Data-Driven Business Process Management

What data can tell us about how organizations really work

UNIVERSITY OF MANNHEIM
Business School

Prof. Dr. Jana-Rebecca Rehse

07.11.2024

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

About Me

2009 – 2015
Saarland University
(Saarbrücken)
B.Sc. and M.Sc. in
Business Informatics

2014
Steven's Institute of Technology
(Hoboken, NJ)
Visiting Researcher

2015 – 2020
German Research Center for Artificial Intelligence
(Saarbrücken)
Researcher & Project Lead

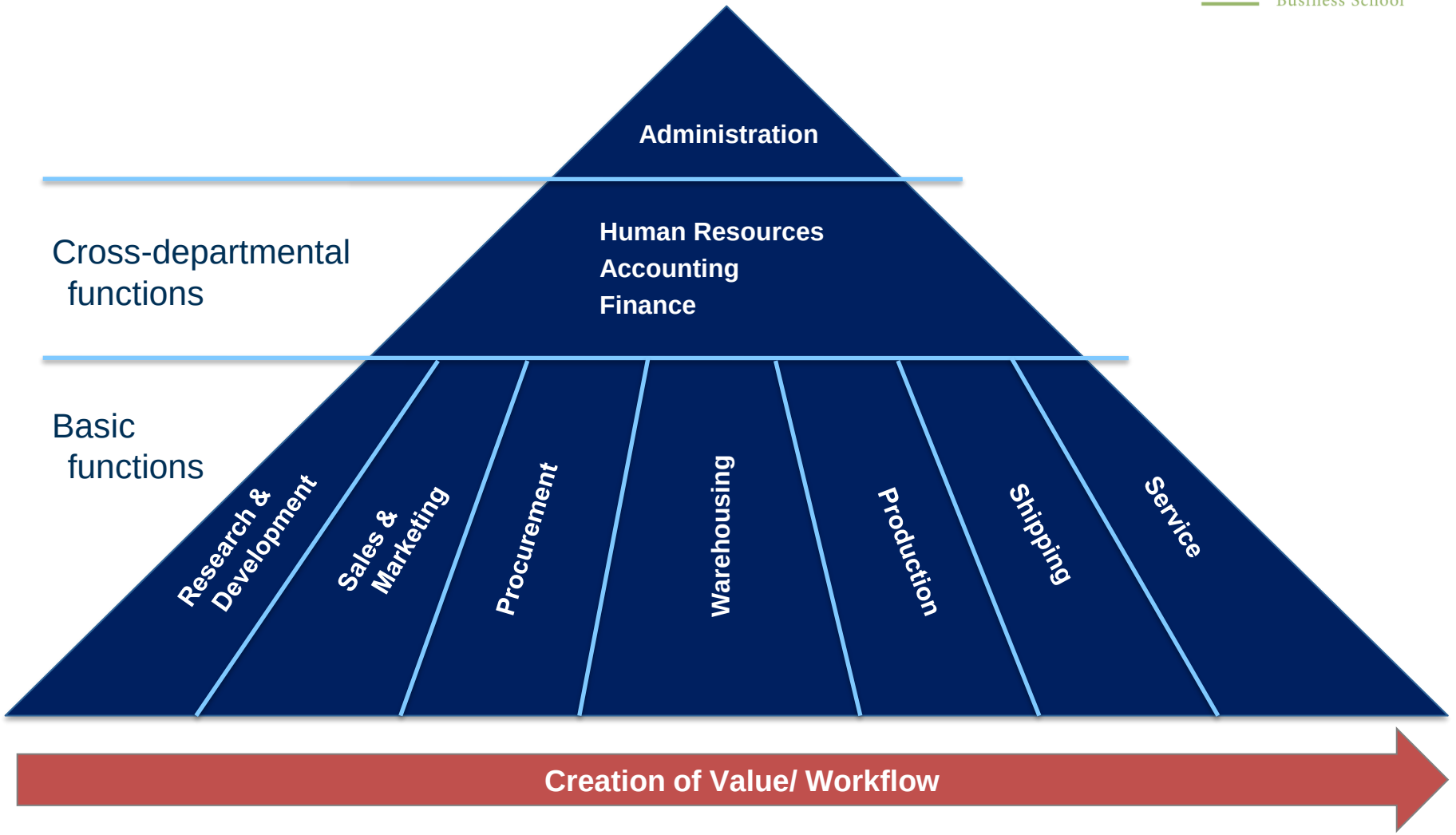
2019
Saarland University
(Saarbrücken)
Doctoral degree in
business

Since 2020
University of Mannheim
Junior Professor &
Research Group Lead

Focus: Data-driven business process management by means of process mining & machine learning

What is Data-driven Business Process Management?

How companies are organized



Processes are everywhere!

- Order-to-Cash
- Quote-to-Order
- Procure-to-Pay
- Issue-to-Resolution
- Application-to-Approval
- Lead-to-Opportunity
- Contract-to-Renewal
- Hire-to-Retire
- …

Processes in a university

- Student administration (customer service): Recruiting, enrollment, class assignment, recording grades, issuing certificates, counseling, …
- Employee administration (human resources): Issue job postings, recruiting, payroll, taxation, travel expense reporting, …
- Finances: Budgeting, accounting, offer handling, purchase acquisition handling, inventory management, …
- IT: Server administration, support, user management, …
- …

Questions about processes

How do you manage processes?

UNIVERSITY OF MANNHEIM
Business School

Ask the people!

- Many processes are enacted by people, who know about how the process works.
- Those people can be asked to learn about the process.

Ask the data!

- Many processes are supported by IT systems, which save data about the process.
- This data can be analyzed to learn about the process

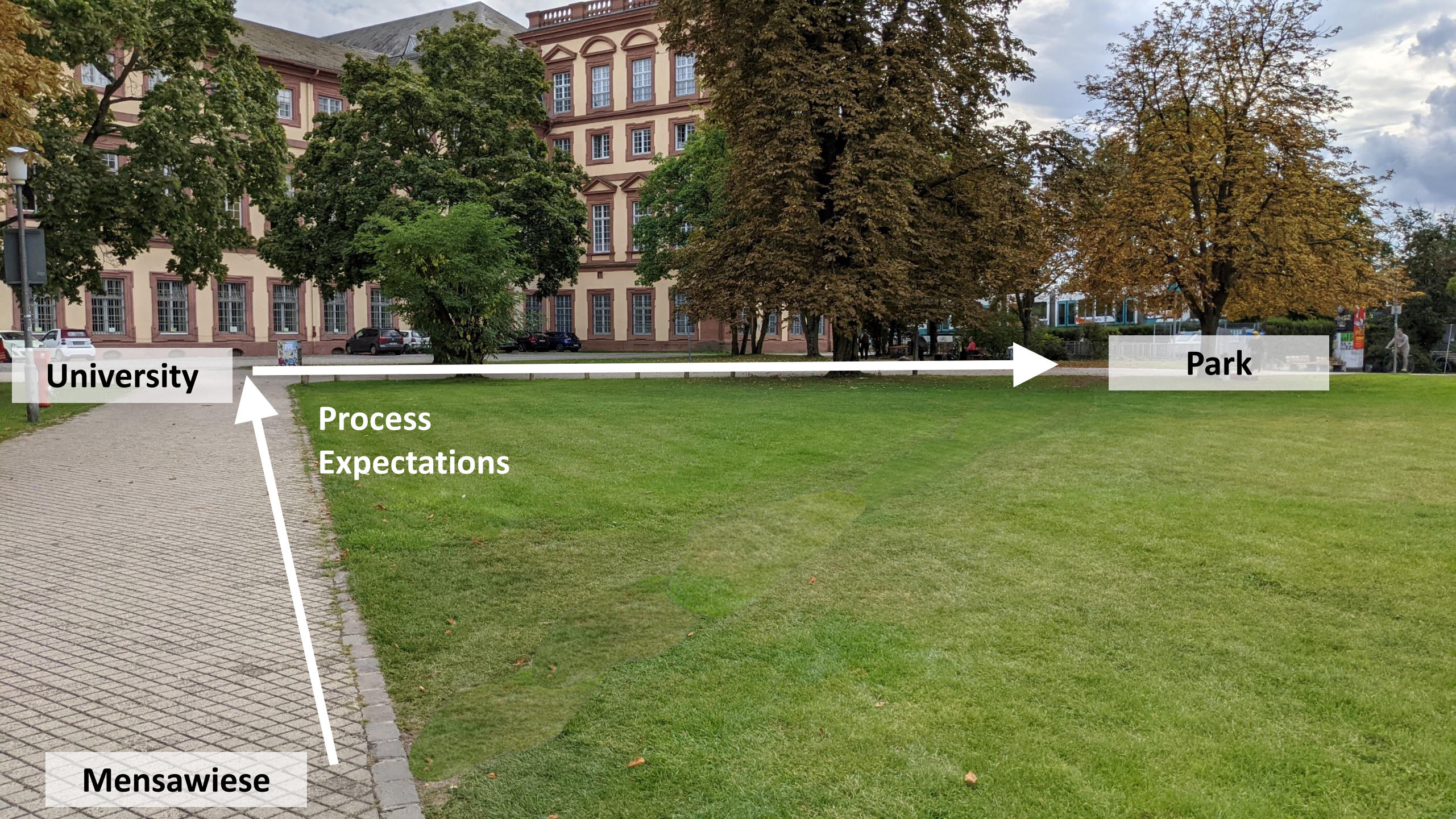
How do you manage those processes?

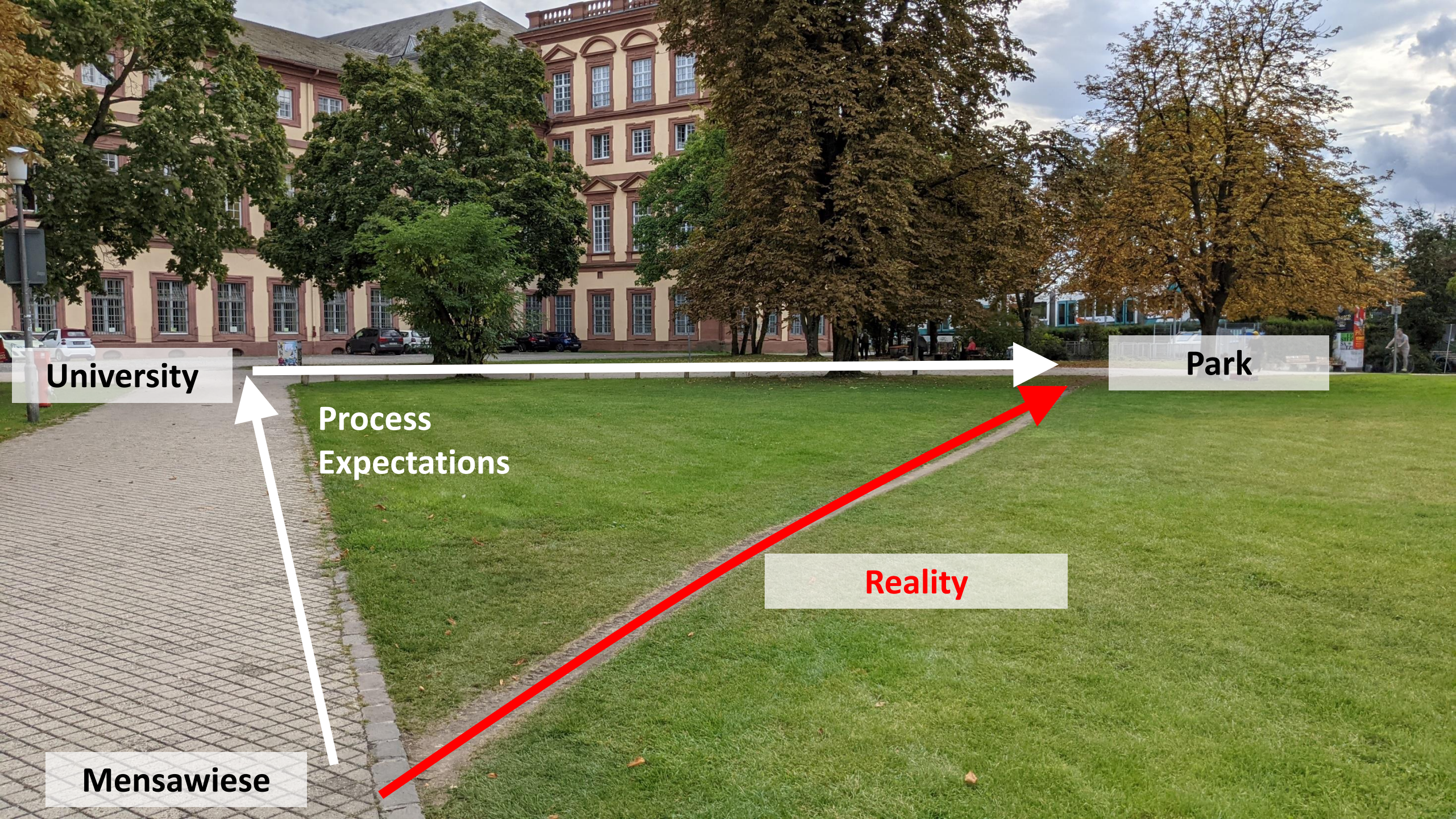
Ask the people!

- Pros:
  - Low technical effort.
  - May offer explanations.
  - Contains „soft“ factors.
- Contras:
  - High organizational effort.
  - May show the to-be-process instead of the as-is-process
  - Only offers individual perspectives.

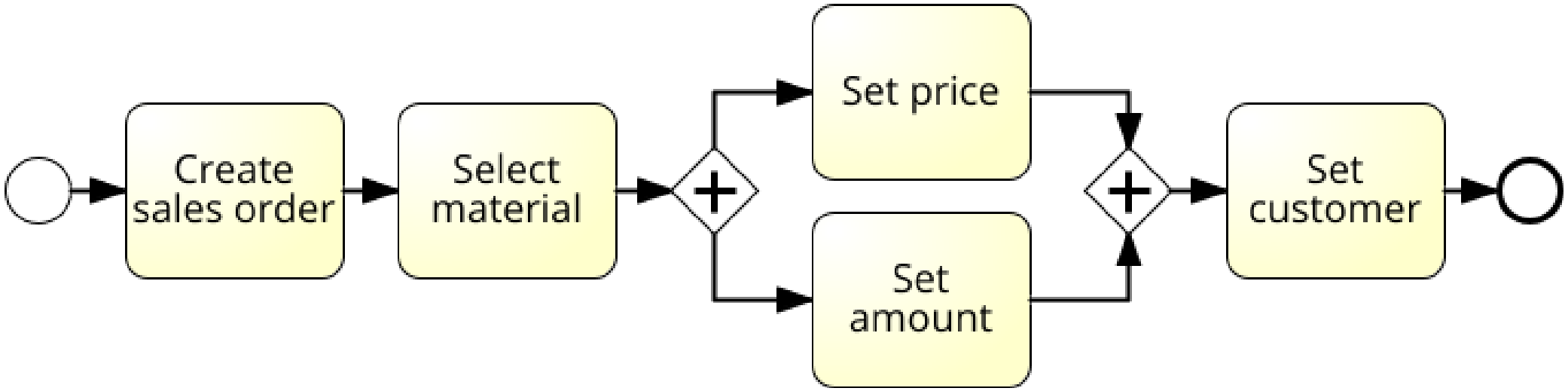
Ask the data!

- Pros:
  - Shows the as-is-process instead of the to-be-process.
  - Contains complete process.
  - Low organizational effort.
- Contras:
  - High technical effort.
  - May not show „soft“ factors.
  - Does not offer explanations.


University
Park
Process
Expectations
Mensawiese


University
Park
Process
Expectations
Reality
Mensawiese

A (Simple) Process



One Recorded Process Execution

Case ID	Activity	Timestamp
SO1	Create sales order	04.06.2023 17:21
SO1	Select material	04.06.2023 17:22
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24

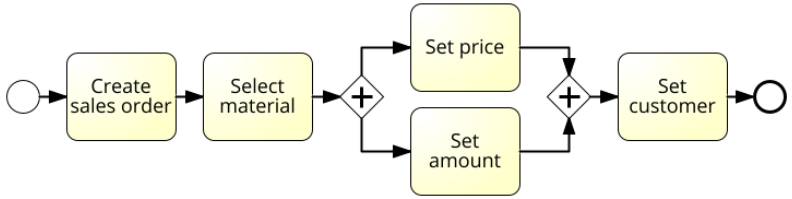
One event

One case

One Event Log

Case ID	Activity	Timestamp
SO1	Create sales order	04.06.2023 17:21
SO1	Select material	04.06.2023 17:22
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24
SO2	Create sales order	04.06.2023 17:21
SO2	Set price	04.06.2023 17:22
SO2	Set price	04.06.2023 17:22
SO2	Select material	04.06.2023 17:23
SO2	Set amount	04.06.2023 17:24
SO3	Create sales order	04.06.2023 17:21
SO3	Select material	04.06.2023 17:22
SO3	Set amount	04.06.2023 17:23
SO3	Set amount	04.06.2023 17:23
SO3	Set price	04.06.2023 17:24
SO4	Create sales order	04.06.2023 17:21
SO4	Set price	04.06.2023 17:22
SO4	Set amount	04.06.2023 17:23
SO4	Select material	04.06.2023 17:24
SO5	Create sales order	04.06.2023 17:21
SO5	Set amount	04.06.2023 17:22
SO5	Set customer	04.06.2023 17:23
SO6	Create sales order	04.06.2023 17:21
SO6	Set customer	04.06.2023 17:22
SO6	Set amount	04.06.2023 17:23

Process Mining



- What is the process that people really follow?
- Where are the bottlenecks in my process?
- Where do users deviate from the designed process?
- …

To-be process model

Process Design

Process Mining

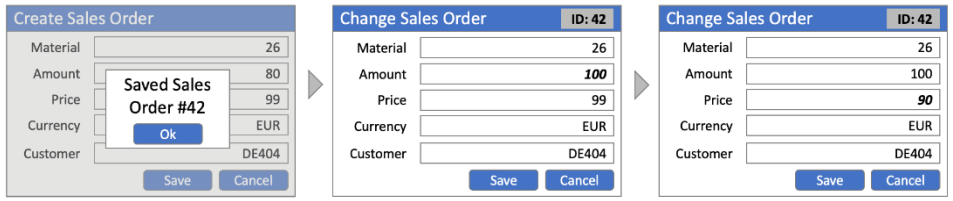
Insights

IT system

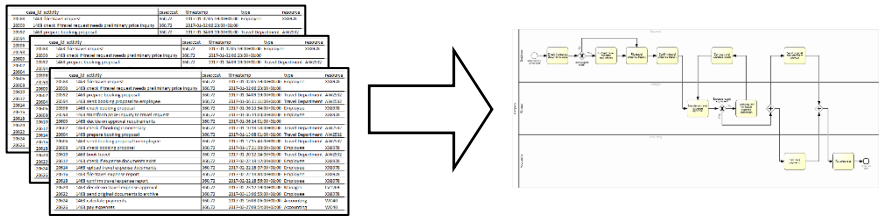
Process Implementation

Process Execution

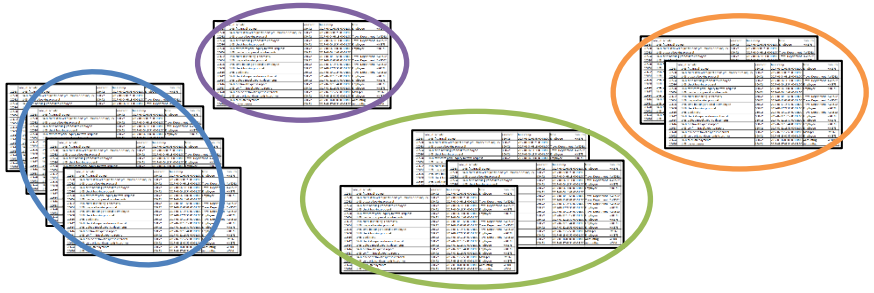
Event log



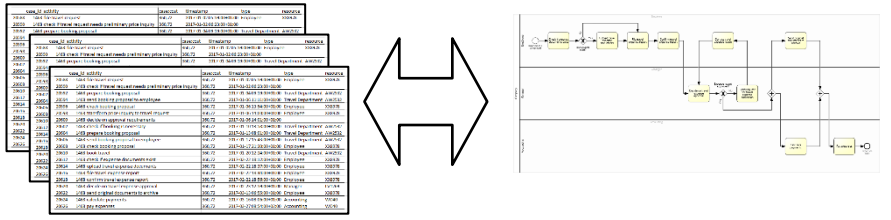
Case ID	Activity	Timestamp
SO1	Create sales order	04.06.2023 17:21
SO1	Select material	04.06.2023 17:22
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set price	04.06.2023 17:23
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24
SO1	Set amount	04.06.2023 17:24



What does the executed process look like?
(*Process Discovery*)

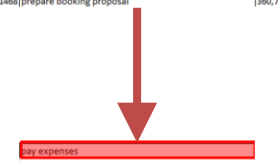


Which variants does the process have?
(*Trace Clustering*)

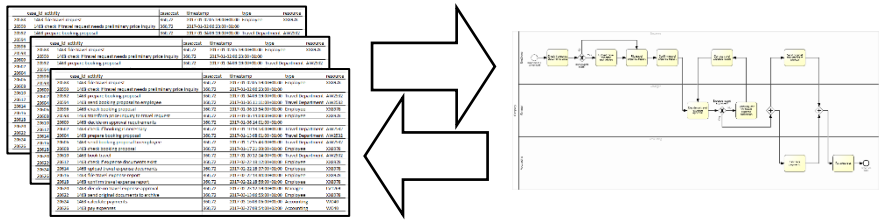


Where does the log differ from the model?
(*Conformance Checking*)

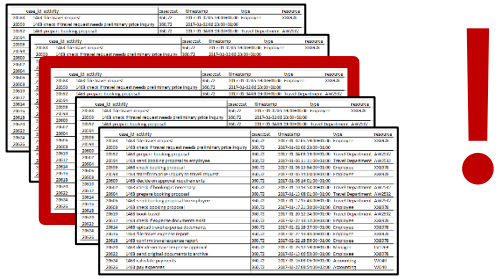
	case_id	activity	case:cost	timestamp	type	resource
20588	1468	file travel request	360,72	2017-01-02 05:59:00+01:00	Employee	XX8978
20590	1468	check if travel request needs preliminary price inquiry	360,72	2017-01-02 06:23:00+01:00		
20592	1468	prepare booking proposal	360,72	2017-01-04 09:19:00+01:00	Travel Department	AW2532
20594	1468	send booking proposal to employee	360,72	2017-01-06 11:11:00+01:00	Travel Department	AW2532
20596	1468	check booking proposal	360,72	2017-01-06 13:54:00+01:00	Employee	XX8978
20598	1468	transform price inquiry to travel request	360,72	2017-01-06 14:00:00+01:00	Employee	XX8978
20600	1468	decide on approval requirements	360,72	2017-01-06 14:01:00+01:00		
20602	1468	check if booking is necessary	360,72	2017-01-10 13:58:00+01:00	Travel Department	AW2532
20604	1468	prepare booking proposal	360,72	2017-01-13 08:01:00+01:00	Travel Department	AW2532



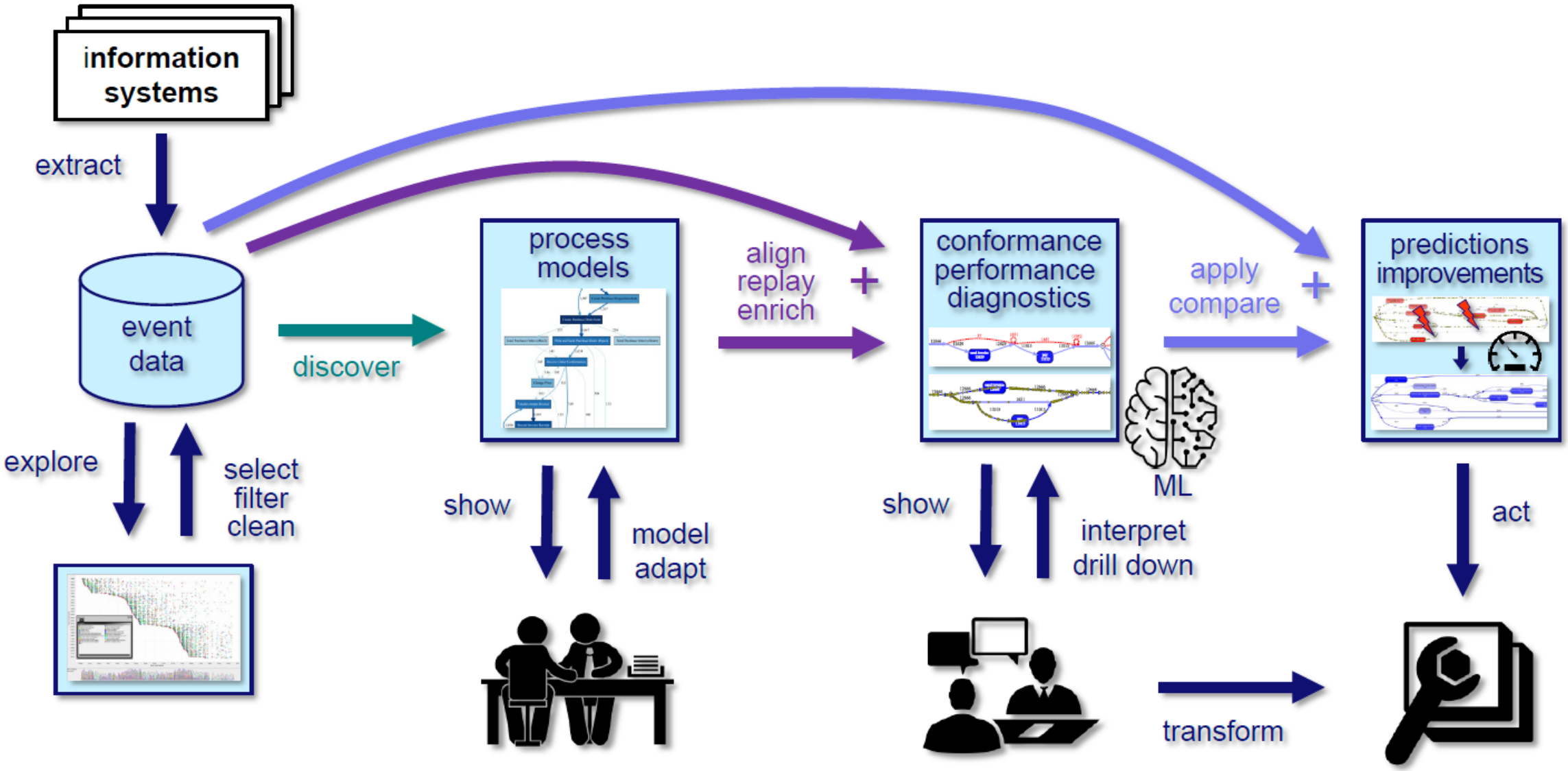
What is the outcome of a process instance?
(*Process Prediction*)



How can we improve the executed process?
(*Process Enhancement)*



Which process instance differs from others?
(*Anomaly Detection*)



Source: W. van der Aalst (2022)

Process mining in the press

UNIVERSITY OF MANNHEIM
Business School

Harvard Business Review

Coronavirus Magazine Popular Topics Podcasts Video Store The Big Idea Visual Library

ANALYTICS

What Process Mining Is, and Why Companies Should Do It

by Thomas H. Davenport and Andrew Spanyi

April 23, 2019

NHS England sees significant results with process mining technology

BY DEVONY HOF

England's National Health Services has started adopting Celonis SE's process mining platform to fix a variety of healthcare bottlenecks.

Jun 30, 2020, 04:56pm EDT

Process Mining: From Analytics To Action

Tom Davenport Contributor
CIO Network

Follow

FINISHED VEHICLE LOGISTICS

How process-mining in the automotive supply chain drives efficiency

By Megan Kelly | 12 September 2024

FORBES > INNOVATION

Why Being Data-Driven Is A Priority For Global Businesses To Succeed

Kerry Brown Forbes Councils Member
Forbes Technology Council
COUNCIL POST | Membership (Fee-Based)

Gartner Research

Market Guide for Process Mining

Published: 17 June 2019

ID: G00387812

Analyst(s): Marc Kerremans

Summary

New forms of automation (e.g., robotic process automation) and knowledge of the underlying processes/interactions are key to digital transformation. Process mining helps enterprise architecture and technology innovation leaders assess operations and performance, increasing these initiatives' value.

Prof. Dr. Jana-Rebecca Rehse

07.11.2024

Celonis & Signavio: Two Success Stories

UNIVERSITY OF MANNHEIM
Business School

PREMIUM • EDITORS' PICK

Celonis Raises $1 Billion At $11 Billion Valuation, Making It New York's —And Germany's — Most Valuable Startup

SAP Signavio

Why BPI, and why now? SAP Signavio's CEO on how process intelligence changes the enterprise - and how it reveals the true agenda of RISE with SAP

celonis

Celonis Ranks 13 on Forbes Cloud 100 Acknowledging Its Game Changing Process Intelligence Platform

August 07, 2024

Siemens Healthcare Chooses RISE with SAP for Digital Transformation

November 8, 2023

REAL ESTATE REALTY CHECK

Celonis becomes newest full-floor tenant at One World Trade Center

By Steve Cuozzo February 10, 2020 | 9:56pm

Prof. Dr. Jana-Rebecca Rehse

07.11.2024

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

And what exactly is your research about…?

My three streams of research

UNIVERSITY OF MANNHEIM
Business School

User Behavior Mining

Idea: Apply process mining techniques to low-level logs of software-user interactions to learn about behavior patterns

Rehse, J.-R., Abb, L., Berg, G., Bormann, C., Kampik, T., & Warmuth, C. (2024): User Behavior Mining: A Research. Business Information Systems Engineering (online first).

Conformance Checking in Practice

Idea: Increase the practical value of conformance checking for managing the compliance of business processes

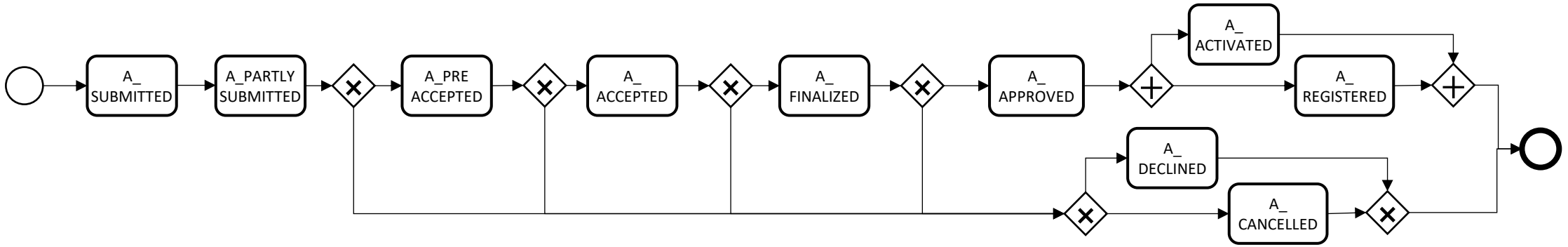
Grohs, M., Pfeiffer, P., & Rehse, J.-R. (2023): Business Process Deviation Prediction: Predicting Unwanted Process Behavior. In International Conference on Process Mining (pp. 113-120). IEEE.

BPM meets AI

Idea: Explore the opportunities that novel AI techniques hold for business process management

Dumas, M., Fournier, F., Limonad, L., Marrella, A., Montali, M., Rehse, J.-R., & others (2023): Augmented Business Process Management Systems: A Research Manifesto. ACM Transactions on Management Information Systems 14 (1), 1-19.

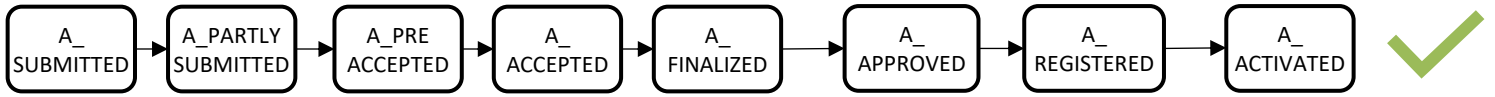
Conformance Checking



*Do **process executions** (traces) **conform** with the process model?*

Trace t_1

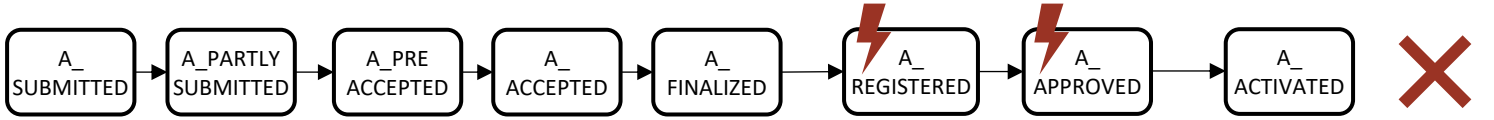
10,000€ Requested Amount



Deviation d_1: A_APPROVED missing **Deviation d_2**: A_APPROVED wrongly executed

Trace t_2

25,000€ Requested Amount



Reactive vs. Proactive Conformance Checking

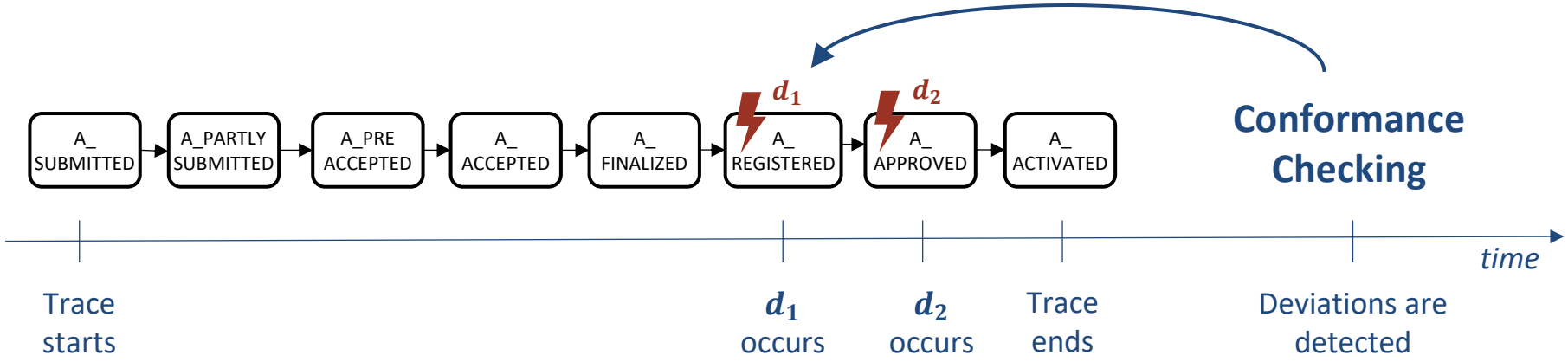
UNIVERSITY OF MANNHEIM
Business School

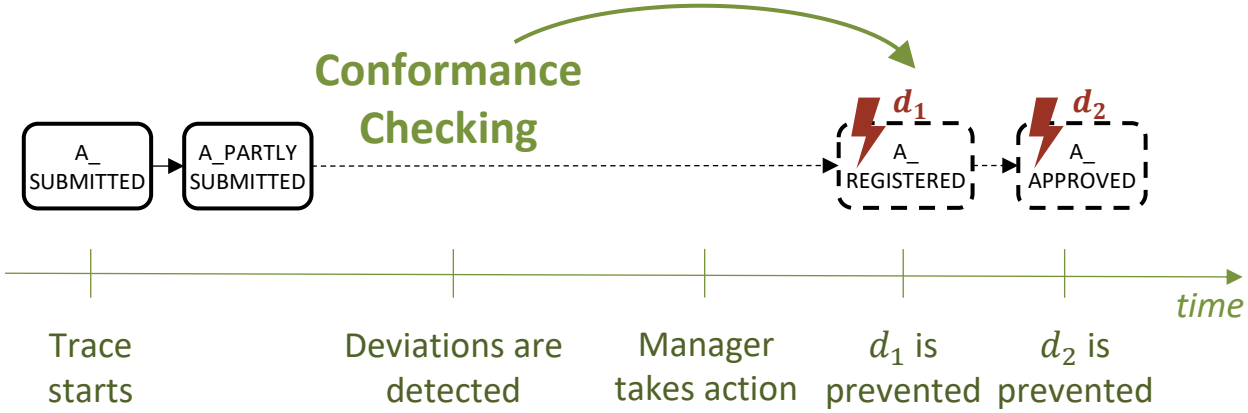
Currently: Reactive

Conformance checking can only detect deviations **after** they occur, so they cannot be prevented.

Vision: Proactive

Conformance checking should detect deviations **before** they occur, so they can be prevented.





Application of (Supervised) Machine Learning

Prof. Dr. Jana-Rebecca Rehse

07.11.2024

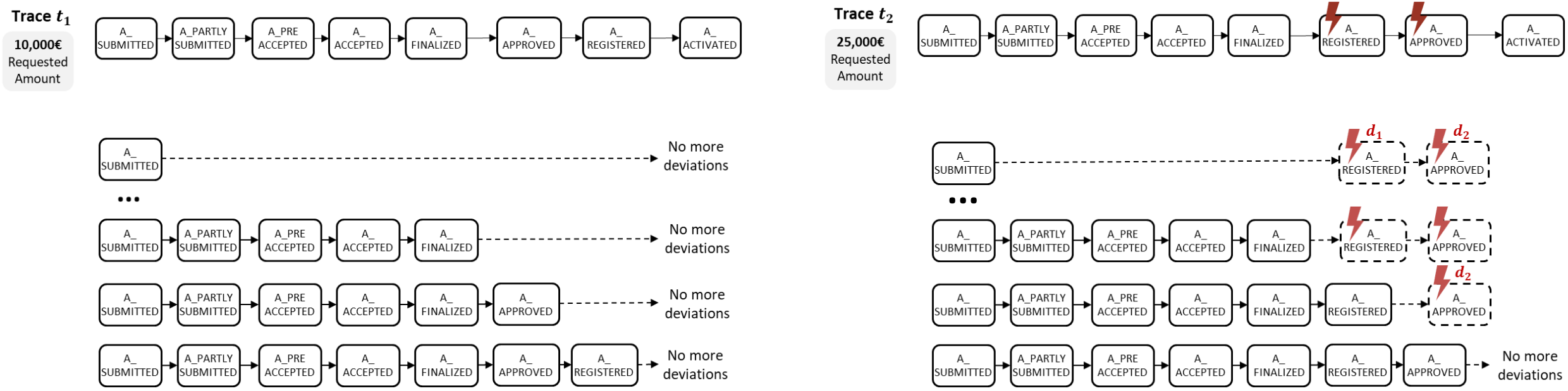
Cf. Grohs, M., Pfeiffer, P., Rehse, J.-R. (2025): Proactive Conformance Checking: An Approach for Predicting Deviations in Business Processes, Information Systems 127, 102461

Business Process Deviation Prediction

UNIVERSITY OF MANNHEIM
Business School

Goal

Apply supervised machine learning to incomplete traces to predict which deviations will occur in the future



Challenges of Deviation Prediction

UNIVERSITY OF MANNHEIM
Business School

1 Explicit Process Knowledge

Assigning labels (i.e., occurring deviations) requires a prescriptive **process model**	→	**Supervised ML approach**, trained with data obtained from comparing log and model

2 Prediction Labels

Multiple (a trace may deviate more than once)	→	**Separate** (binary) ML model per deviation
Dynamic (labels change over trace duration)		Each trace prefix labeled **individually**
Imbalanced (deviations occur infrequently)		**Undersampling** applied to training data

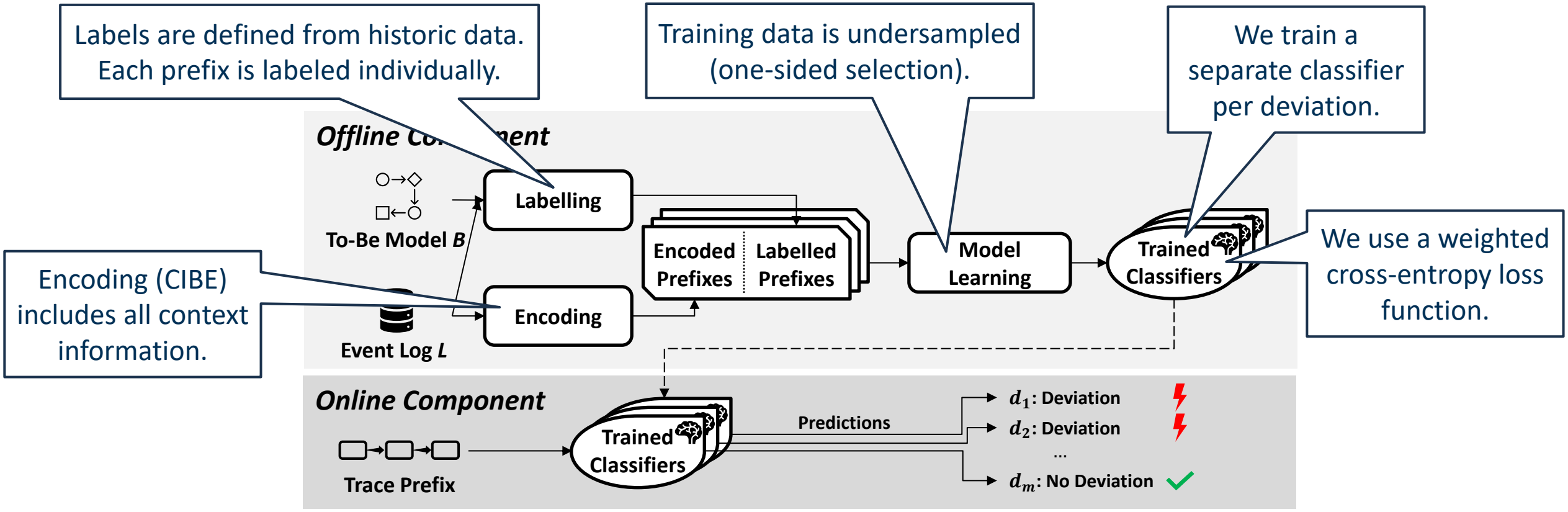
3 Context Importance

Deviations in the control flow may be caused by **context attributes** (e.g., amount)	→	Traces are encoded with all relevant context attributes (**Complex Index-Based Encoding**)

4 Action Orientation

High recall (recognize deviations) preferred over high precision (not misclassifying conform traces)	→	**Weighted cross-entropy loss function** that reduces false negatives (missed deviations)

Our Approach



Evaluation

Alternative prediction approach not based on ML

Alternative ML-based prediction approach

Our approach

Logs	Genga et al. (2017)	CatBoost	BPDP
BPI12 A			
BPI12 O			
BPI20 Dom.			
BPI20 Int.			
BPI20 RfP			
BPI20 Prep.			
MobIS			

Public event log data used as benchmarks

AUC_{ROC} = Area under the curve of the receiving operator characteristic

- Standard evaluation metric for binary classifications
- Relates true positive to false positive rate for different thresholds
- Varies between 0 and 1, with 0.5 representing a random classifier
- Values > 0.7 are acceptable, values > 0.8 are excellent

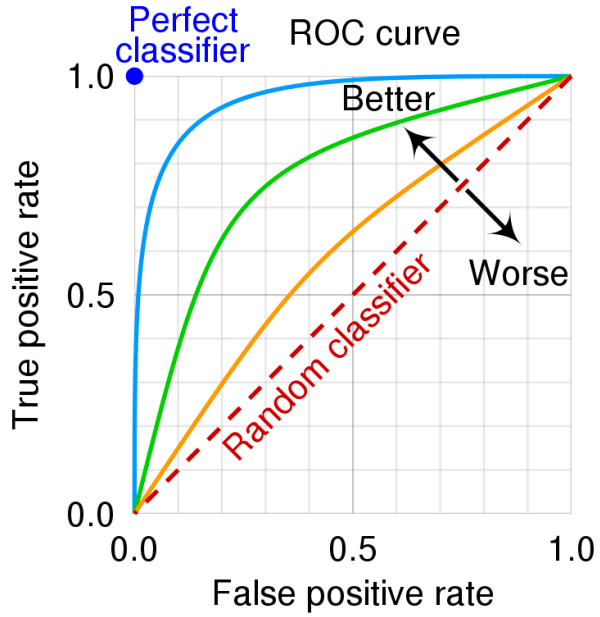


Figure: https://commons.wikimedia.org/wiki/File:Roc_curve.svg

Evaluation

BPDP outperforms alternative approaches

Logs	Genga et al. (2017)	CatBoost	BPDP
BPI12 A	0.5166	0.6954	**0.7324**
BPI12 O	0.4872	0.6591	**0.6665**
BPI20 Dom.	0.6372	0.7511	**0.8258**
BPI20 Int.	0.5796	0.6969	**0.7270**
BPI20 RfP	0.5762	0.6961	**0.7620**
BPI20 Prep.	0.5521	0.6333	**0.7040**
MobIS	0.5461	0.5729	**0.6534**

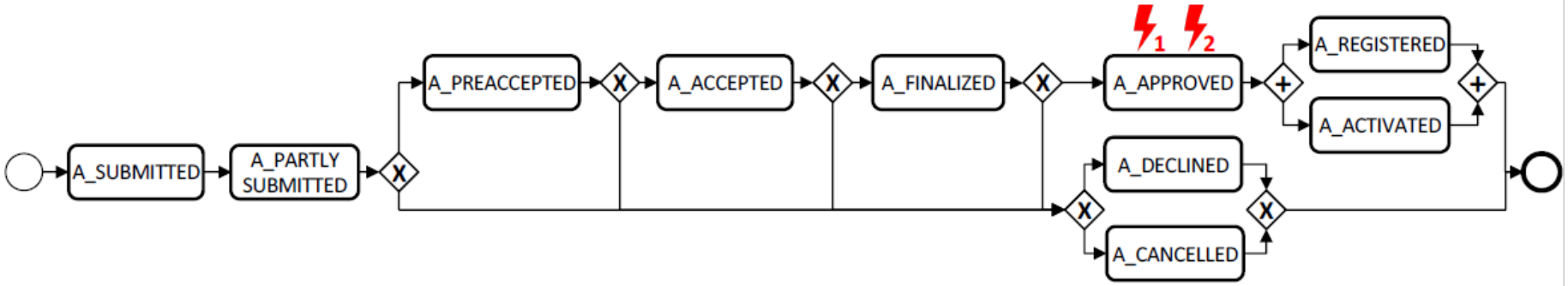
Performance differs between logs

Figure: https://commons.wikimedia.org/wiki/File:Roc_curve.svg

User Output

UNIVERSITY OF MANNHEIM
Business School

Process



Selected case	12012237	*select other case*

Customer	Mrs. Jane Doe	**Loan amount requested**	50,000€
Current state	A_SUBMITTED → A_PARTLY SUBMITTED		

Deviation Predictions	**Pred. Probability**	**Top Features for Prediction**
1 **A_APPROVED** will be **skipped**	80%	➢ **High loan amount, Resource 112** at event 2
2 **A_APPROVED** will be **inserted**	85%	➢ **Resource 45** at event 1, **high loan amount**

Towards Proactive Conformance Checking

Contributions	
	Novel machine learning approach for predicting deviations in the future of ongoing traces
	Outperforms existing state-of-the-art by considering the specific challenges of deviation prediction
	Equipped with graphical user interface as well as multiple features to support process managers

Remaining Research Questions	
	How can we increase the precision and overall accuracy of our predictive models?
	How can we consider deviation patterns, i.e., sets of frequently occurring deviations?
	How can we establish true causes for occurring deviations?

Thank you!

Questions?

Jun.-Prof. Dr.
Jana-Rebecca Rehse

Juniorprofessur für Management Analytics
Universität Mannheim
rehse@uni-mannheim.de