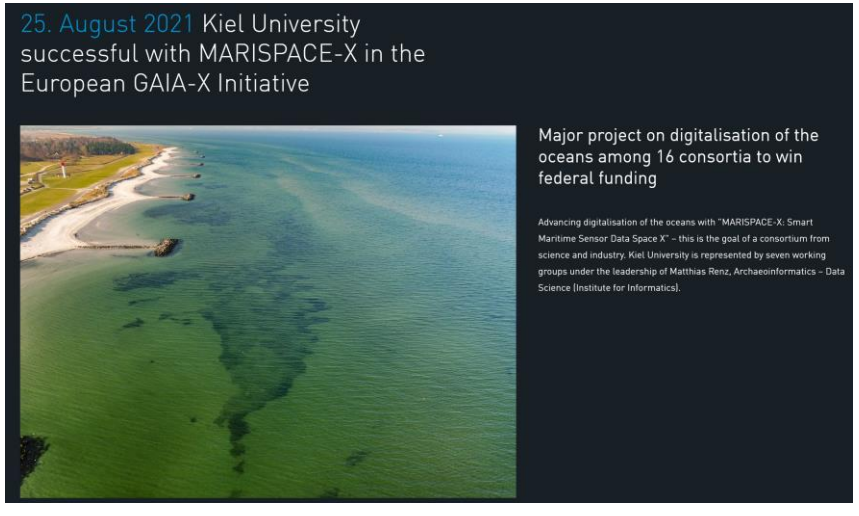
Further projects in which we are involved…

Christian-Albrechts-Universität zu Kiel

Quelle: https://captn.sh/



MARISPACE-X: Smart Maritime Sensor Data Space X



https://www.futureocean.org/en/cluster/aktuelles/meldungen/2021/2021_08_25_MARISPACE-X_erfolgreich.php

Further projects in which we are involved…

Christian-Albrechts-Universität zu Kiel

time series data

- predict the hazards of cyanobacteria with the purpose of exploratory visualization of diverse conditions
- predict the behavior of viromes

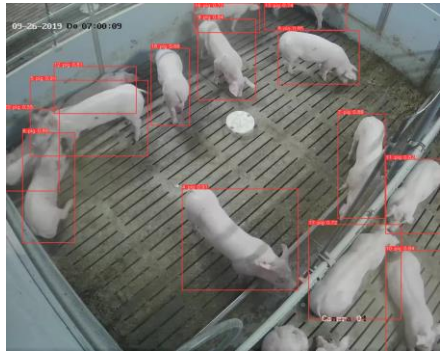
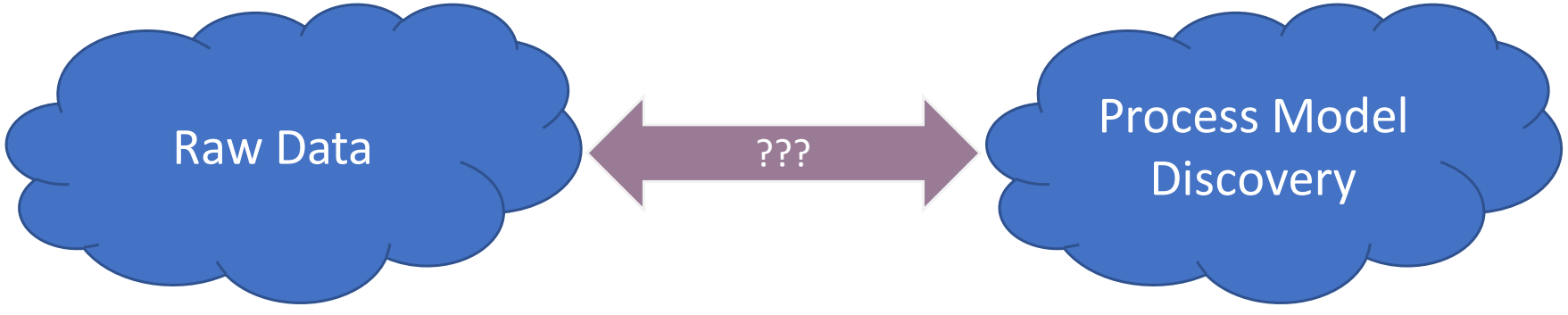


image data

- Process mining on image data to detect anomalies in the behavior of pigs

Challenge

- How to bridge the gap between raw data and process (model) discovery?



Characteristics of Business Processes

- activities are connected to each other via acting persons, machines, document flows, resources, etc.
- activities are executed by persons or by machines in a specific order to perform certain tasks.
- a business process might have *structured*, *weakly structured* and *unstructured* components (*sub-processes*).

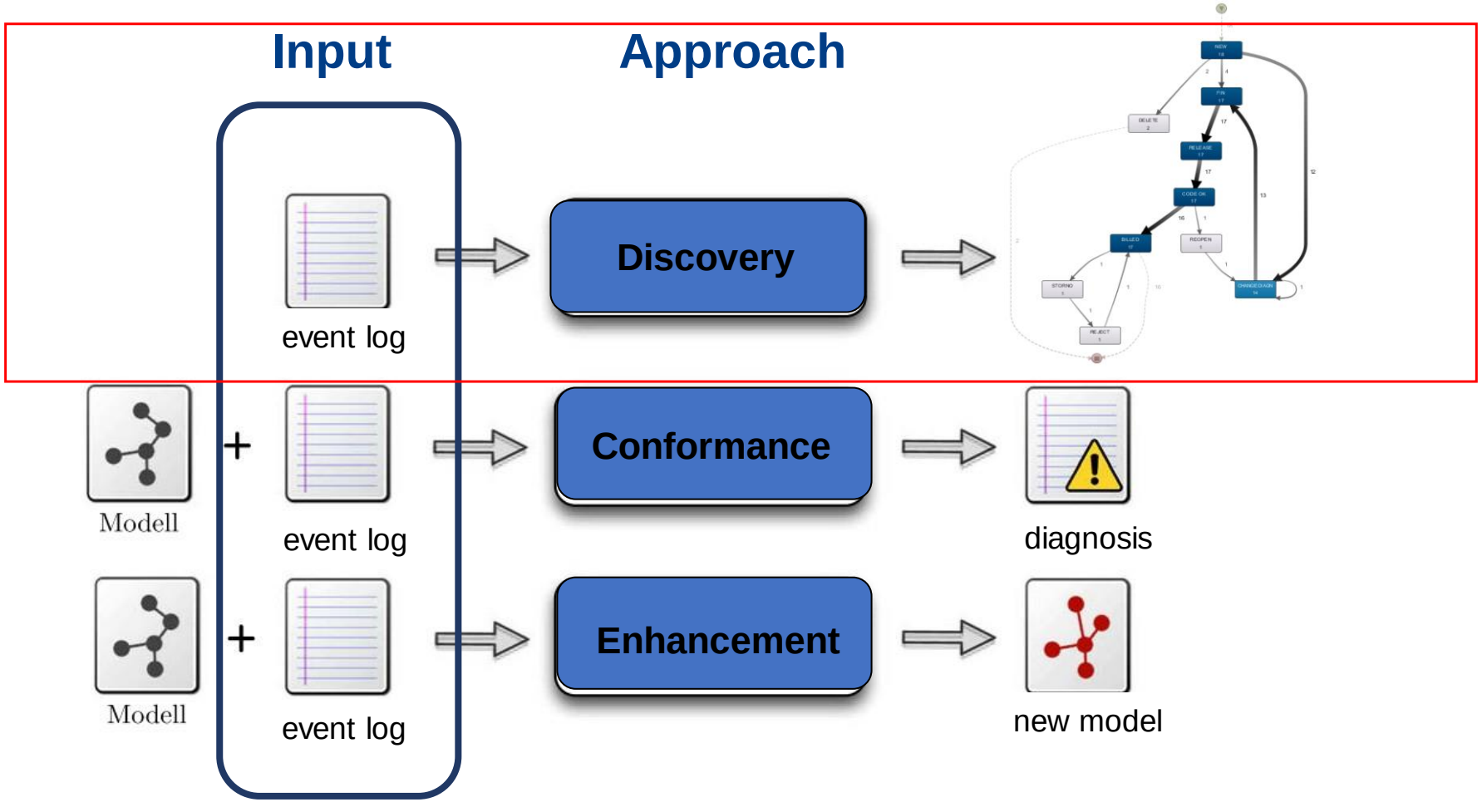
Process Model

Process model
(or process definition, process schema)

- describes the structure of a real business process
- specifies all possible paths along a business process
- specifies the rules for choosing a path
- specifies all activities that must be executed

A process model is a template.
Starting from there all process instances are initiated.

Short Recap: Process Mining

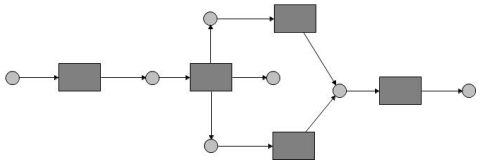


Process Mining – Event Log

Christian-Albrechts-Universität zu Kiel

- Case ID
- Activity
- Time stamp (start and/or end)
- Additional attributes (e.g., role, costs)

Fall ID	Ereignis ID	Eigenschaften				
		Zeitstempel	Aktivität	Rolle	Kosten	...
1	35654423	30-12-2010:11.02	Registriere Anfrage	Pete	50	...
	35654424	31-12-2010:10.06	Ausführliche Beurteilung	Sue	400	...
	35654425	05-01-2011:15.12	Überprüfe Ticket	Mike	100	...
	35654426	06-01-2011:11.18	Treffe Entscheidung	Sara	200	...
	35654427	07-01-2011:14.24	Lehne Antrag ab	Pete	200	...
2	35654483	30-12-2010:11.32	Registriere Anfrage	Mike	50	...
	35654485	30-12-2010:12.12	Überprüfe Ticket	Mike	100	...
	35654487	30-12-2010:14.16	Schnelle Beurteilung	Pete	400	...
	35654488	05-01-2011:11.22	Treffe Entscheidung	Sara	200	...
	35654489	08-01-2011:12.05	Zahle Kompensation	Ellen	200	...
3	35654521	30-12-2010:14.32	Registriere Anfrage	Pete	50	...
	35654522	30-12-2010:15.06	Schnelle Beurteilung	Mike	400	...
	35654524	30-12-2010:16.34	Überprüfe Ticket	Ellen	100	...
	35654525	06-01-2011:09.18	Treffe Entscheidung	Sara	200	...
	35654526	06-01-2011:12.18	Behandle Anfrage erneut	Sara	200	...
	35654527	06-01-2011:13.06	Ausführliche Beurteilung	Sean	400	...
	35654530	08-01-2011:11.43	Überprüfe Ticket	Pete	100	...
	35654531	09-01-2011:09.55	Treffe Entscheidung	Sara	200	...
	35654533	15-01-2011:10.45	Zahle Kompensation	Ellen	200	...



Process Mining – Event Log

Christian-Albrechts-Universität zu Kiel

```
<log xes.version="1.0" xmlns="http://code.deckfour.org/xes">
        <trace>
                <event>
                        <date key="time:timestamp" value="2010-12-30T11:02:00.000+01:00"/>
                        <string key="Activity" value="register request"/>
                        <string key="Resource" value="Pete"/>
                        <string key="Costs" value="50"/>
                </event>
                <event>
                        <date key="time:timestamp" value="2010-12-31T10:06:00.000+01:00"/>
                        <string key="Activity" value="examine thoroughly"/>
                        <string key="Resource" value="Sue"/>
                        <string key="Costs" value="400"/>
                </event>
                <event>
                        <date key="time:timestamp" value="2011-01-05T15:12:00.000+01:00"/>
                        <string key="Activity" value="check ticket"/>
                        <string key="Resource" value="Mike"/>
                        <string key="Costs" value="100"/>
                </event>
        </trace>
        <trace>
                <event>
                        <date key="time:timestamp" value="2011-01-06T15:02:00.000+01:00"/>
                        <string key="Activity" value="register request"/>
                        <string key="Resource" value="Pete"/>
                        <string key="Costs" value="50"/>
                </event>
                <event>
                        <date key="time:timestamp" value="2011-01-07T12:06:00.000+01:00"/>
                        <string key="Activity" value="check ticket"/>
                        <string key="Resource" value="Mike"/>
                        <string key="Costs" value="100"/>
                </event>
                <event>
                        <date key="time:timestamp" value="2011-01-08T14:43:00.000+01:00"/>
                        <string key="Activity" value="examine thoroughly"/>
                        <string key="Resource" value="Sean"/>
                        <string key="Costs" value="400"/>
                </event>
        </trace>

</log>
```

Log

Trace

Event

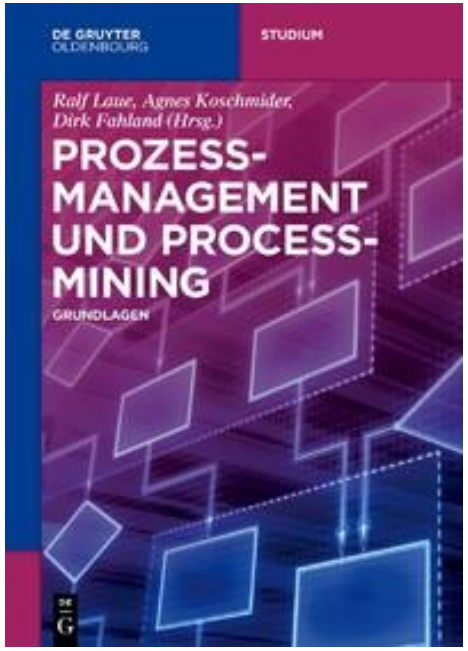
Trace

Event

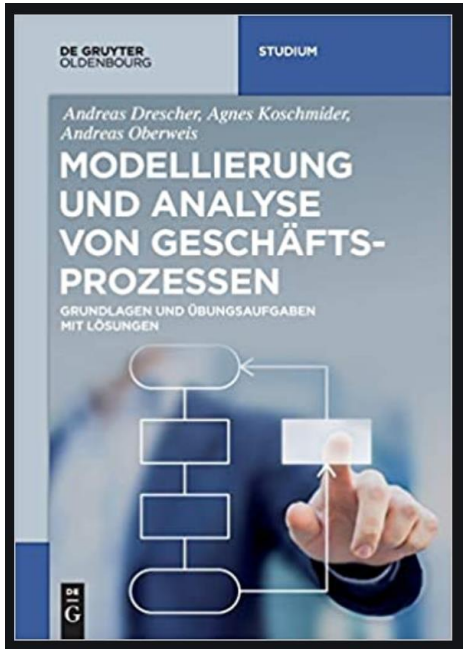
Event

Event

Literature



Laue, Koschmider, Fahland:
Prozessmanagement und
Process-Mining,
De Gruyter Oldenbourg (Verlag)
978-3-11-050015-8 (ISBN)

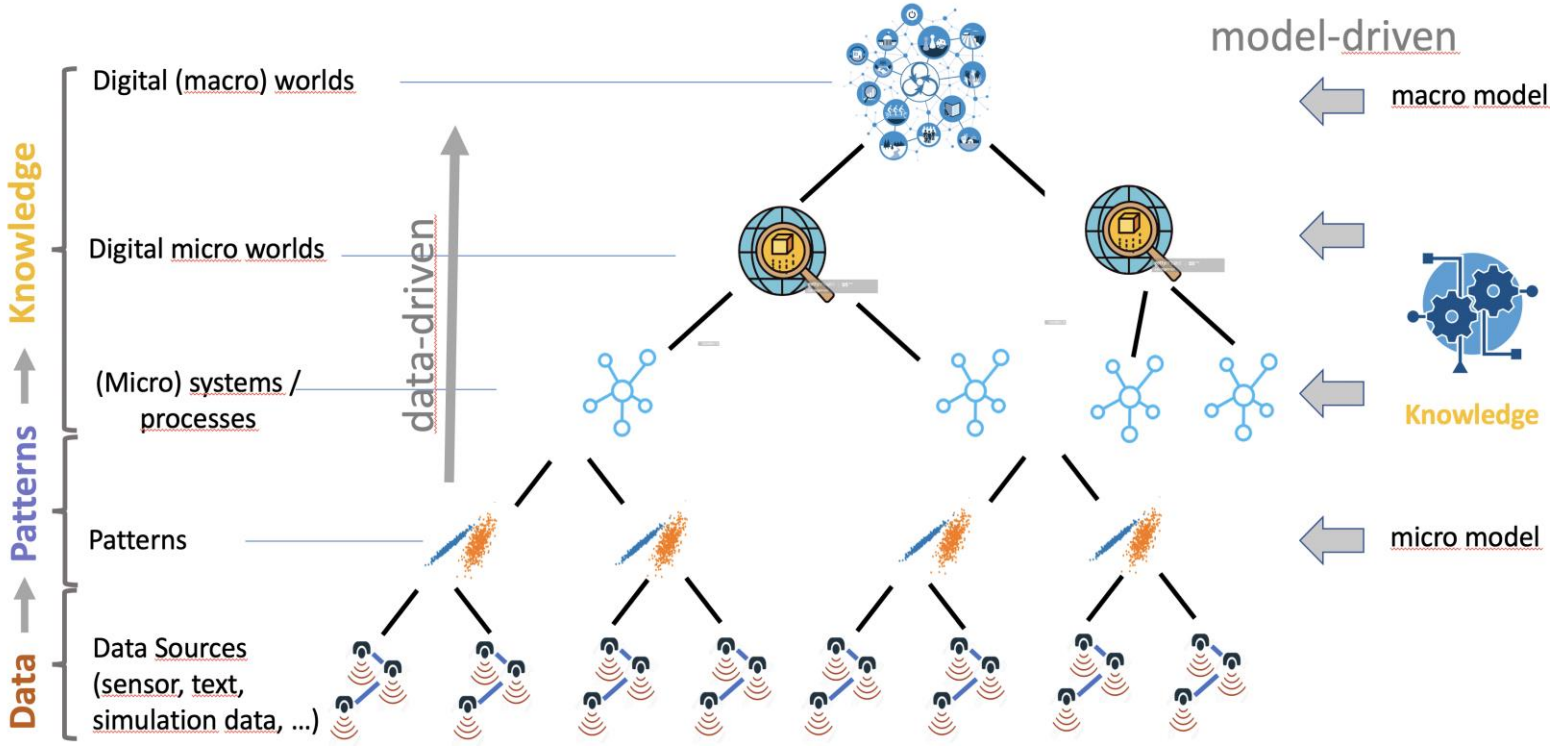


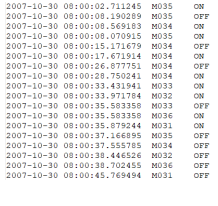
Drescher, Koschmider, Oberweis:
Modellierung und Analyse von
Geschäftsprozessen,
De Gruyter Oldenbourg (Verlag)
978-3-11-049449-5 (ISBN)

Data driven Identification of Process (Models)

Christian-Albrechts-Universität zu Kiel

- Focus
  - Data acquisition and sensor network methods (IoT)
  - Pattern mining
  - Process mining





Data-driven Process Analysis: Approach

Christian-Albrechts-Universität zu Kiel

Data Source

Raw Data

Data Aggregation

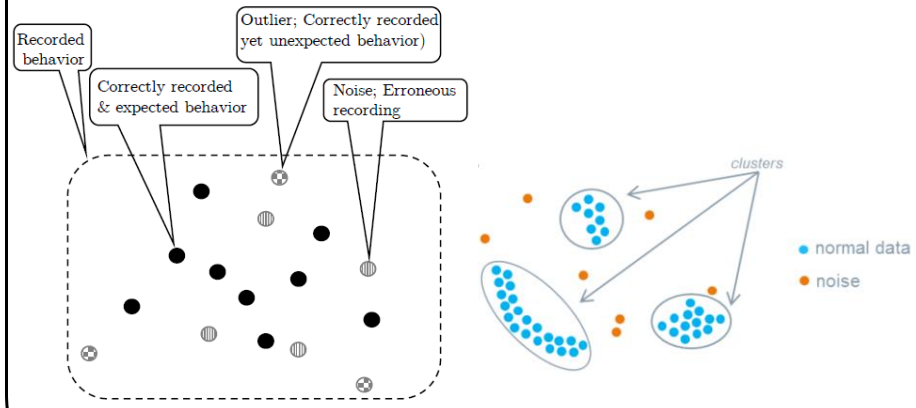
Event log

Process Model

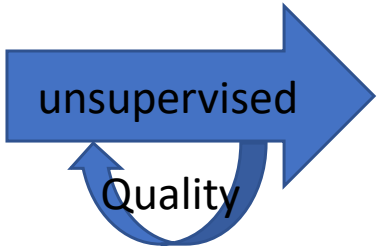
Three Areas of Research

Christian-Albrechts-Universität zu Kiel

Data Quality: Noise/Outlier



Human-in-the Loop



Synthetic Log Generator

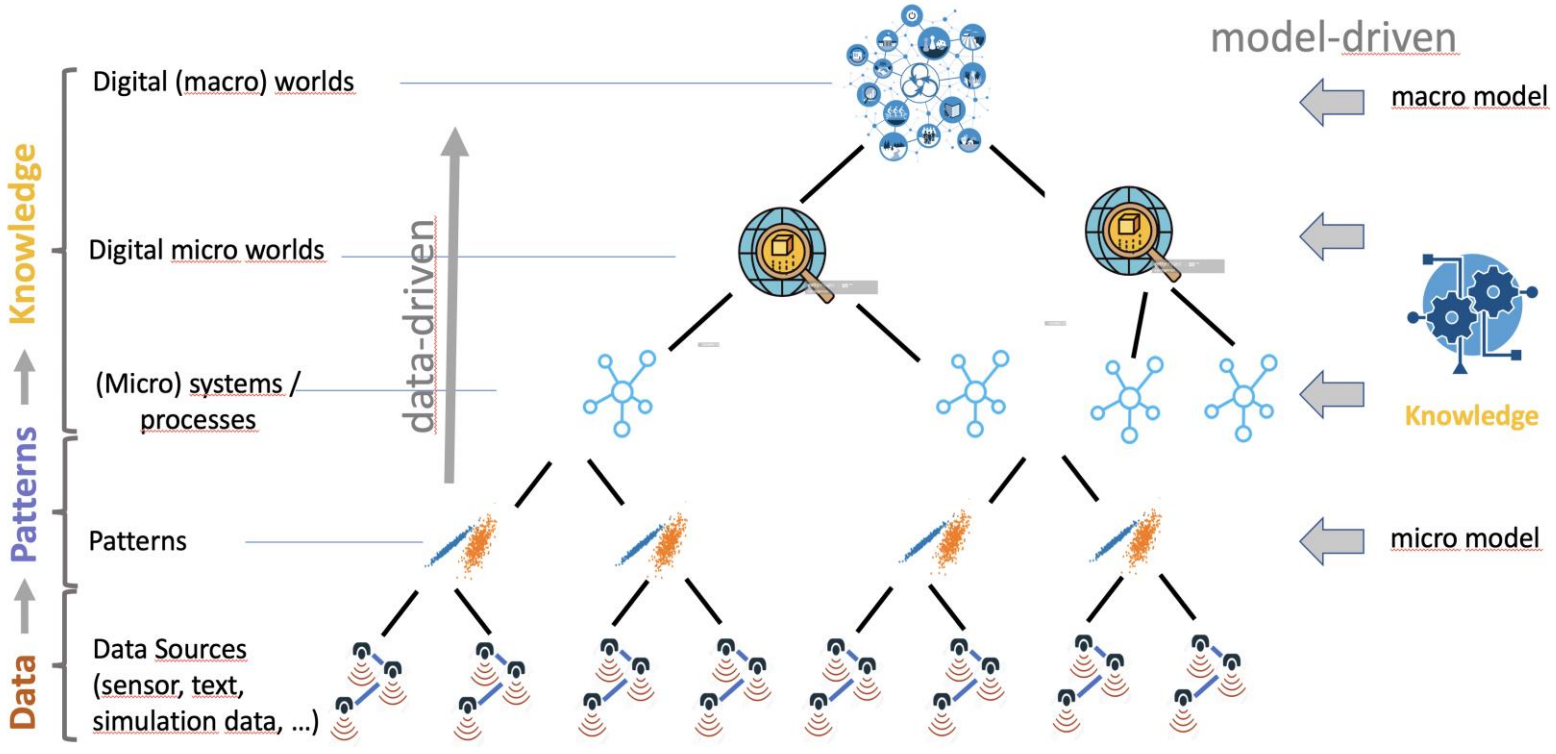
True or False ?

- A business process has only structured components ✗
- A process model describes the structure of a real business process ✓
- Process mining is restricted to the discovery of process models from business events ✗
- Human-in-the-loop requires human interaction ✓

Data driven Identification of Process (Models)

Christian-Albrechts-Universität zu Kiel

- Focus
  - Data acquisition and sensor network methods (IoT)
  - Pattern mining
  - Process mining

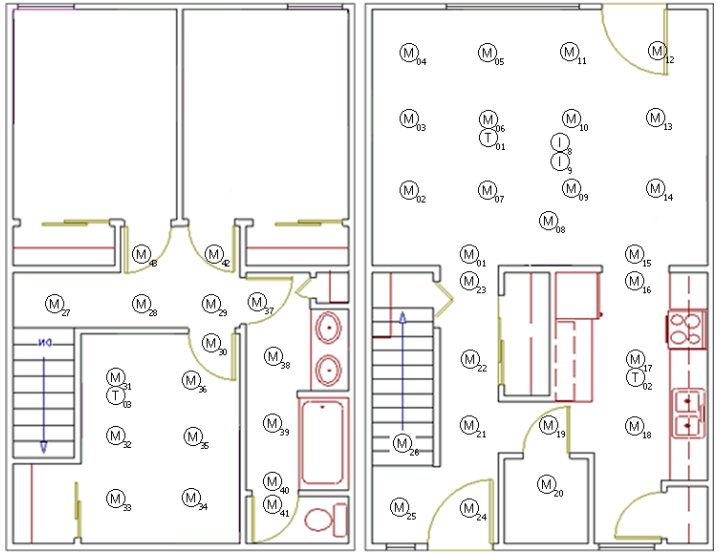


2007-10-30	08:00:02.711245	M035	ON
2007-10-30	08:00:08.190289	M035	OFF
2007-10-30	08:00:08.569183	M034	ON
2007-10-30	08:00:08.070915	M035	ON
2007-10-30	08:00:15.171679	M034	OFF
2007-10-30	08:00:17.671914	M034	ON
2007-10-30	08:00:26.877751	M034	OFF
2007-10-30	08:00:28.750241	M034	ON
2007-10-30	08:00:33.431941	M033	ON
2007-10-30	08:00:33.971784	M032	ON
2007-10-30	08:00:35.583358	M033	OFF
2007-10-30	08:00:35.583358	M036	ON
2007-10-30	08:00:35.879244	M031	ON
2007-10-30	08:00:37.166895	M035	OFF
2007-10-30	08:00:37.555785	M034	OFF
2007-10-30	08:00:38.446526	M032	OFF
2007-10-30	08:00:38.702455	M036	OFF
2007-10-30	08:00:45.769494	M031	OFF

Process Discovery from Sensor Event Data

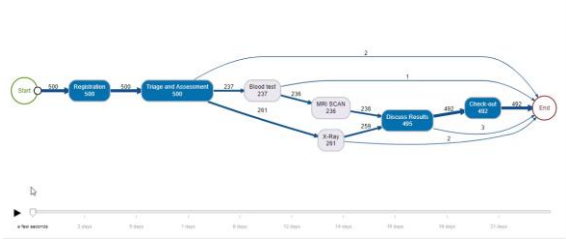
Christian-Albrechts-Universität zu Kiel

- Room plan
- Sensor types
  - M = Motion
  - T = Temperature



- data set: http://casas.wsu.edu/datasets/

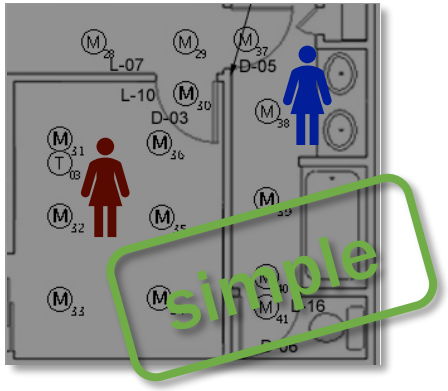
```
2007-10-30 08:00:02.711245  M035    ON
2007-10-30 08:00:08.190289  M035    OFF
2007-10-30 08:00:08.569183  M034    ON
2007-10-30 08:00:08.070915  M035    ON
2007-10-30 08:00:15.171679  M034    OFF
2007-10-30 08:00:17.671914  M034    ON
2007-10-30 08:00:26.877751  M034    OFF
2007-10-30 08:00:28.750241  M034    ON
2007-10-30 08:00:33.431941  M033    ON
2007-10-30 08:00:33.971784  M032    ON
2007-10-30 08:00:35.583358  M033    OFF
2007-10-30 08:00:35.583358  M036    ON
2007-10-30 08:00:35.879244  M031    ON
2007-10-30 08:00:37.166895  M035    OFF
2007-10-30 08:00:37.555785  M034    OFF
2007-10-30 08:00:38.446526  M032    OFF
2007-10-30 08:00:38.702455  M036    OFF
2007-10-30 08:00:45.769494  M031    OFF
```

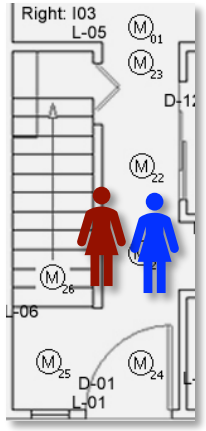


Challenges

- What is an activity?
  - What is a start/end activity?
  - Unlabeled log hampers validity
- A lot of noise/ data quality issues
- Find suitable cluster technique
  - Distance between sensors
  - Activity duration of sensors
  - Order of activities

- How to validate the discovered processes?
- Several entities

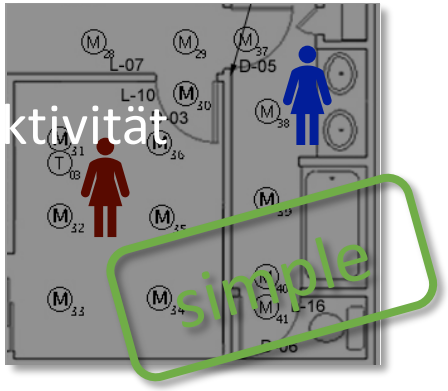


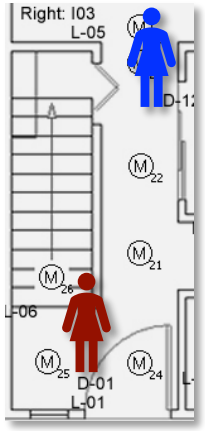


Challenges

- What is an activity?
  - What is a start/end activity?
  - Unlabeled log hampers validity
- A lot of noise/ data quality issues
- Find suitable cluster technique
  - Distance between sensors
  - Activity duration of sensors
  - Order of activities

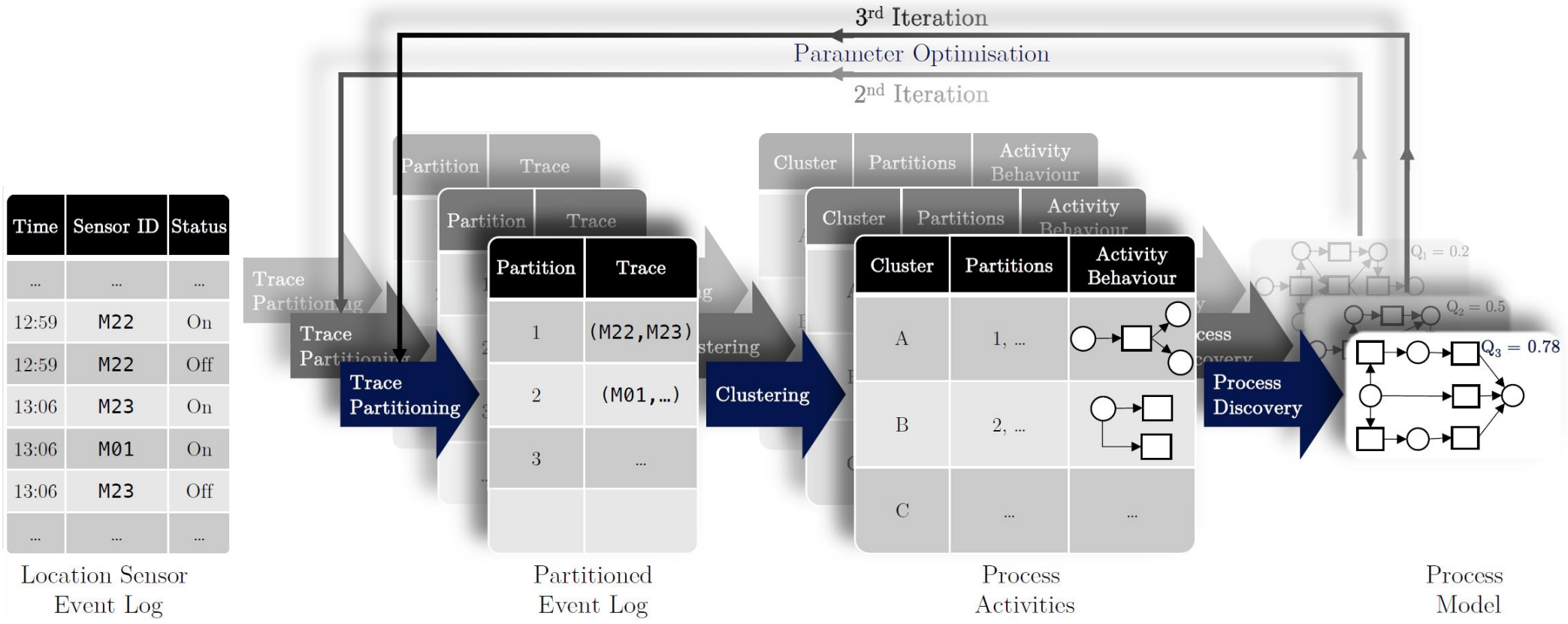
- How to validate the discovered processes?
- Several entities





From Sensor Event Data to Process Models

Christian-Albrechts-Universität zu Kiel



D. Janssen, A. Koschmider, F. Mannhardt: The Road from Sensor Data to Process Discovery and back, submitted, 2022

D. Janssen, F. Mannhardt, A. Koschmider, S.: Process Model Discovery from Sensor Event Data. ICPM Workshops 2020: 69-81

Sensor Event Data → Partitioned Event Log

Christian-Albrechts-Universität zu Kiel

Data extraction

- Data set
- Filtering & adjacency matrix

Vectoring

- Construct blocks (according to sensors, time, room)
- Construct vectors (according to quantity, time)

Clustering

- K-means/Self-Organizing Map/Auto encoder
- Own distance calculation
- Threshold
- Pseudo labeling

Sensor Event Data → Partitioned Event Log

Christian-Albrechts-Universität zu Kiel

- Quantity fix, time variable: 5 sensors are considered a block

11	2011-06-12 01:51:16.024247	M003	ON	1
12	2011-06-12 01:51:21.487317	M003	OFF	
13	2011-06-12 01:53:55.914712	M003	ON	2
14	2011-06-12 01:53:56.320168	M002	ON	3
15	2011-06-12 01:53:57.858869	M002	OFF	
16	2011-06-12 01:54:01.701934	M003	OFF	
17	2011-06-12 01:54:02.086128	M003	ON	4
18	2011-06-12 01:54:02.931032	M002	ON	5
19	2011-06-12 01:54:04.704464	M003	OFF	
20	2011-06-12 01:54:07.57761	M002	OFF	

Block: {M003, M003, M002, M003, M002}

- Time fix, quantity variable: 20 seconds are considered a block

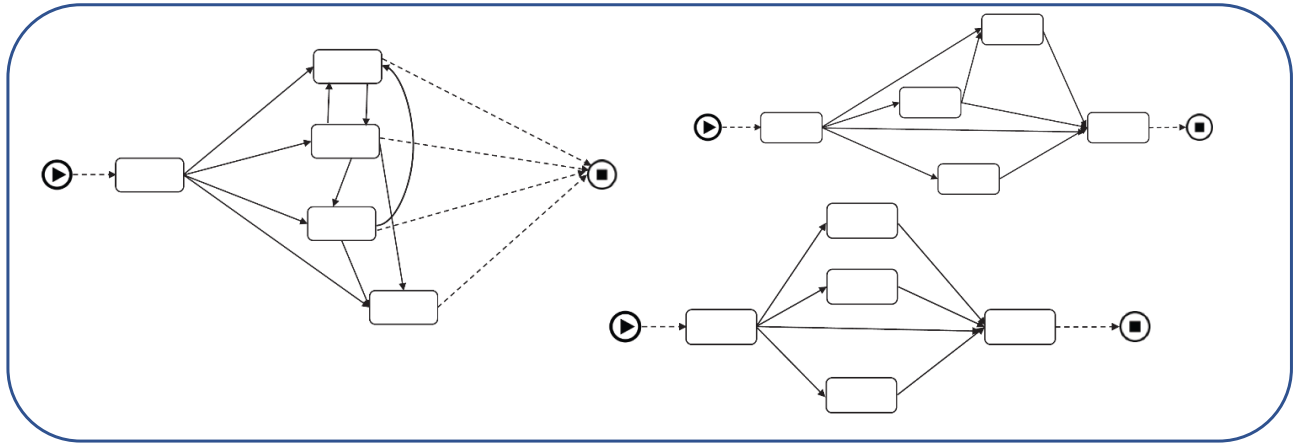
11	2011-06-12 01:51:16.024247	M003	ON	5 Sek.
12	2011-06-12 01:51:21.487317	M003	OFF	
13	2011-06-12 01:53:55.914712	M003	ON	6 Sek.
14	2011-06-12 01:53:56.320168	M002	ON	1 Sek.
15	2011-06-12 01:53:57.858869	M002	OFF	
16	2011-06-12 01:54:01.701934	M003	OFF	
17	2011-06-12 01:54:02.086128	M003	ON	2 Sek.
18	2011-06-12 01:54:02.931032	M002	ON	5 Sek.
19	2011-06-12 01:54:04.704464	M003	OFF	
20	2011-06-12 01:54:07.57761	M002	OFF	

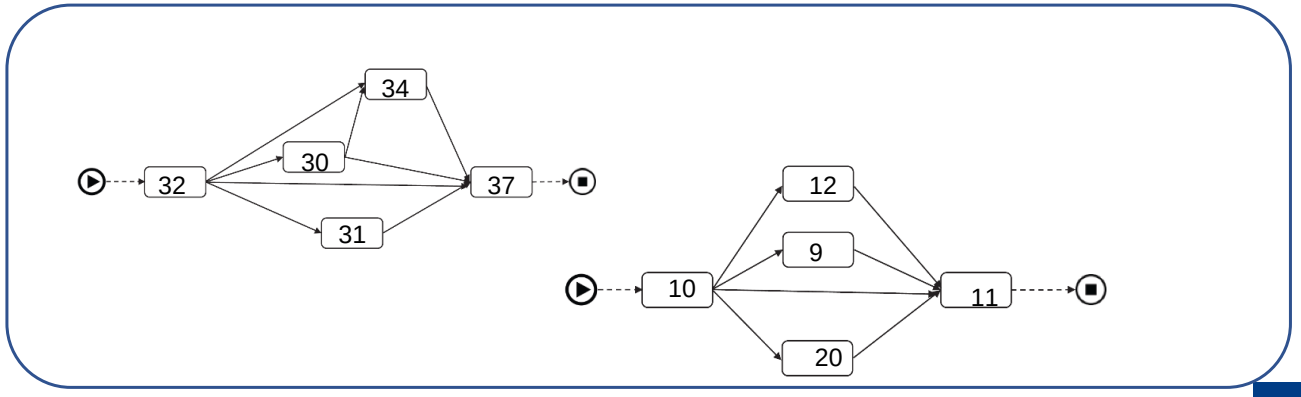
Block: {M003, M003, M002, M003, M002}

Partitioned Event Log → Process Activities

Christian-Albrechts-Universität zu Kiel

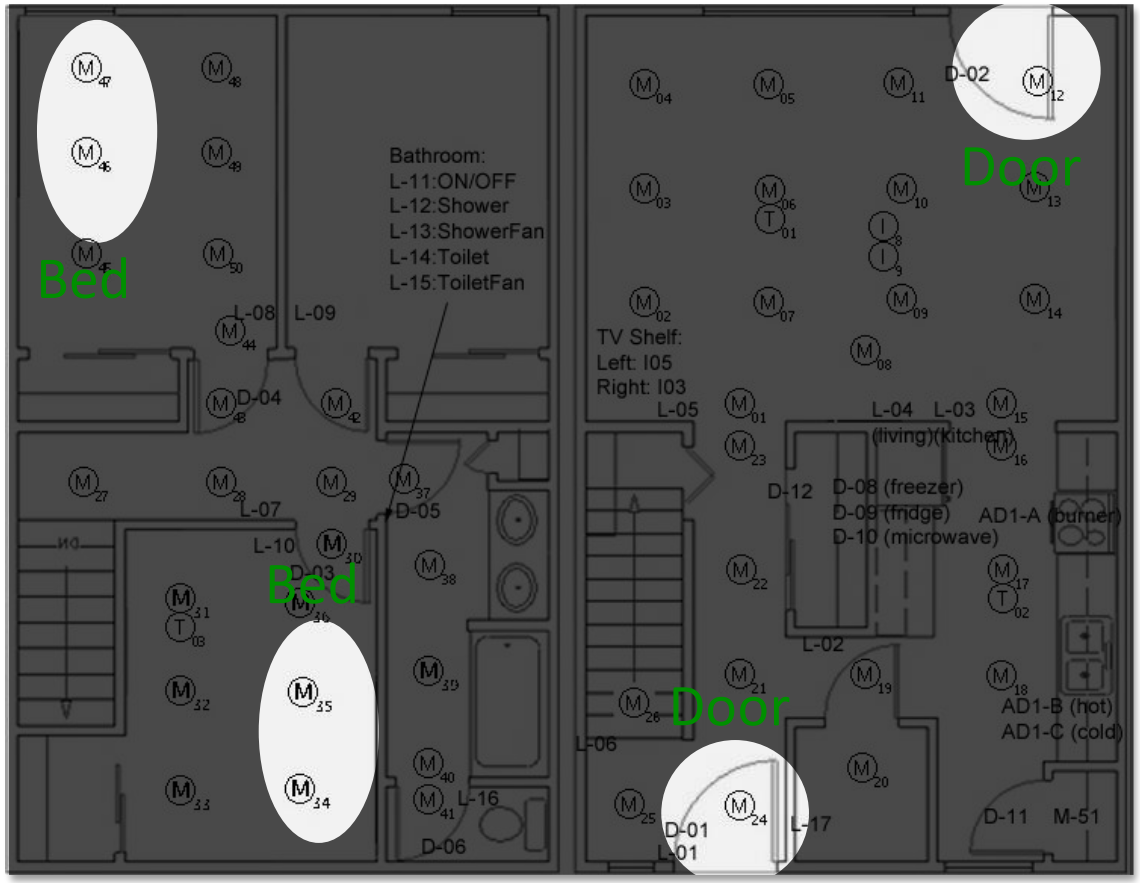
[0, 0, 0, 10, 0, 0, 0, 0, 21, 6, 0, 0, 0, 32, 772, 63050, 3803, 3969, 5189, 45839, 243, 358, 67, 0, 4, 0, 0, 0, 0, 2, 3, 12]

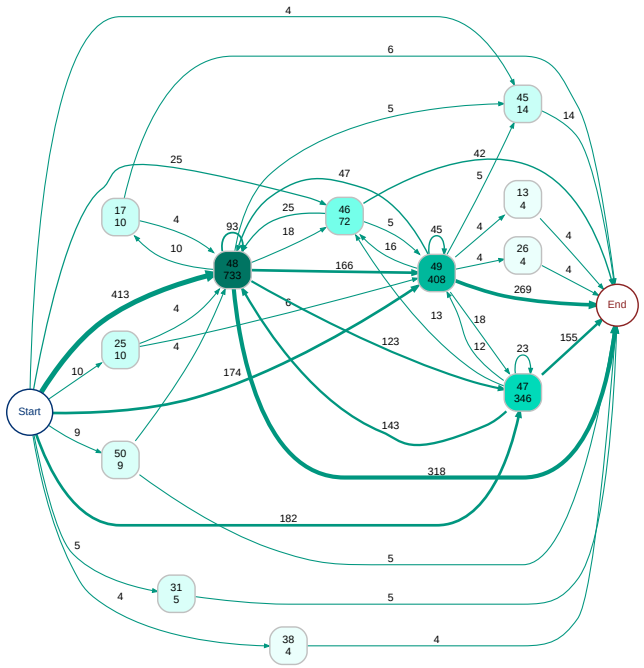




What is a suitable Start Activity?

Christian-Albrechts-Universität zu Kiel



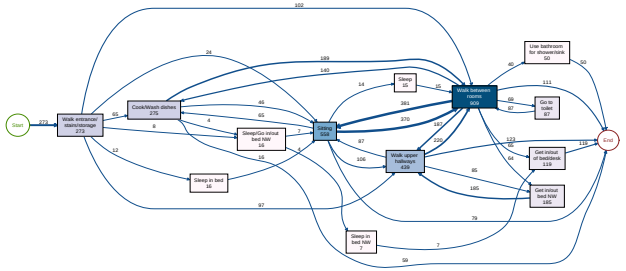


Process Activity → Process Model

Christian-Albrechts-Universität zu Kiel

	Activity			
Case	Time	Label	Inst.	Life Cycle
1	11:01	Cooking	1	S
2	11:01	TV	2	S
1	11:33	Cooking	1	C
...	...	...	...	...
...	...	...	...	...

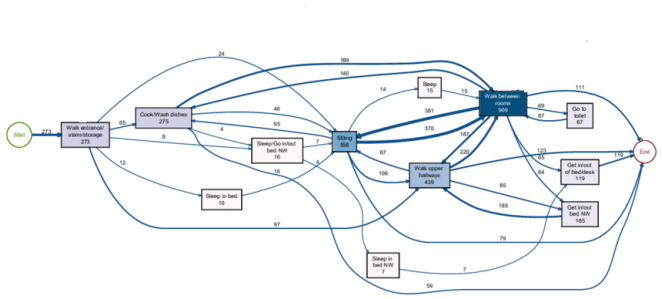
Process Mining



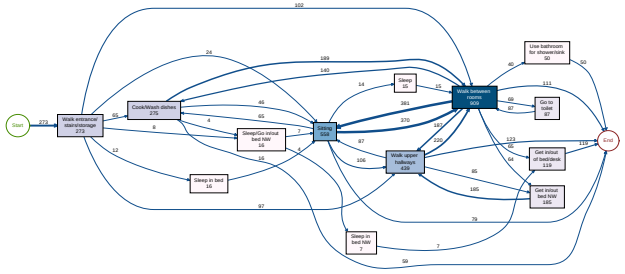
Process Model

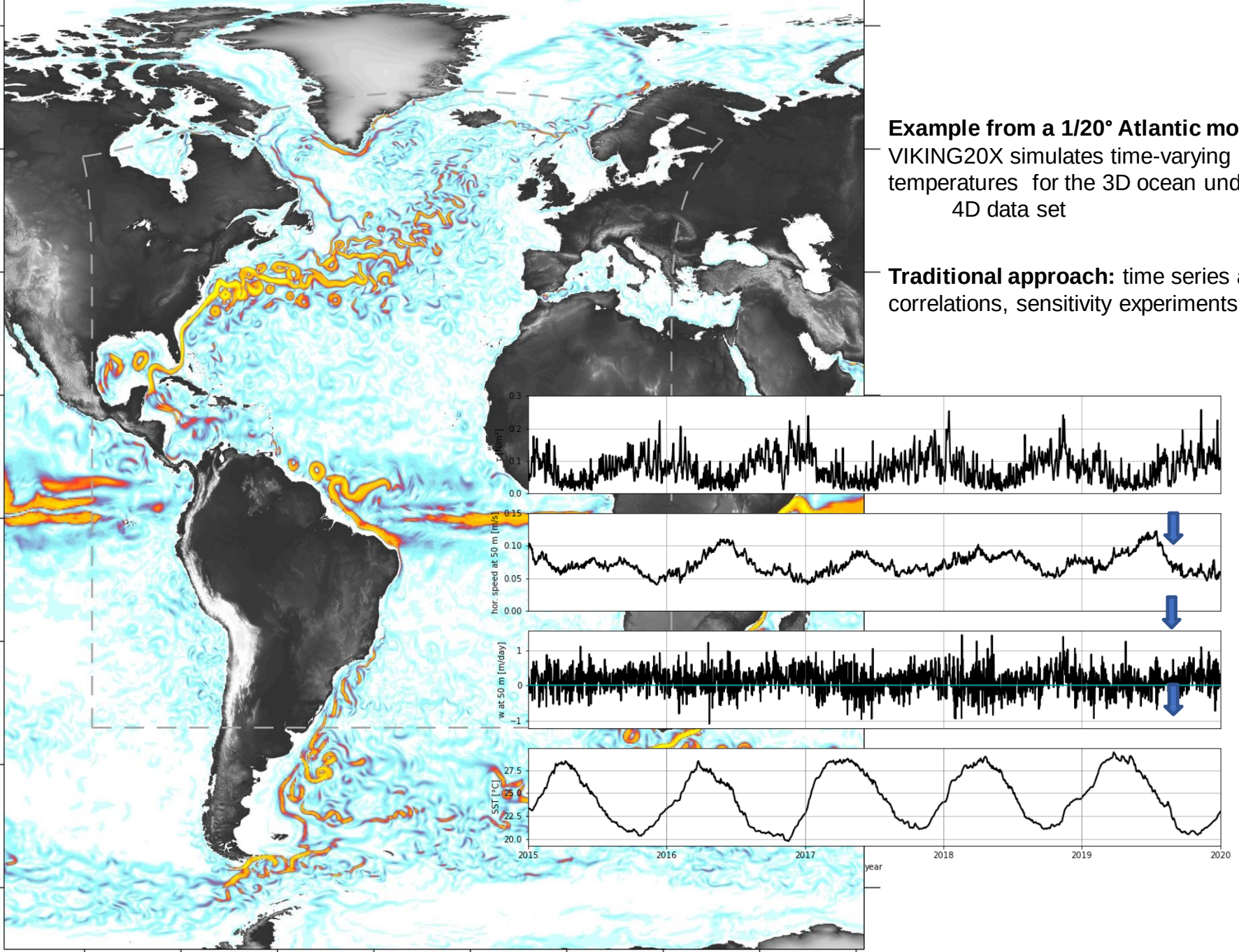
Model Quality

Christian-Albrechts-Universität zu Kiel



Entropia & Hyperparameter Optimization





Example from a 1/20° Atlantic model
VIKING20X simulates time-varying currents & temperatures for the 3D ocean under changing winds
4D data set →

Traditional approach: time series analyses, spatial correlations, sensitivity experiments

Integral timeseries:

Variable winds…

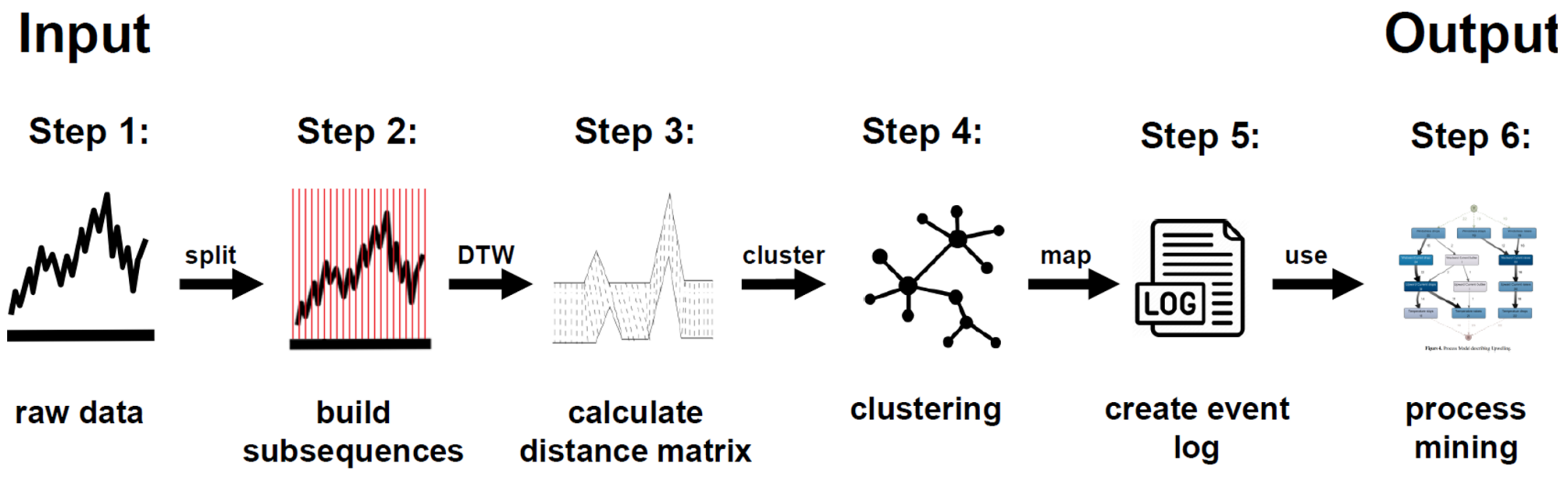
…change currents…

…cause upwelling…

…with impact on temperature

Process Discovery for Time Series Data

Christian-Albrechts-Universität zu Kiel



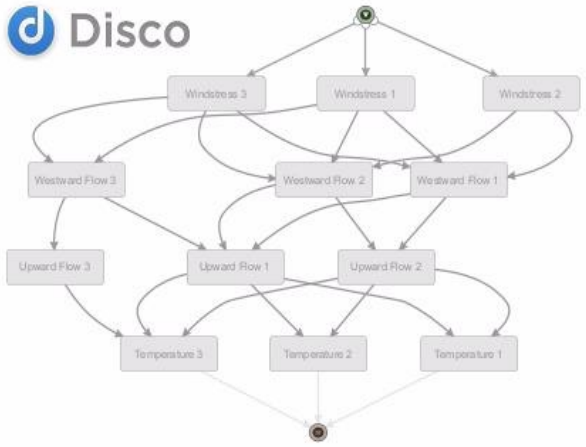
T. Ziolkowski, A. Koschmider, R. Schubert, M. Renz: Process Mining for Time Series Data: Insights into Ocean Processes, submitted, 2022

Traditional Analysis vs. Process Mining-Based

Christian-Albrechts-Universität zu Kiel

- **Predictive analytics:** our approach allows understanding temporal pattern/trend in what is being measured. In natural science like ocean science it can even give an early indication on the overall direction of a typical ocean cycle, which is hardly to predict with traditional approaches in ocean science

- **Outlier detection:** outliers detected in a dataset can help prevent unintended consequences and point to new processes.



Raw Data → Data Aggregation

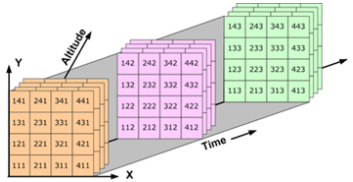
Christian-Albrechts-Universität zu Kiel

Region

Climate data
(2015-2019) of
Region:
21-17 °W
14-18 °N

generate

Data



Measurements of:

- water temperature
- windstress
- vertical and horizontal flow rate

convert

Domain specific format

```
2015-01-01,24.47704315185547, 0,05, 0,6, 0,8
2015-01-02,24.438640594482422, 0,12, 0,48, 0,66
2015-01-03,24.402971267700195, 0,2, 0,74, 0,44
2015-01-04,24.337369918823242, 0,04, 0,64, 0,81
```

Extract and map
values in csv-Format to
prepare for further
process analytics

foto source: https://desktop.arcgis.com/de/arcmap/10.3/manage-data/netcdf/fundamentals-of-netcdf-data-storage.htm

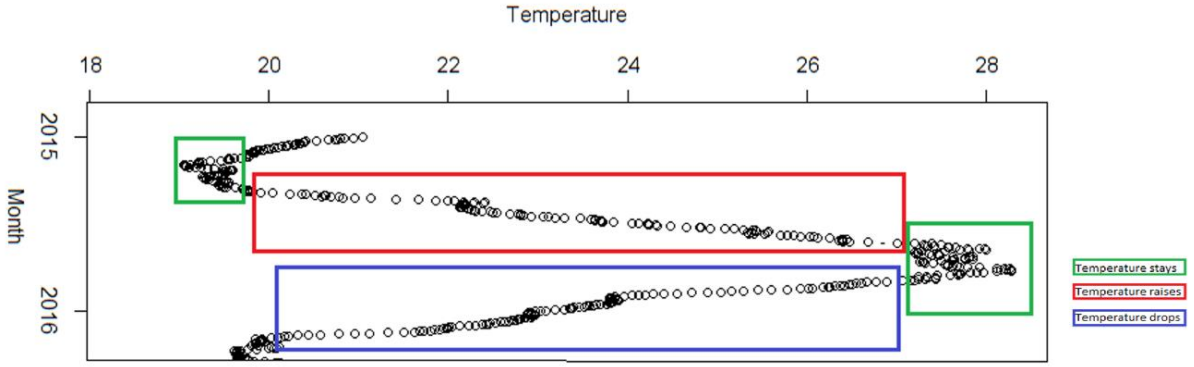
Data Aggregation → Event Log

Christian-Albrechts-Universität zu Kiel

Distance matrix:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
1	0.00000000	0.51955468	1.1712184	1.5897739	0.5241456	1.73360815	2.8770152	3.64472418	4.08577449	4.31753257	3.25987113	1.97672542	1.3153814	0.31869539	0.2371217	0.56255112	0.1582033
2	0.51955468	0.00000000	0.4079769	0.5237244	0.8383750	2.49683552	3.9453226	4.48751119	4.96705011	5.41851435	4.37665898	3.09351327	2.3500787	1.20730683	0.1105747	0.21793248	0.4558184
3	1.17121840	0.40797691	0.0000000	0.5230939	1.7363864	3.44214306	4.6326914	5.36793493	5.80595814	6.07925994	5.02832270	3.74517698	3.0659574	1.85897055	0.7419530	0.80461095	1.1361262
4	1.58977388	0.52372440	0.5230939	0.0000000	1.0913348	2.72935203	4.4123218	4.78979810	5.30018797	5.97230562	5.48480857	4.04867805	2.8457935	2.33367481	0.7235108	0.61265637	1.0924312
5	0.52414560	0.83837496	1.7363864	1.0913348	0.0000000	1.01430260	2.4481959	3.01236931	3.52774665	3.97198919	3.27588437	1.90217126	0.9436152	0.66782274	0.5649572	0.31198428	0.3603147
6	1.73360815	2.49683552	3.4421431	2.7293520	1.0143026	0.00000000	0.6014207	1.06203934	1.58538154	2.15397050	1.54353863	0.59594564	0.4381323	1.14224370	2.1556413	1.78590657	1.7570683
7	2.87701524	3.94532264	4.6326914	4.4123218	2.4481959	0.60142072	0.0000000	0.42854080	0.67415425	1.13488022	0.55689859	0.39155874	1.3390981	1.87130648	3.6145899	3.45098066	3.2997005
8	3.64472418	4.48751119	5.3679349	4.7897981	3.0123693	1.06203934	0.4285408	0.00000000	0.14843721	0.44867074	0.62062159	1.12125870	1.8193276	2.74320170	4.1554118	3.82443412	3.7644890
9	4.08577449	4.96705011	5.8059581	5.3001880	3.5277466	1.58538154	0.6741542	0.14843721	0.00000000	0.16048231	0.44851309	1.56824983	2.2990101	3.08565779	4.6289084	4.32323014	4.2440845
10	4.31753257	5.41851435	6.0792599	5.9723056	3.9719892	2.15397050	1.1348802	0.44867074	0.16048231	0.00000000	0.38513193	1.81771549	2.8302924	3.31038151	5.1022418	4.97259445	4.7884010
11	3.25987113	4.37665898	5.0283227	5.4848086	3.2758844	1.54353863	0.5568986	0.62062159	0.44851309	0.38513193	0.00000000	0.75781959	1.7893787	2.25580819	4.0476927	4.26967319	3.7600474
12	1.97672542	3.09351327	3.7451770	4.0486780	1.9021713	0.59594564	0.3915587	1.12125870	1.56824983	1.81771549	0.75781959	0.00000000	0.5062330	0.97266248	2.7645470	2.97974021	2.4769017
13	1.31538143	2.35007868	3.0659574	2.8457935	0.9436152	0.43813225	1.3390981	1.81932761	2.29901009	2.83029245	1.78937867	0.50623296	0.0000000	0.42347889	2.1025497	1.83773985	1.7290441

Result of Clustering:

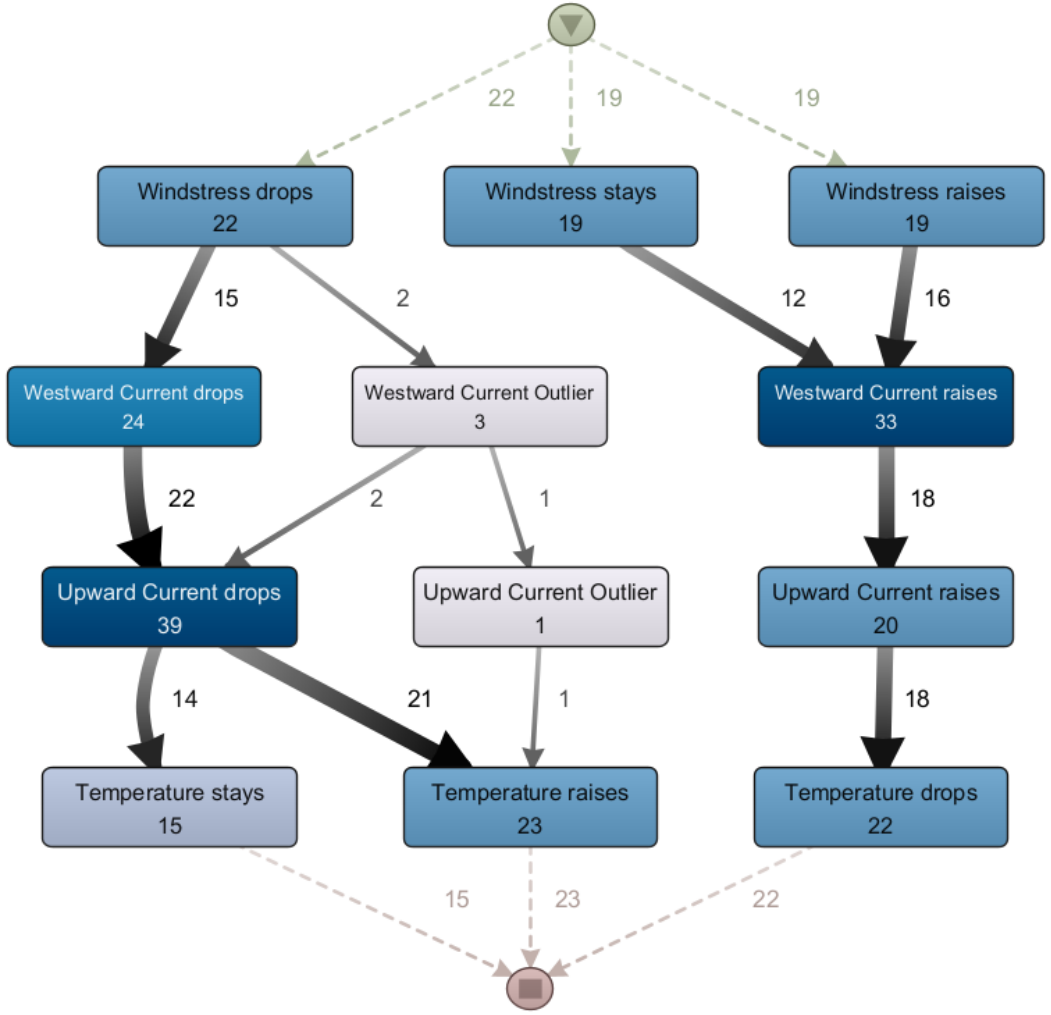


Event log:

Timestamp	Activity	CaseID
Jan 15	Erhöhter Wind	2015
Jan 15	Gleiche Strömung	2015
Jan 15	Gleiches Upwelling	2015
Jan 15	Gleiche Temperatur	2015
Feb 15	Erhöhter Wind	2015
Feb 15	Erhöhte Strömung	2015
Feb 15	Gleiches Upwelling	2015
Feb 15	Gleiche Temperatur	2015

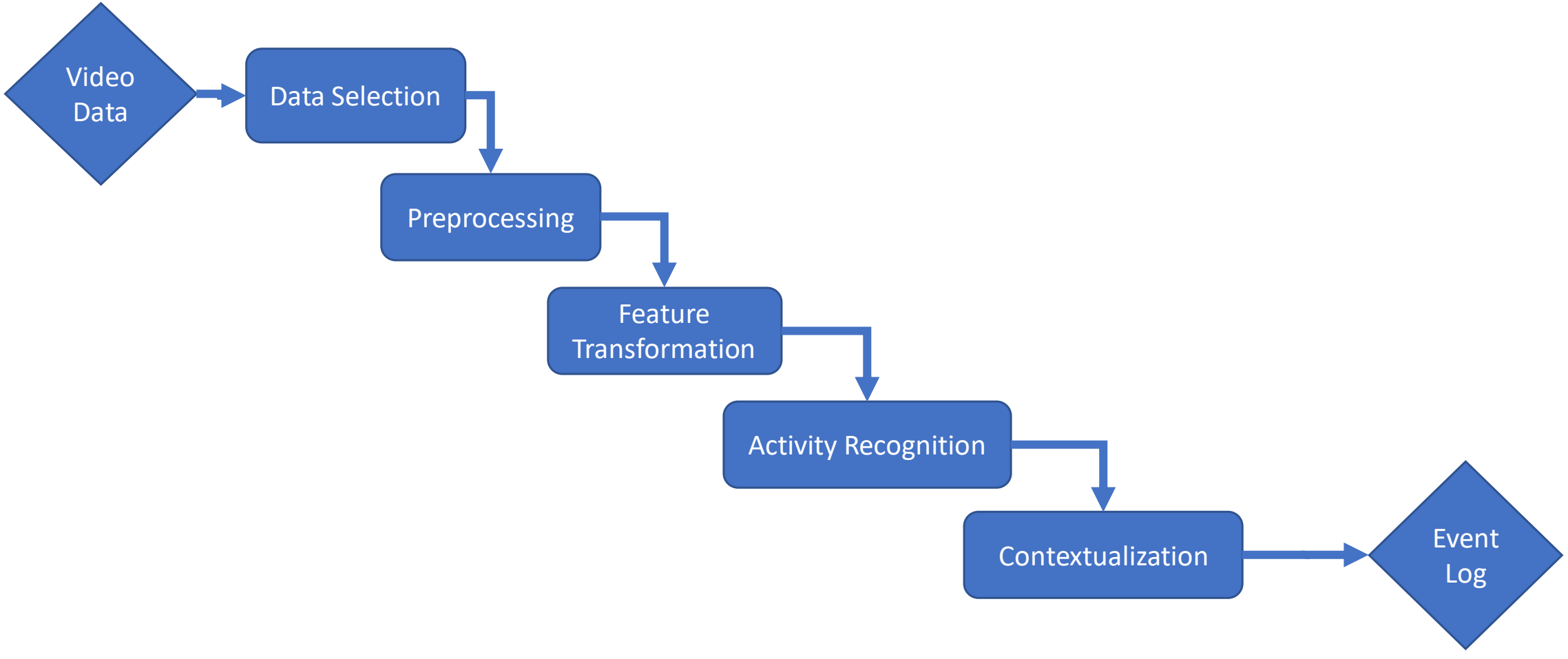
Event Log → Process Model

Christian-Albrechts-Universität zu Kiel



Process Mining on Image Data

Christian-Albrechts-Universität zu Kiel



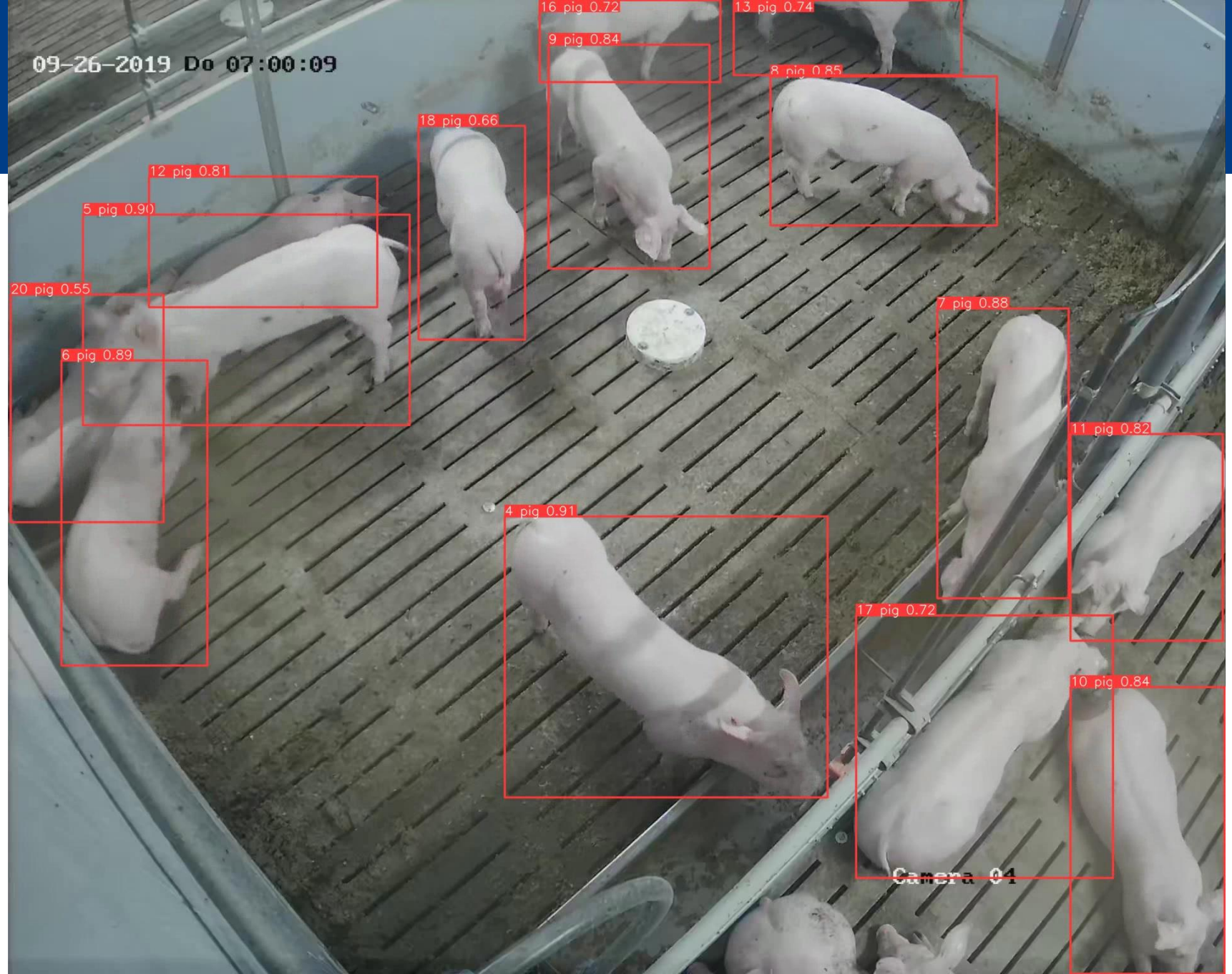
Example (128x128 Pixel)

Example (128x128 Pixel)

Example (128x128 Pixel)

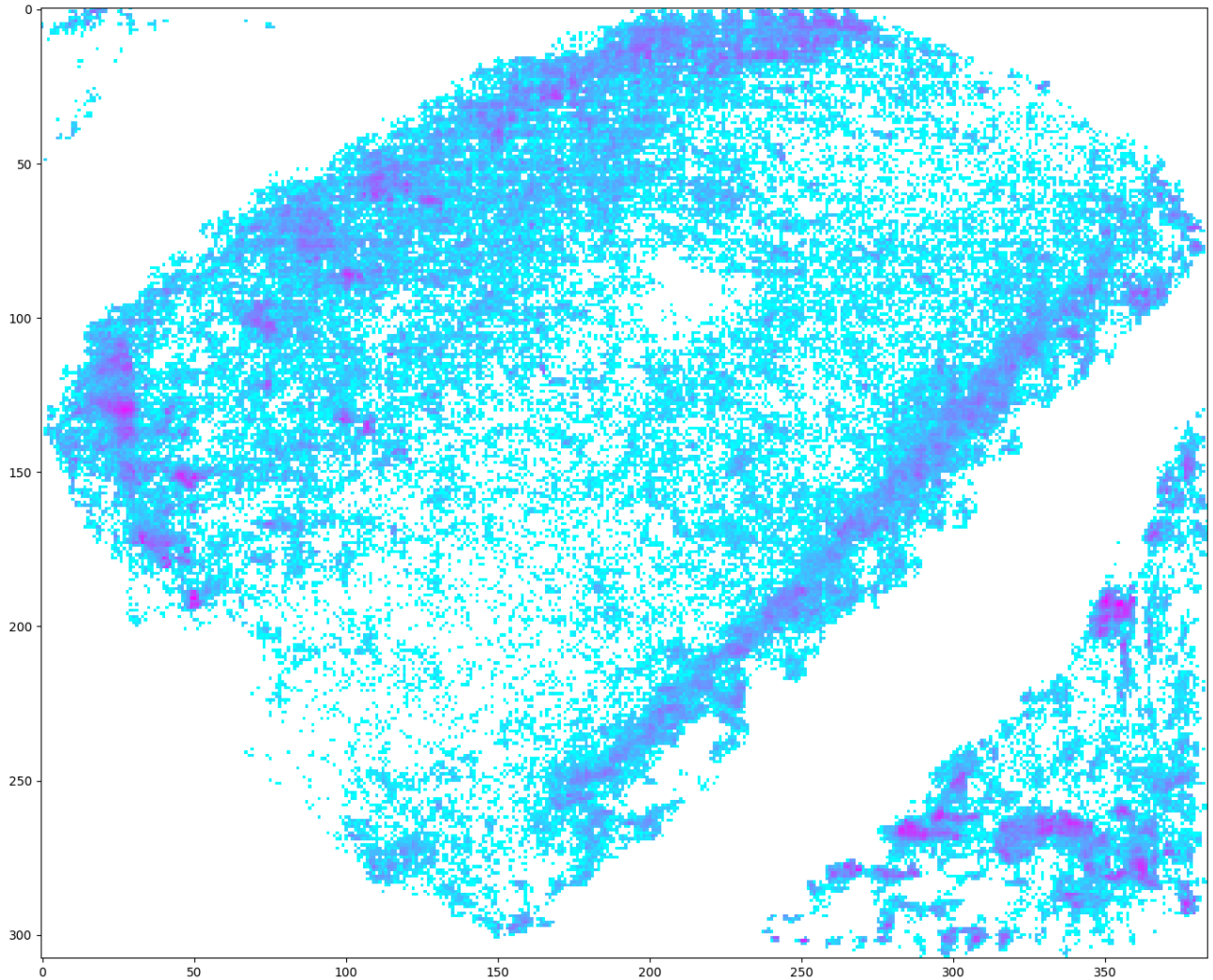
Example (128x128 Pixel)

Christian-Albrechts-Universität zu Kiel

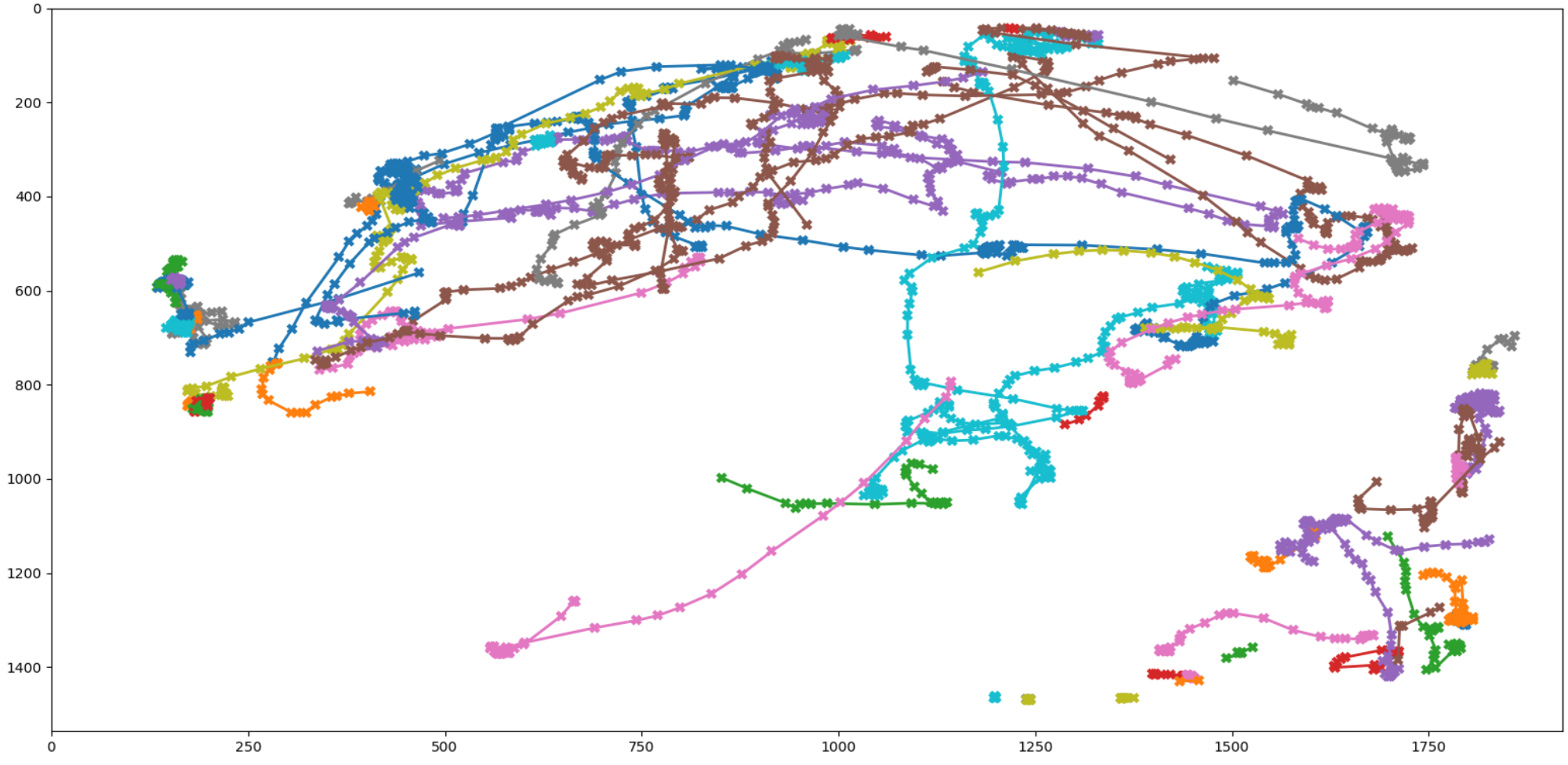


Christian-Albrechts-Universität zu Kiel

Pig positions over 1h

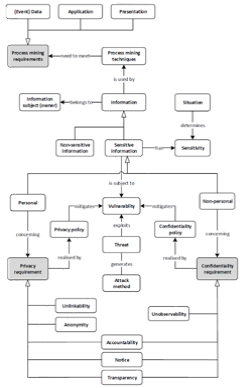


Pig movement traces

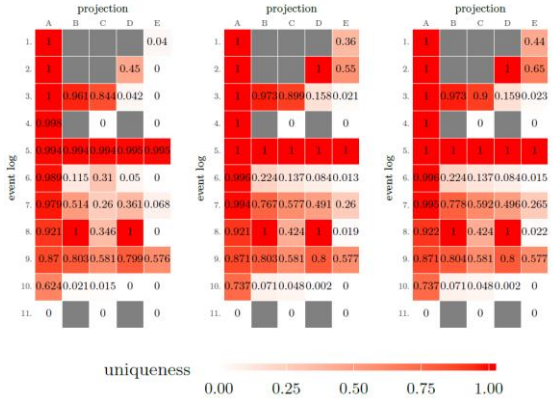


Privacy in Data Analytics

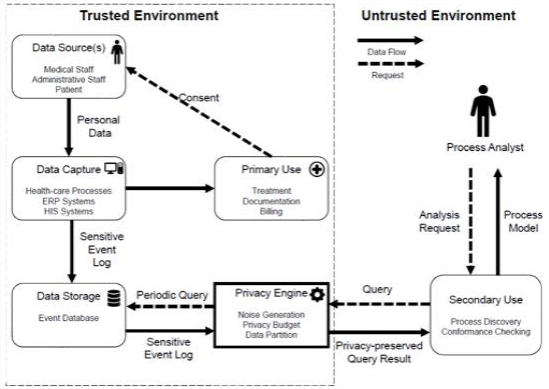
Christian-Albrechts-Universität zu Kiel



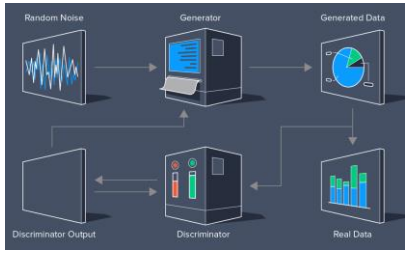
G. Elkoumy, S.A. Fahrenkrog-Petersen, M. Fani Sani, A. Koschmider, F. Mannhardt, S. Nuñez von Voigt, M. Rafiei, L. von Waldthausen: Privacy and Confidentiality in Process Mining - Threats and Research Challenges, ACM Transactions of Management Information Systems, 2022 akzeptiert



S. Nuñez von Voigt, S.A. Fahrenkrog-Petersen, D. Janssen, A. Koschmider, F. Tschorsch, F. Mannhardt, O. Landsiedel, M. Weidlich: Quantifying the Re-identification Risk of Event Logs for Process Mining. CAiSE 2020: 252-267

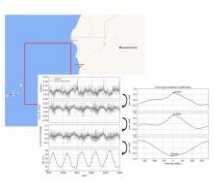


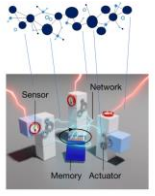
F. Mannhardt, A. Koschmider, N. Baracaldo, M. Weidlich, J. Michael: Privacy-Preserving Process Mining - Differential Privacy for Event Logs. Business & Information Systems Engineering 61(5): 595-614 (2019)

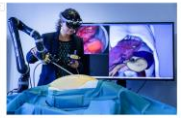


K. Kaczmarek. Koschmider: Conceptualizing a Log Generator for Privacy-aware Event Logs. In: TPSA Workshop 2021

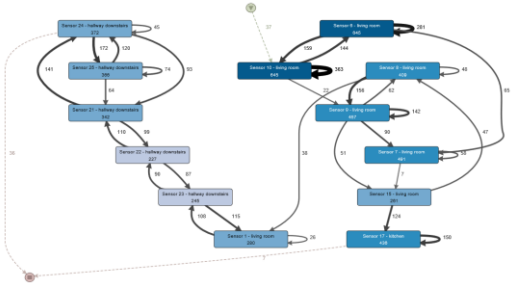
Summary







- What unknown processes are acting
  (i.e., did we found all processes that exist)?



- Whether the found processes actually work as thought