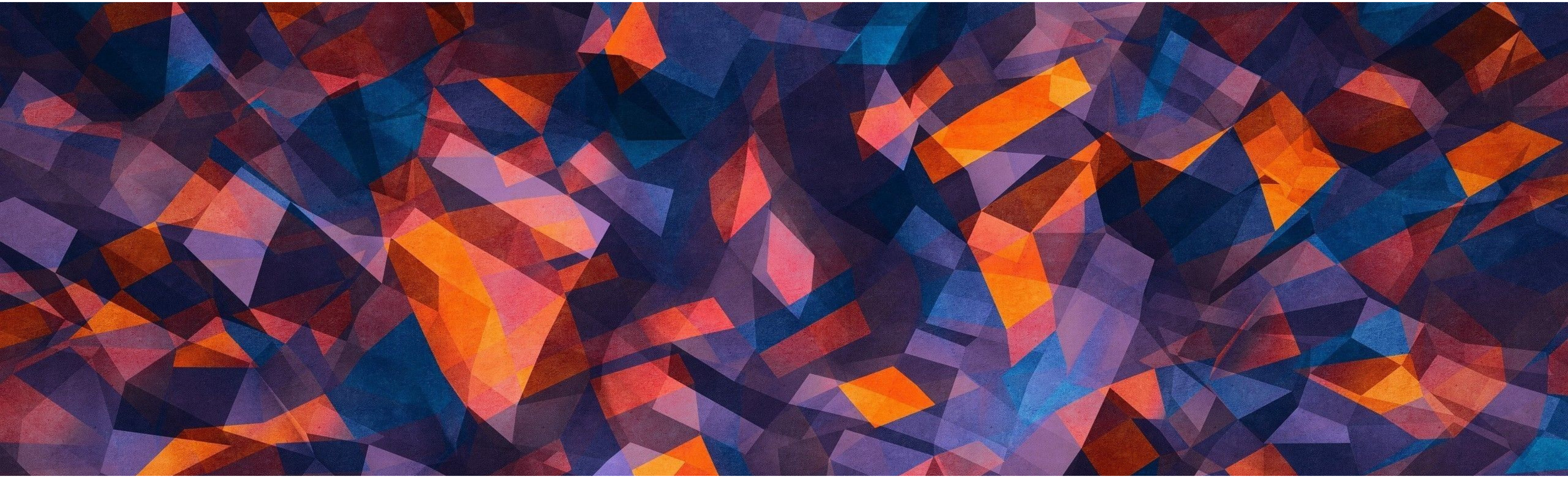
Universität St.Gallen



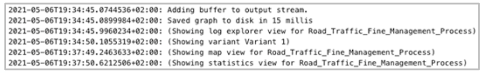
Leveraging Digital Trace Data to Investigate and Support Human-Centered Work Processes

Presented by Barbara Weber

Universität St.Gallen

Digital Trace Data

Digital technologies create an ever-increasing volumne of **digital traces.**



```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```

Observation

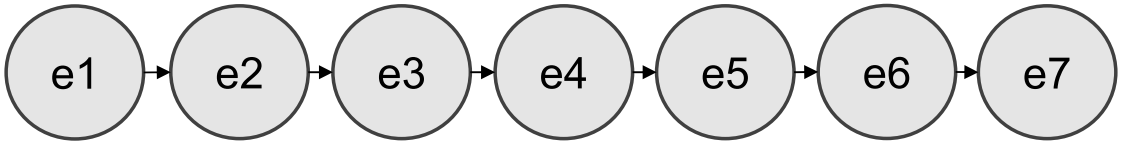
👁 👂 ✋ 👅 👃

Information Systems, Sensors

Multiple events can be related by time, causality, abstraction, or other relationships

Occurence
(Events)

Record about past occurence
Event Sequence



Process Mining on the Rise


EDITORS' PICK | Jun 2, 2021, 06:00am EDT | 63.454 views

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


IBM acquires Italy's myInvenio to integrate process mining directly into its suite of automation tools

Ingrid Lunden @ingridlunden / 2:03 PM GMT+2 • April 15, 2021

Comment


Appian acquires process mining company Lana Labs

Kyle Wiggers @Kyle_L_Wiggers August 5, 2021 2:05 PM


INSIGHTS

SAP to acquire Business Process automation startup Signavio for a reported US$1.2 billion

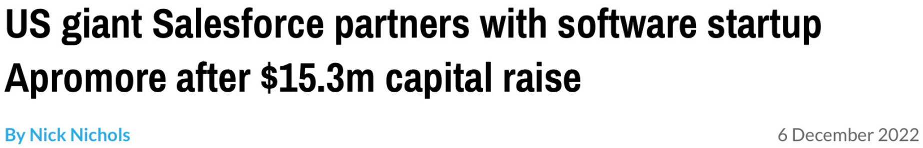
Staff Writers / Thu 28 Jan 2021

Insights > SAP to acquire Business Process automation startup Signavio for a reported US$1.2 billion


Microsoft acquires process mining vendor Minit to grow its automation offerings

Kyle Wiggers @kyle_l_wiggers / 9:05 PM GMT+2 • March 31, 2022

Comment


US giant Salesforce partners with software startup Apromore after $15.3m capital raise

By Nick Nichols

6 December 2022

Magic Quadrant for Process Mining Tools

- Gartner published a market guide for process mining in 2018
- Inaugural publication of a Magic Quadrant for Process Mining tools in 2023

Universität St.Gallen

The Potential of Process Mining

Creation of «current state» processes

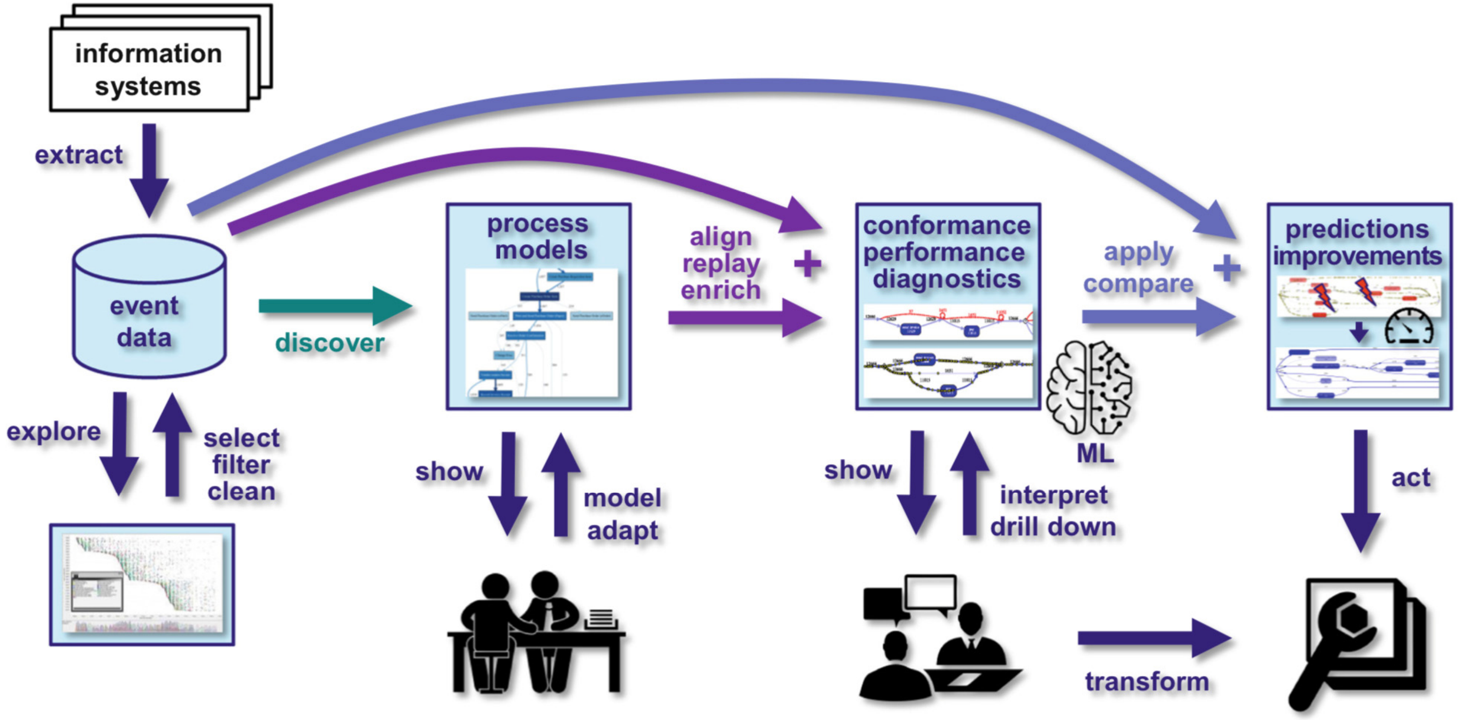
Connecting BPM with Data

„Process mining software can help organizations easily capture information from enterprise transaction systems and provides detailed — and data-driven — information about how key processes are performing.“

Source: Davenport and Spanyi, What Process Mining Is, and Why Companies Should Do It

Universität St.Gallen

Process Mining: The Big Picture



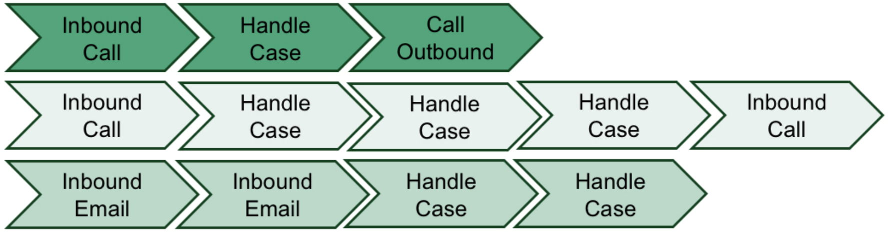
Source: van der Aalst & Carmona: Process Mining Handbook

Event Data: The Starting Point for Process Mining

Case	Activity	Start Date	End Date
Case 17	Inbound Call	04.03.2010 07:35	04.03.2010 07:46
Case 17	Handle Case	04.03.2010 07:53	04.03.2010 07:55
Case 17	Handle Case	08.03.2010 11:16	08.03.2010 11:18
Case 1	Inbound Call	09.03.2010 08:05	09.03.2010 08:10
Case 1	Handle Case	11.03.2010 10:30	11.03.2010 10:32
Case 17	Handle Case	11.03.2010 11:15	11.03.2010 11:19
Case 1	Call Outbound	11.03.2010 11:45	11.03.2010 11:52
Case 19	Inbound Email	14.03.2010 14:08	18.03.2010 08:04
Case 17	Inbound Call	14.03.2010 17:53	14.03.2010 17:56
Case 19	Inbound Email	18.03.2010 08:06	18.
Case 19	Handle Case	18.03.2010 08:07	18.
Case 19	Handle Case	18.03.2010 08:09	18.

An event log contains traces
Each trace is a sequence of events belonging to the same case

Traces of Case 1, 17 and 19



Inbound Call → Handle Case → Call Outbound

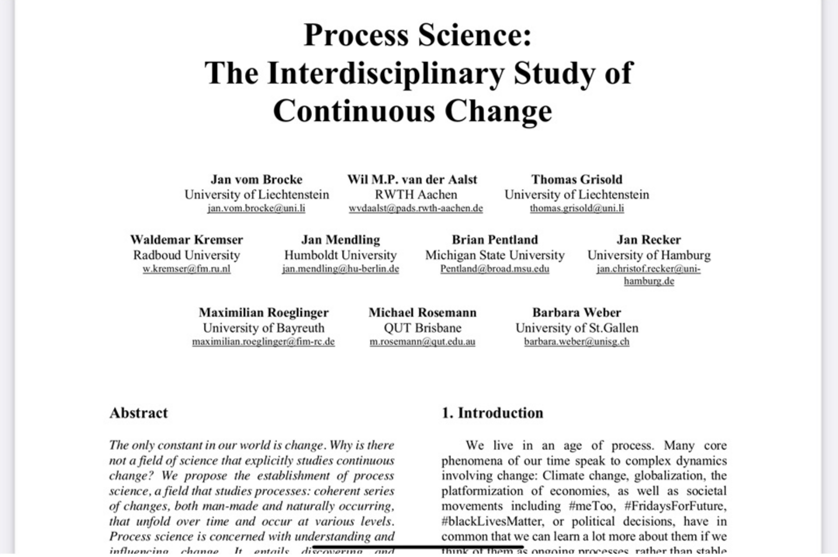
Inbound Call → Handle Case → Handle Case → Handle Case → Inbound Call

Inbound Email → Inbound Email → Handle Case → Handle Case

Case ID, activity and at least one timestamp per event are the minimum requirements for an event log

Process Science

Event Data for Studying Continuous Change



Process Science: The Interdisciplinary Study of Continuous Change

Jan vom Brocke
University of Liechtenstein
jan.vom.brocke@uni.li

Wil M.P. van der Aalst
RWTH Aachen
wvdaalst@pads.rwth-aachen.de

Thomas Grisold
University of Liechtenstein
thomas.grisold@uni.li

Waldemar Kremser
Radboud University
w.kremser@fm.ru.nl

Jan Mendling
Humboldt University
jan.mendling@hu-berlin.de

Brian Pentland
Michigan State University
Pentland@broad.msu.edu

Jan Recker
University of Hamburg
jan.christof.recker@uni-hamburg.de

Maximilian Roeglinger
University of Bayreuth
maximilian.roeglinger@fim-rc.de

Michael Rosemann
QUT Brisbane
m.rosemann@qut.edu.au

Barbara Weber
University of St.Gallen
barbara.weber@unisg.ch

Abstract

The only constant in our world is change. Why is there not a field of science that explicitly studies continuous change? We propose the establishment of process science, a field that studies processes: coherent series of changes, both man-made and naturally occurring, that unfold over time and occur at various levels. Process science is concerned with understanding and

1. Introduction

We live in an age of process. Many core phenomena of our time speak to complex dynamics involving change: Climate change, globalization, the platformization of economies, as well as societal movements including #meToo, #FridaysForFuture, #blackLivesMatter, or political decisions, have in common that we can learn a lot more about them if we

> *“Process science is the interdisciplinary study of continuous change. By process, we mean a coherent series of changes that unfold over time and occur at multiple levels.“*

Digital trace data offer new opportunities to study **how phenomena evolve** in terms of underlying **sequences of events**.

vom Brocke et al., Process Science: The Interdisciplinary Study of Continuous Change

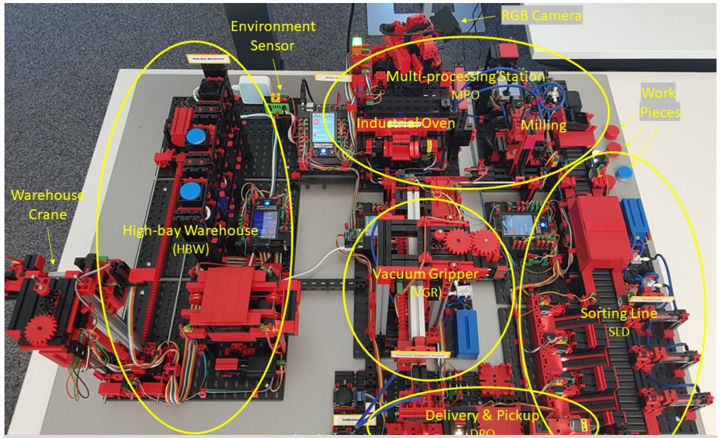
Universität St.Gallen

Process Science Activities

	Discovery	Explanation	Intervention
	Goal. Capture and describe processes.	**Goal.** Understand why, how and when a process unfolds.	**Goal.** Intervene and shape the process into desired directions.
Methods	*Example.* Methods to create process representations from digital trace data and to identify patterns in processes.	*Example.* Methods to study the context in which a pattern is situated.	*Example.* Methods to develop and evaluate interventions.

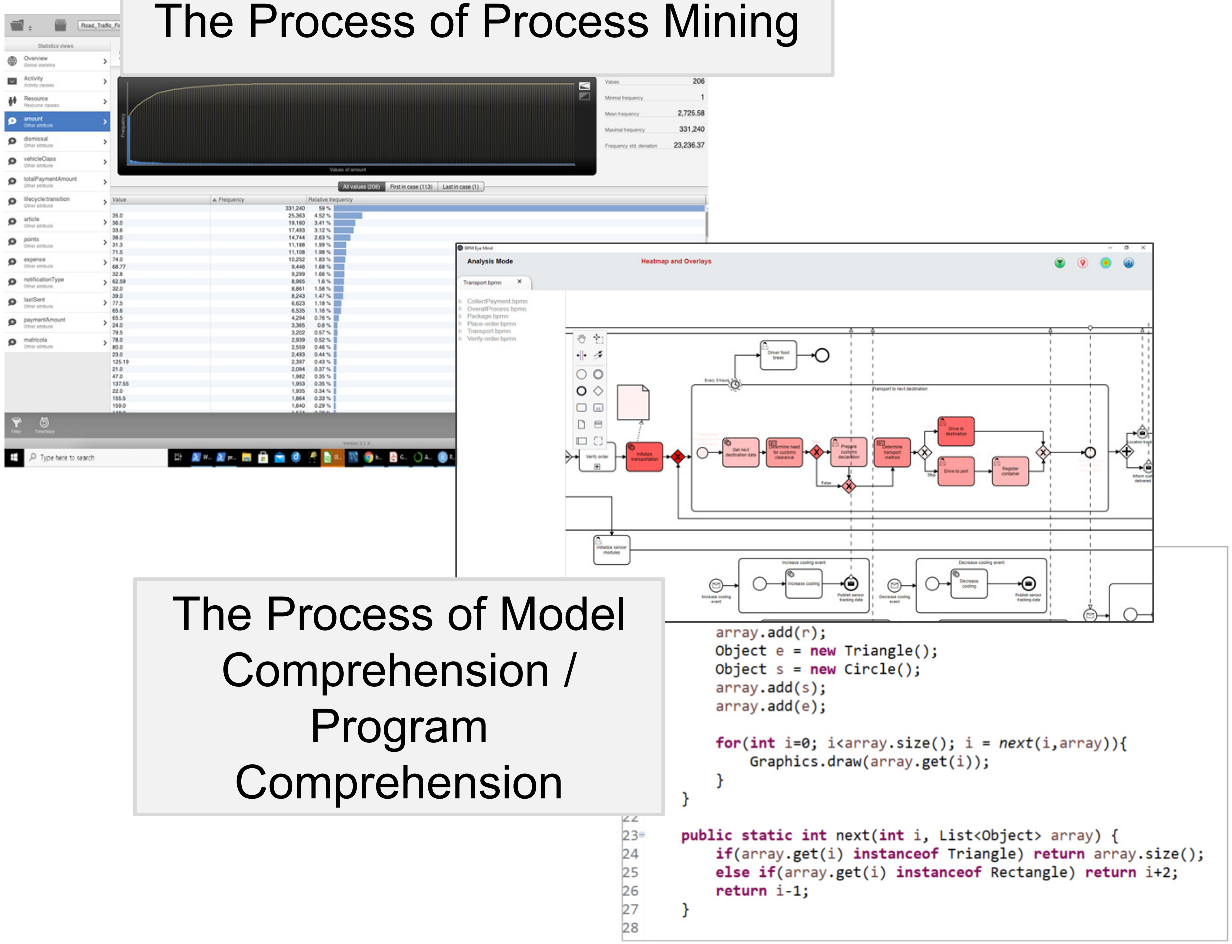
Universität St.Gallen

Examples of Different Processes



The Process of Storage and Production in a Smart Factory

Phlebotomy: The Process of Drawing Blood



The Process of Process Mining

The Process of Model Comprehension / Program Comprehension

```
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object> array) {
        if(array.get(i) instanceof Triangle) return array.size();
        else if(array.get(i) instanceof Rectangle) return i+2;
        return i-1;
    }
```

Universität St.Gallen

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery and Exploration

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Universität St.Gallen

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Process Observability Largely Differs

Example event from factory:

Topic: FTFactory/HBW_1

```
{  "id": "FTFactory/HBW_1",  "timestamp": "2020-
12-11 13:35:35.50",  "i1_light_barrier_interrupted":
false,  "i2_light_barrier_interrupted": true,
"i3_light_barrier_interrupted": true,
"i4_light_barrier_interrupted": false,
"i5_position_switch_pressed": true,
"i6_position_switch_pressed": true,
"i7_position_switch_pressed": false,
"i8_position_switch_pressed": true,  "m1_speed": 0,
"m2_speed": 0,  "m3_speed": 0,  "m4_speed": 0,
"current_state": "ready",  "current_task": "",
"current_task_elapsed_seconds_since_start": 0,
"current_sub_task": "",  "failure_label": "",
"current_pos_x": 0,  "current_pos_y": 0,
"target_pos_x": 0,  "target_pos_y": 0,
"amount_of_stored_workpieces": 0}
```

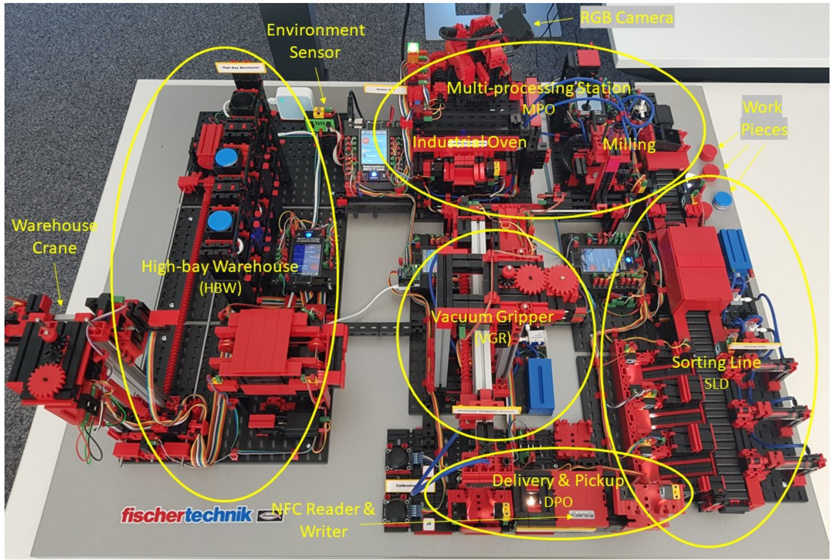
Sensors:

- Switches
- Light barriers
- Color sensors
- Environment
- Camera
- NFC

Actuators:

- Motors
- Compressors
- Valves

Smart Factory equipped with **sensors** and **actuators** emitting **events**



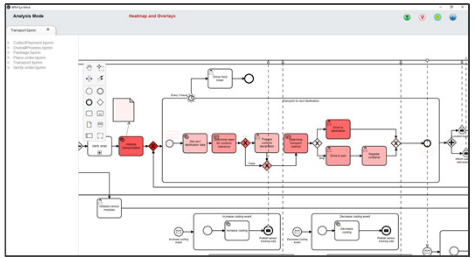
Smart Factory @ UNISG

Low

High

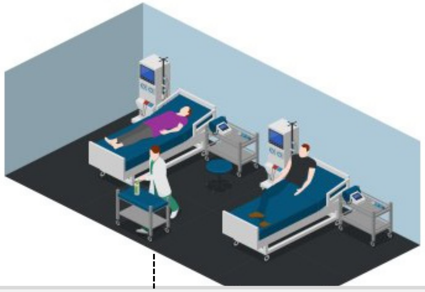
Source: R. Seiger, L. Malburg, B. Weber, R. Bergmann, Integrating process management and event processing in smart factories: A systems architecture and use cases.

Process Observability Largely Differs

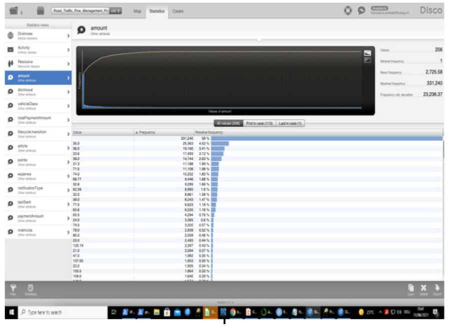


Navigation, scrolling and zooming events during model comprehension (depending on tool); large parts of the process occur in the **reader's mind**

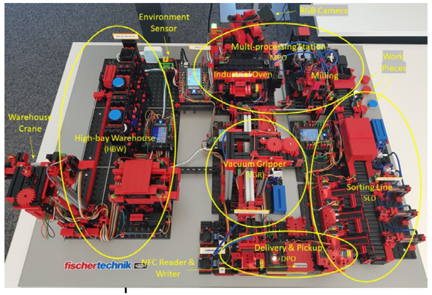
Process is largely manual; **no events** since most parts performed **outside of any IT system**



Tool interaction events during analysis (depending on the tool); large parts of the process occur in the **analyst's mind**



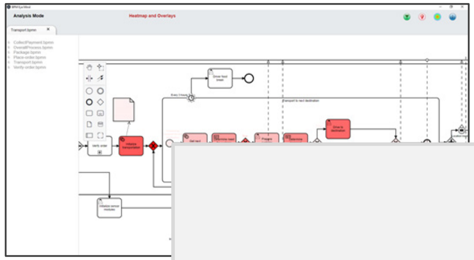
Sensors and **actuators** emitting **events**



Low

High

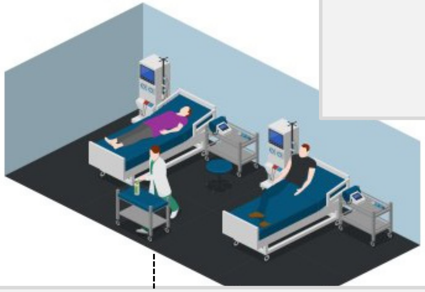
Process Observability Largely Differs

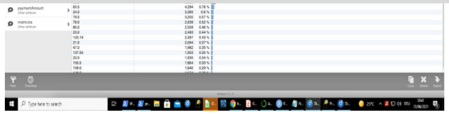


Navigation, scrolling and zooming events during model comprehension (depending on tool): large parts of the

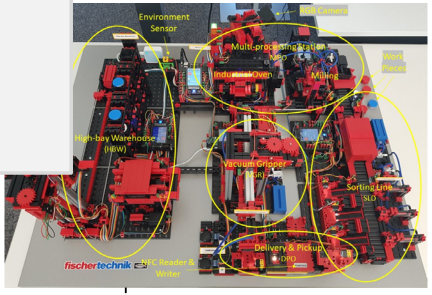
Usage of sensors and additional forms of data collection to increase process observability.

Process is largely m
no events since mo
performed **outside o**
system





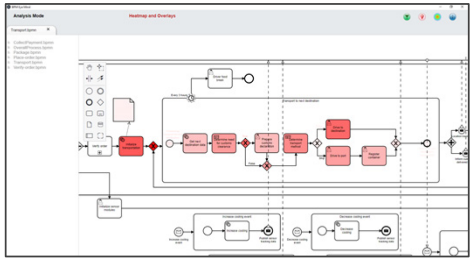
nsors and
tuators emitting
ents



Low

High

Process Observability Largely Differs



Navigation, scrolling and zooming events during model comprehension (depending on tool); large parts of the process occur in the **reader's mind**

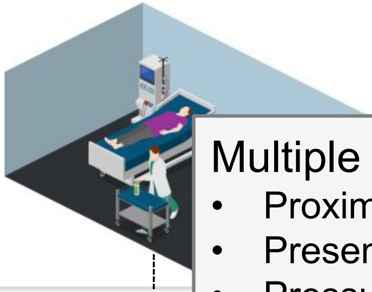
Eye-tracker

Galvanic Skin Response

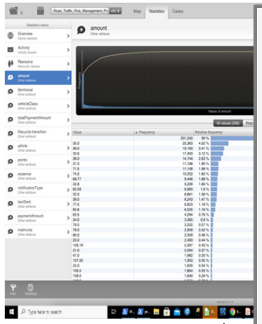
Tool interaction events during analysis (depending on the tool); large parts of the process occur in the **analyst's mind**

Sensors and **actuators** emitting **events**

Process is largely manual; **no events** since most parts performed **outside of any IT system**



Multiple Sensors
- Proximity sensor
- Presence sensor
- Pressure sensor
- Touch sensor
- Flow sensor

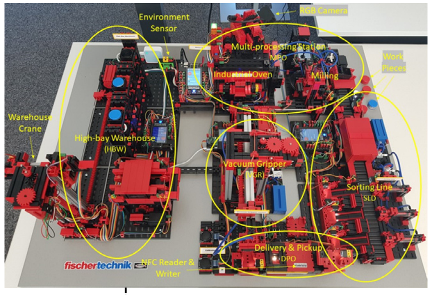


Application logs
(where available)

Screen recordings
to derive user interactions

Think-aloud data

Retrospective Interviews



Low

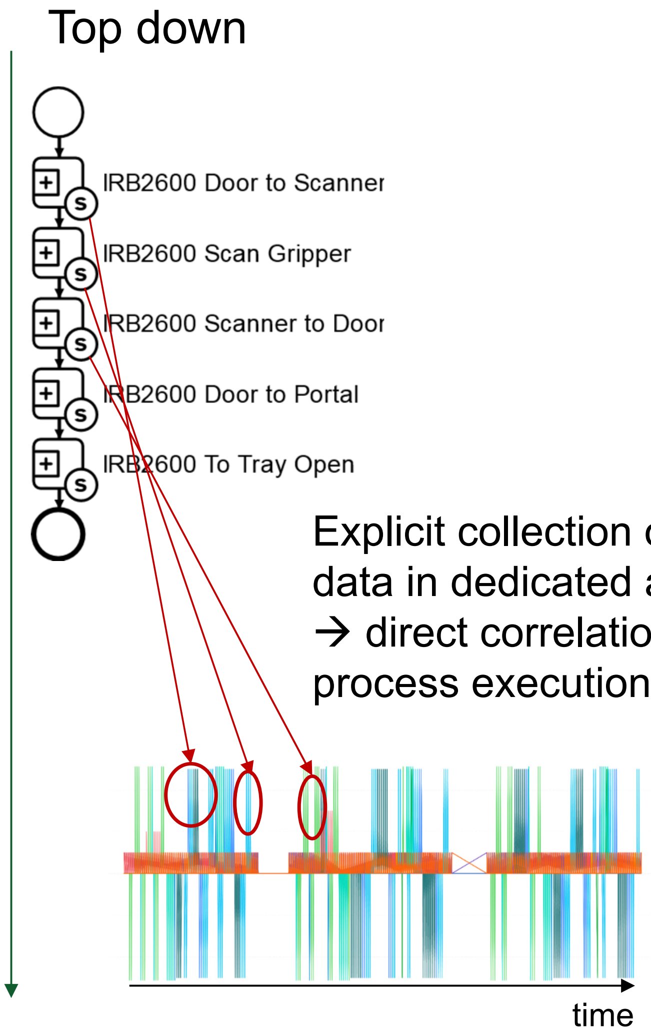
High

The Importance of Data Collection

- Data collection needs to be carefully planned to enable the linking of the collected data with the different elements of the process
  - Collecting data in a process context
  - Synchronized data collection

Universität St.Gallen

Process-driven Execution and Collection of IoT Data in Context



Explicit collection of sensor data in dedicated activities → direct correlation with process execution

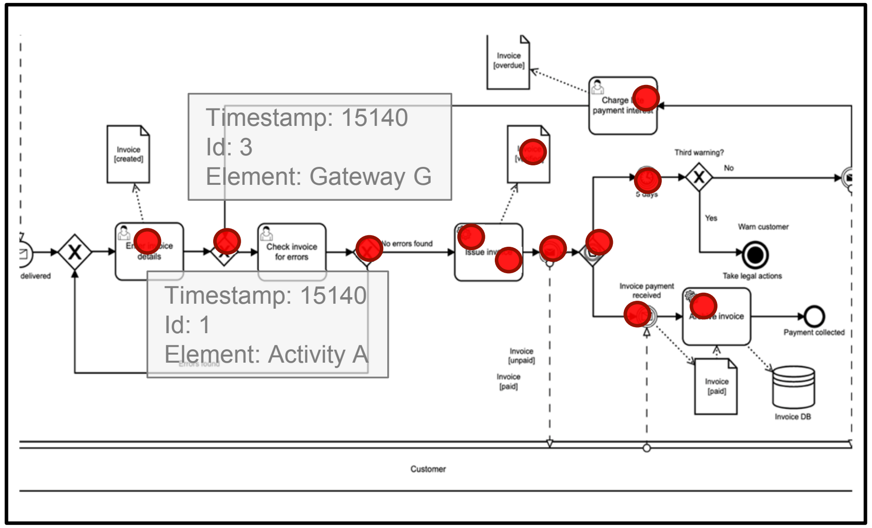
- **IoT data** is collected during process execution and gets **embedded in the broader process context**
- This results in **IoT-enriched event logs** which associate sensor data with the corresponding process execution events

Source: Mangler et al., DataStream XES Extension: Embedding IoT Sensor Data into Extensible Event Stream Logs

Automated Mapping of Attentional Processes to Software Design Artifacts

Eye-tracker

Data collection



Timestamp: 15140
Id: 3
Element: Gateway G

Timestamp: 15140
Id: 1
Element: Activity A

Automated mapping of gazes to elements of the artifact

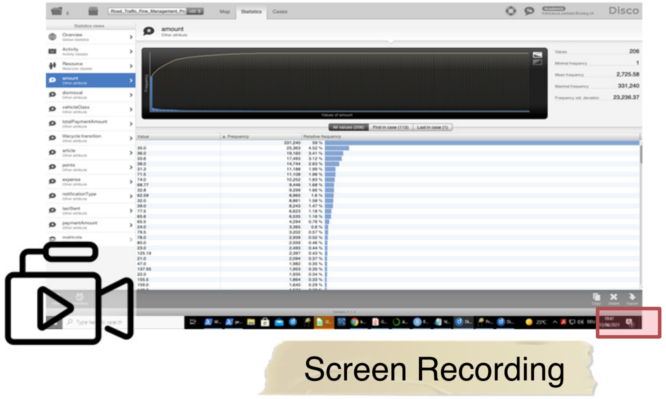
- **Gaze events** are automatically mapped during data collection to the elements of the software design artifact (here process model)
- This results in an **enriched log of gaze events** which associates gaze data with the corresponding elements of the artifact

Source: Prototype developed by Amine Abbad Andaloussi

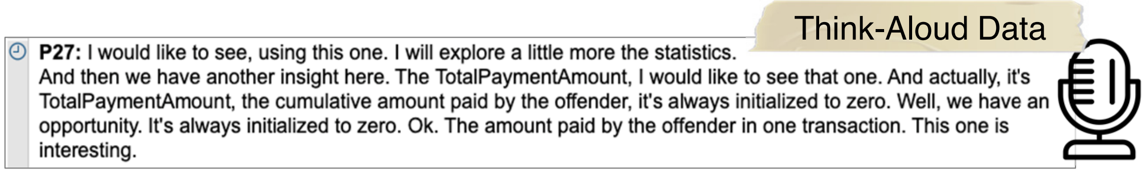
Synchronized Collection of Data

Collecting data in a process context is not always feasible.

If collected in a synchronized manner, links between different modalities can be established at later stages, e.g., using timestamps.

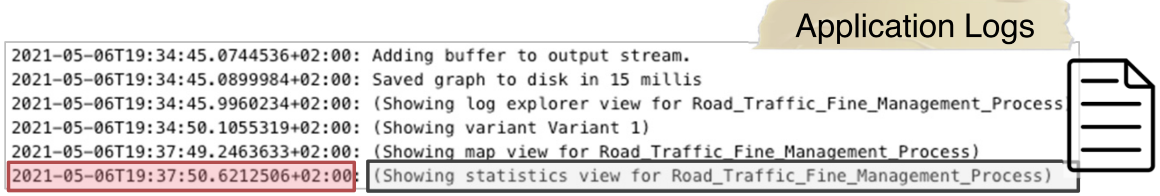


Screen Recording



Think-Aloud Data

P27: I would like to see, using this one. I will explore a little more the statistics. And then we have another insight here. The TotalPaymentAmount, I would like to see that one. And actually, it's TotalPaymentAmount, the cumulative amount paid by the offender, it's always initialized to zero. Well, we have an opportunity. It's always initialized to zero. Ok. The amount paid by the offender in one transaction. This one is interesting.



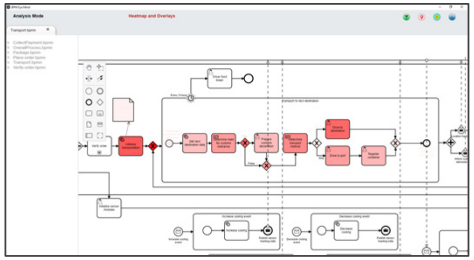
Application Logs

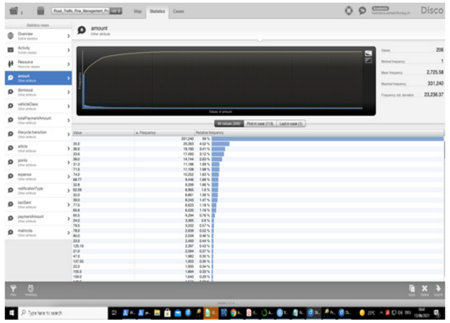
```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```

Universität St.Gallen

Availability of Process Knowledge

Process and **activities** largely **unknown;** high flexibility and variability



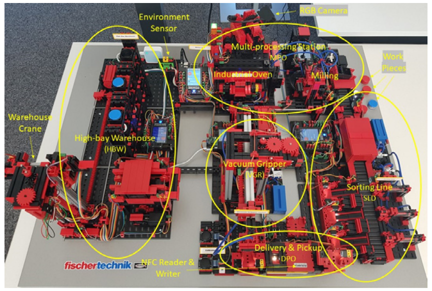


Guidelines including **process steps;** indication for hand hygiene (**business rules**); some flexibility and variability

WHO guidelines on drawing blood:

best practices in phlebotomy

Process and **activities known**; end-to-end visibility due to process-driven execution and data collection; repetitive and well structured



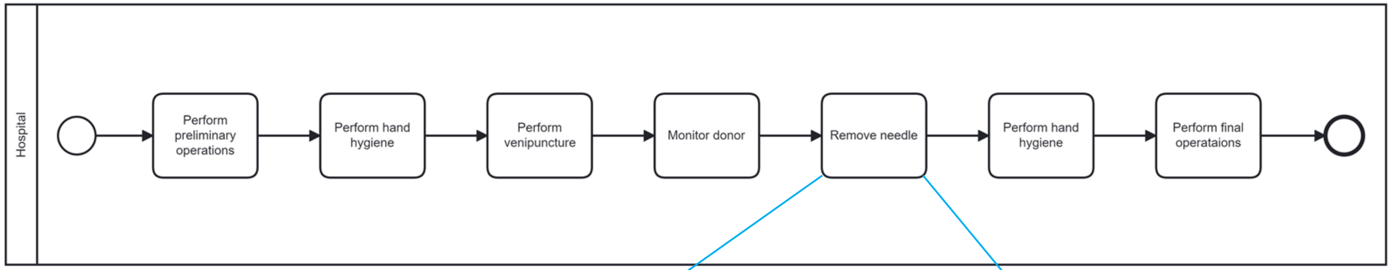
Low → High

Process Activity / Event Detection from Sensors

Known process and activities

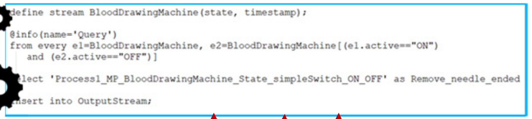
WHO guidelines on drawing blood:

best practices in phlebotomy

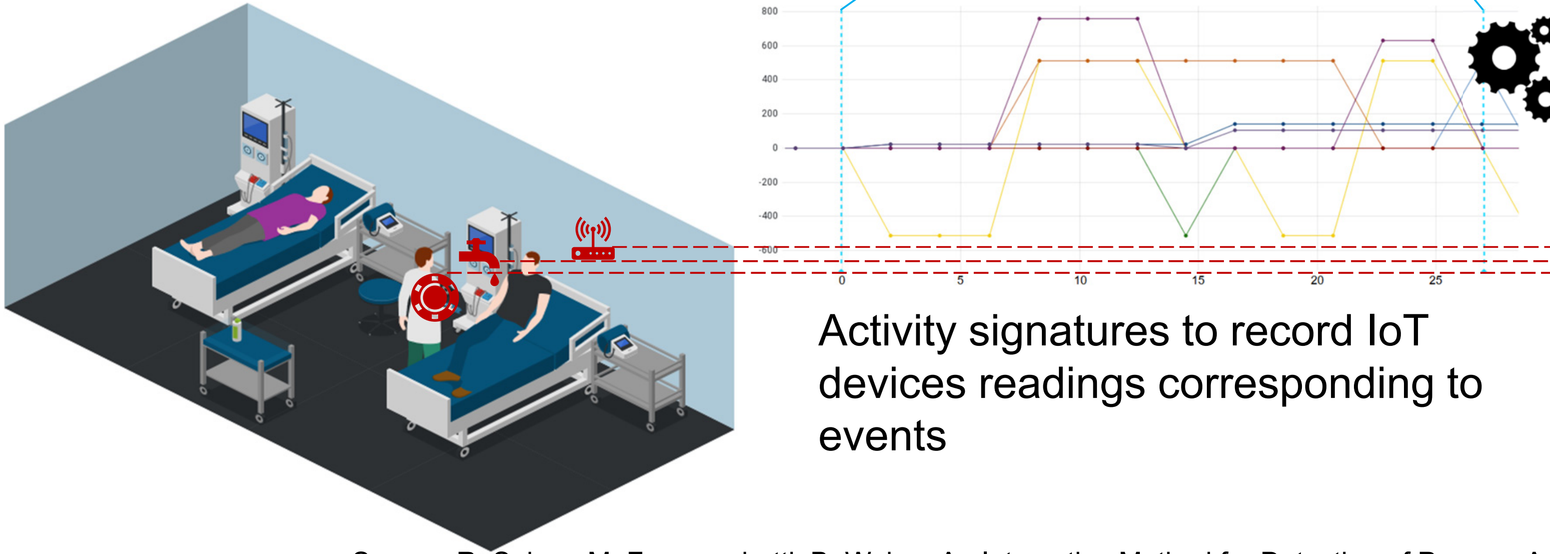


IoT to detect relevant state changes

Complex Event Processing queries



```
define stream BloodDrawingMachine(state, timestamp);

@info(name='Query')
from every e1=BloodDrawingMachine, e2=BloodDrawingMachine[(e1.active=="ON")
   and (e2.active=="OFF")]

elect 'Process1_MP_BloodDrawingMachine_State_simpleSwitch_ON_OFF' as Remove_needle_ended

nsert into OutputStream;
```

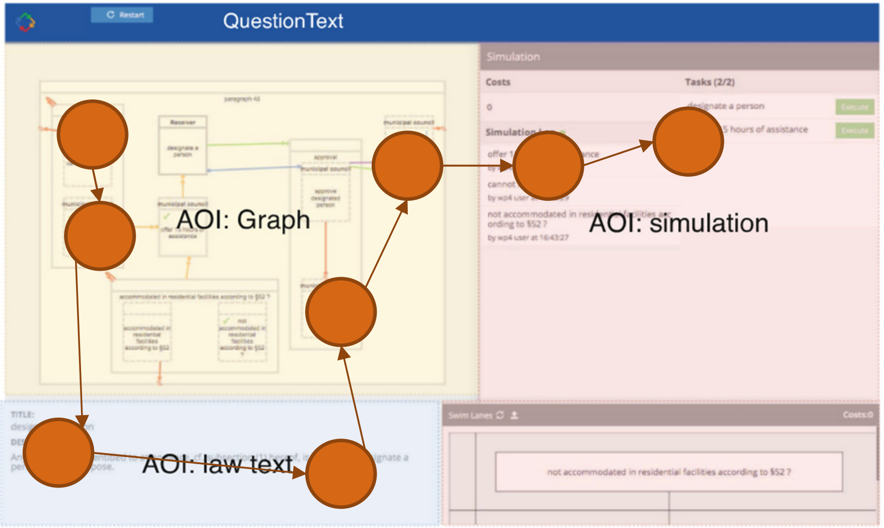


Activity signatures to record IoT devices readings corresponding to events

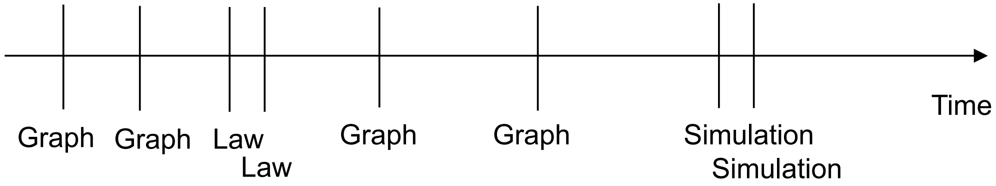
Source: R. Seiger, M. Franceschetti, B. Weber, An Interactive Method for Detection of Process Activity Executions from IoT Data

Areas of Interest as Activity Proxies

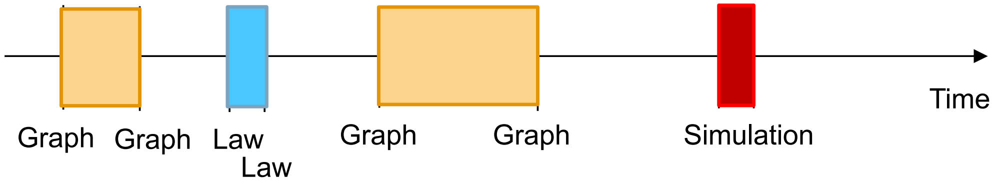
- Each comprehension task performed by a participant (i.e., trial) is considered a process instance
- Visits to Areas of Interest (corresponding to elements of the artifact) are used as proxies for activities



Fixations with associated AOI



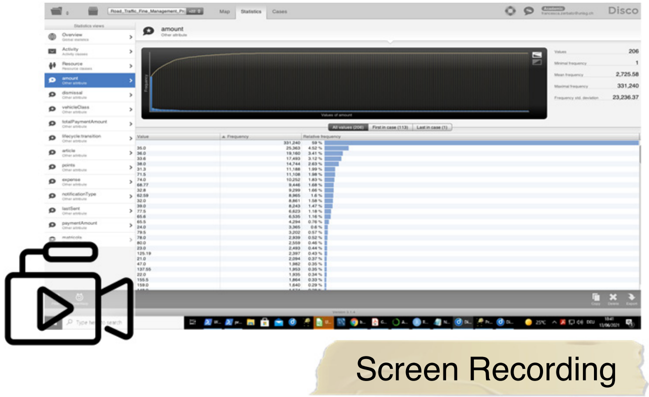
Event Sequence as Sequence of AOI Visits



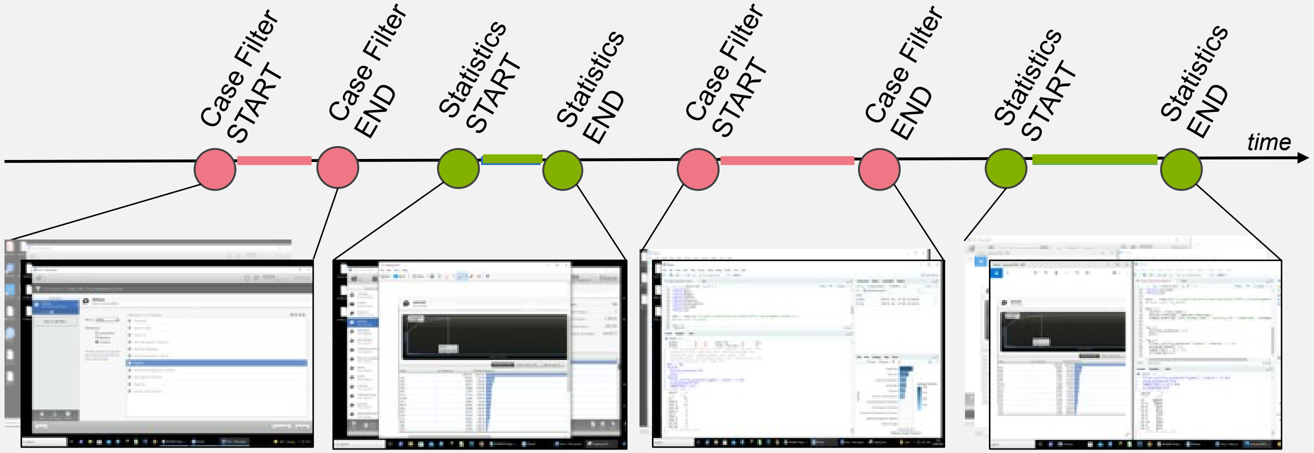
Source: Abbad Andaloussi et al., Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Universität St.Gallen

Creating User Interaction Logs From Screen Recordings



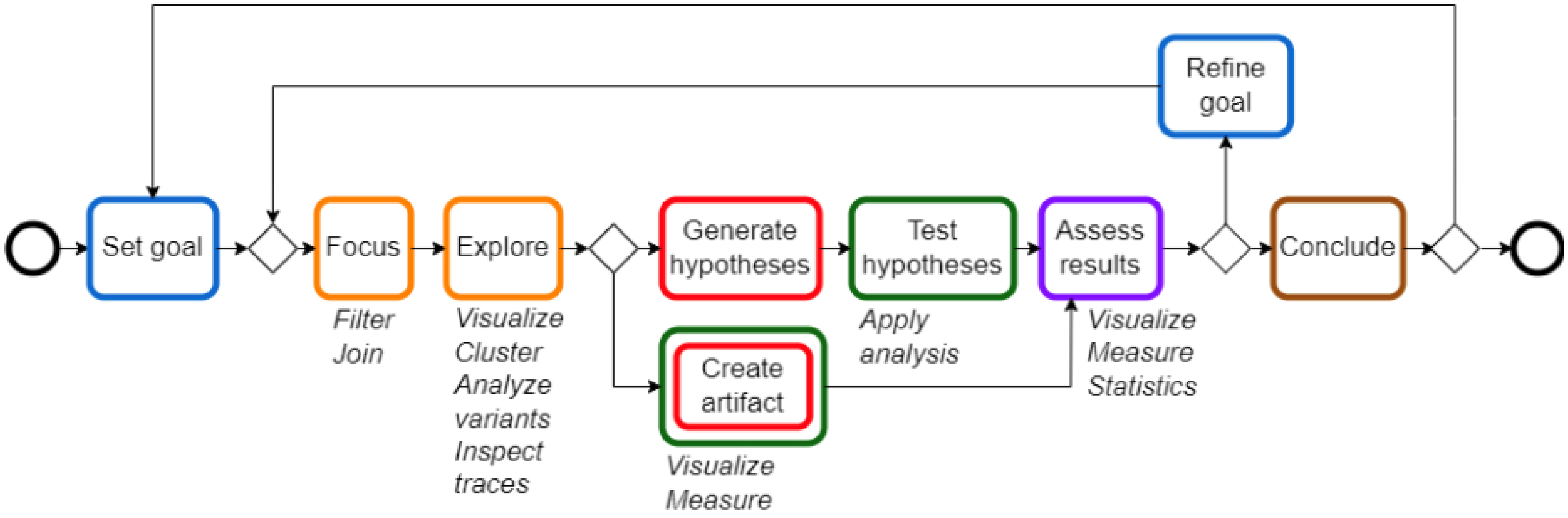
Since processes and activities are largely unknown, decisions on **what to consider as events** is left to the researcher.



I	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:16:37,1	00:16:59,8
P27	Statistics	Disco	00:16:37,1	00:16:59,8
…	…	…	…	…

Guiding Log Creation with Procss Knowledge

- PEM4PPM Model based on Prediction Error Minimization Theory (PEM)
- PEM4PPM activities can be used for log creation



Color legend: Handle goal Create attention Create prediction Test prediction Minimize error Act

Source: P. Soffer and I. Hadar: Israel Science Foundation project under grant agreement 2005/21

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery and Exploration

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

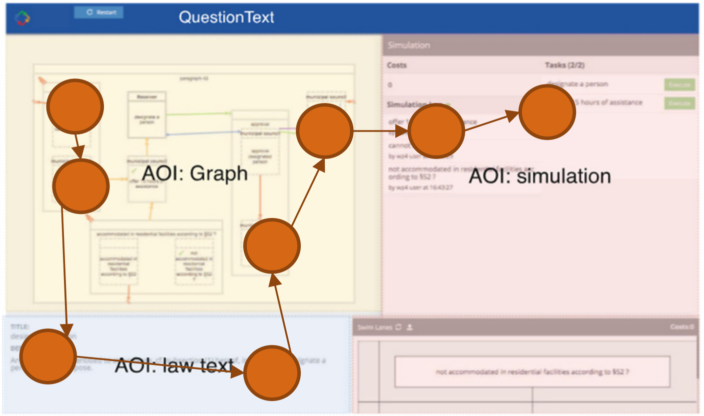
Conformance Checking

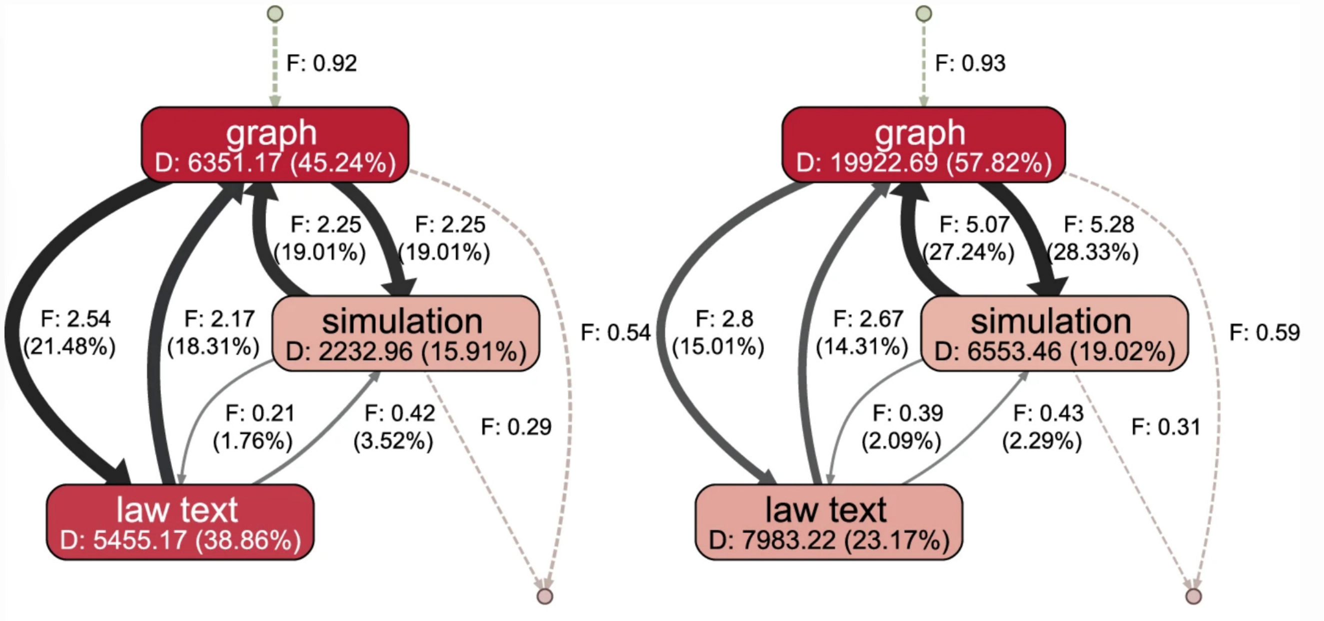
Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Mining User Behavior Patterns Example: Hybrid Process Artifacts



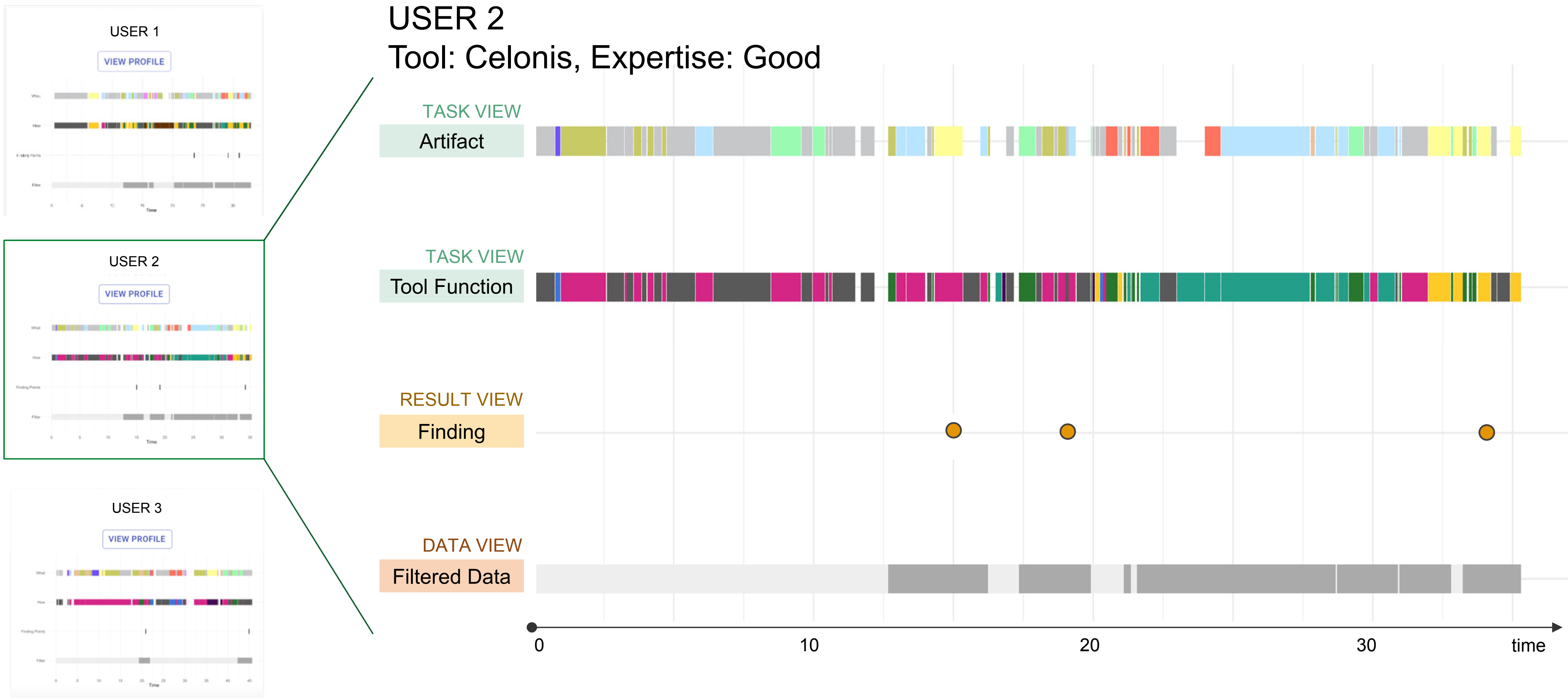


(a) Municipal employees **(b)** Academics

Attention maps in form of Directly-Follow-Graphs comparing the attentional processes for municipal employees and academics. *D* is the mean fixation duration, and *F* is the mean transition frequency between two AOIs.

Source: Abbad Andaloussi et al., Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Visualizing Event Sequences
Creation of Multi-Perspective Profiles



USER 2
Tool: Celonis, Expertise: Good

Visualizing Event Sequences

Focus on Subsequences of Interest

1. Offender is ignoring the fine and not paying

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery and Exploration

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

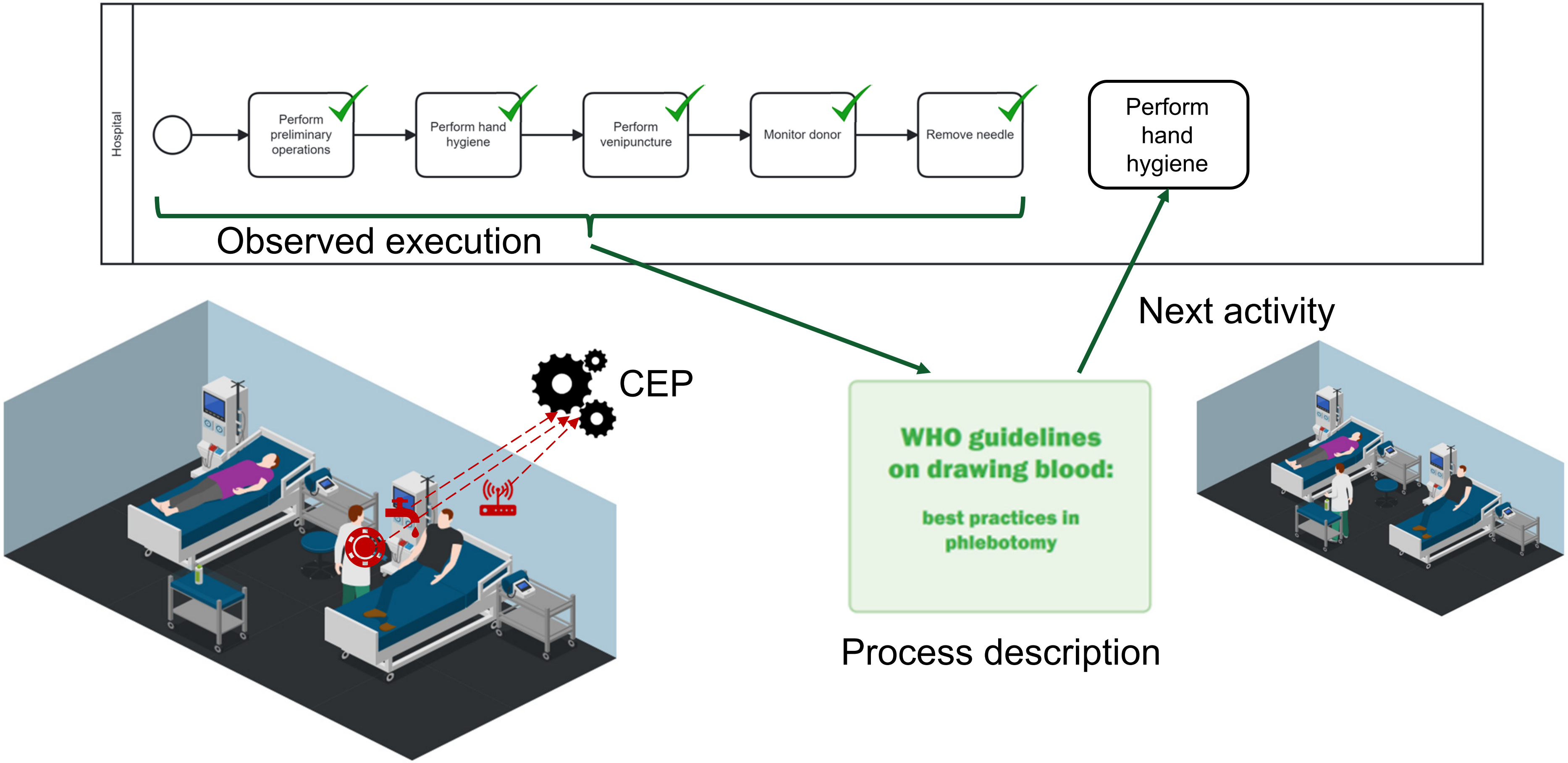
Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Monitoring for Hand Hygiene Indications



Hospital

Perform preliminary operations ✓

Perform hand hygiene ✓

Perform venipuncture ✓

Monitor donor ✓

Remove needle ✓

Perform hand hygiene

Observed execution

Next activity

CEP

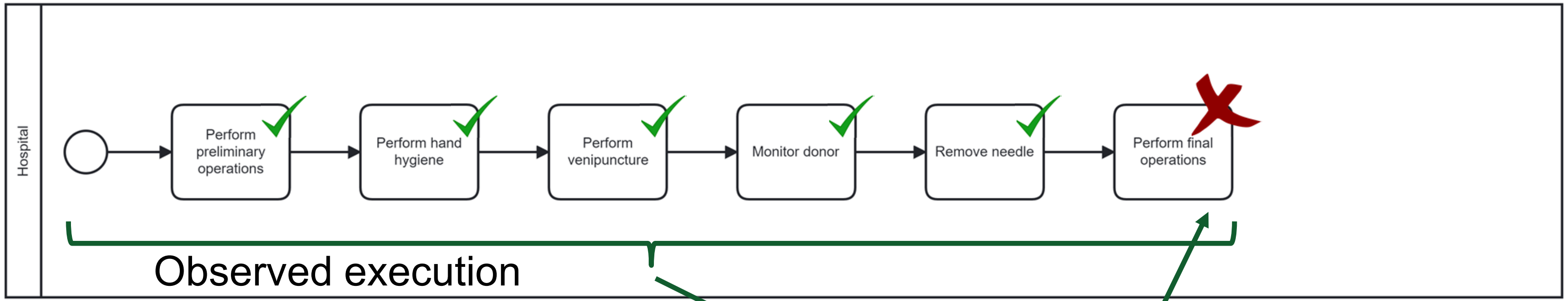
WHO guidelines on drawing blood:

best practices in phlebotomy

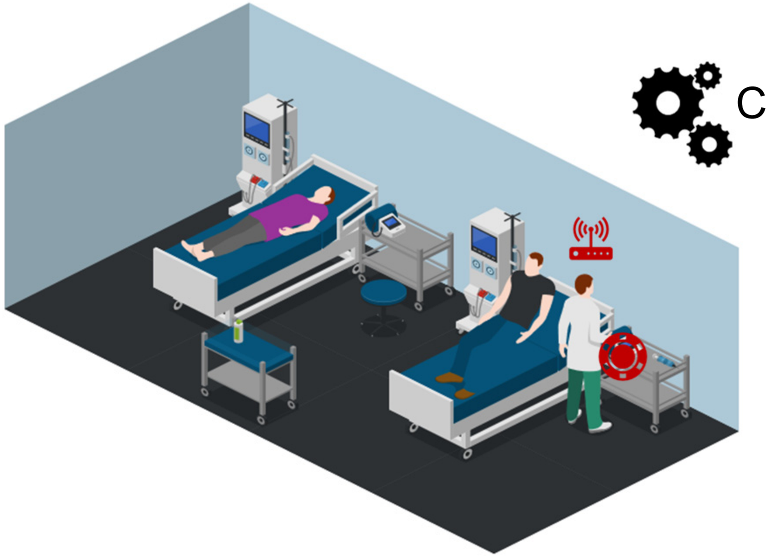
Process description

Universität St.Gallen

Process Conformance Checking



Observed execution



CEP

WHO guidelines on drawing blood:

best practices in phlebotomy

Process description

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery and Exploration

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

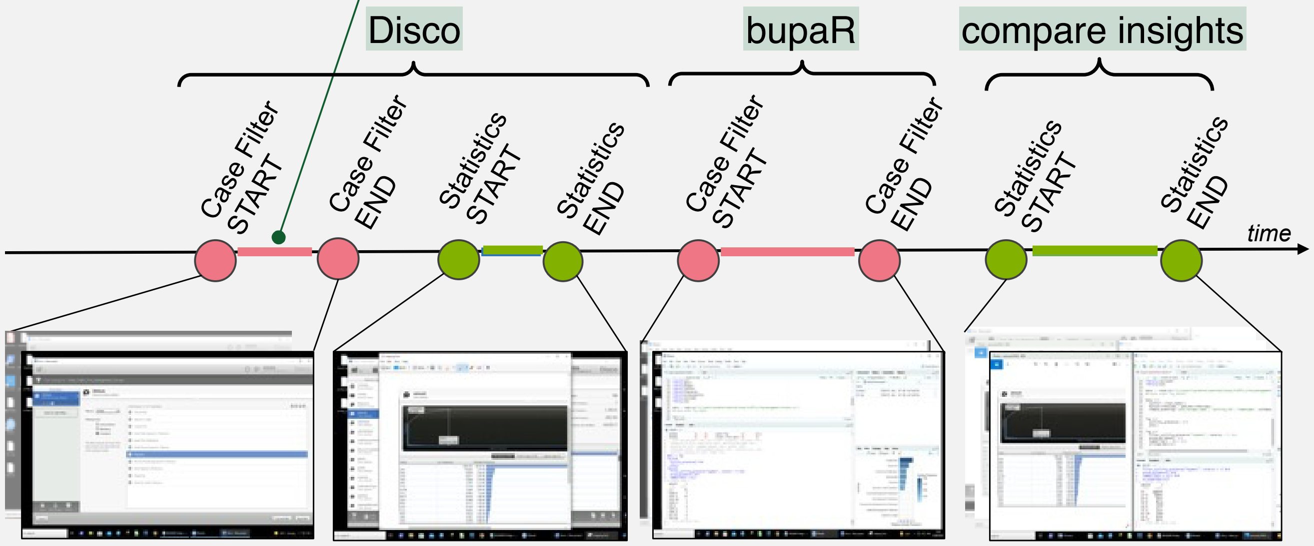
Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Universität St.Gallen

Providing Context to User Interaction Logs

Think-Aloud Data

[00:09:38,3 – 00:11:09,9]
"I'm trying to filter the Payment activity to see all the cases that we don't have a payment. I've tried using the filter forbidden..."

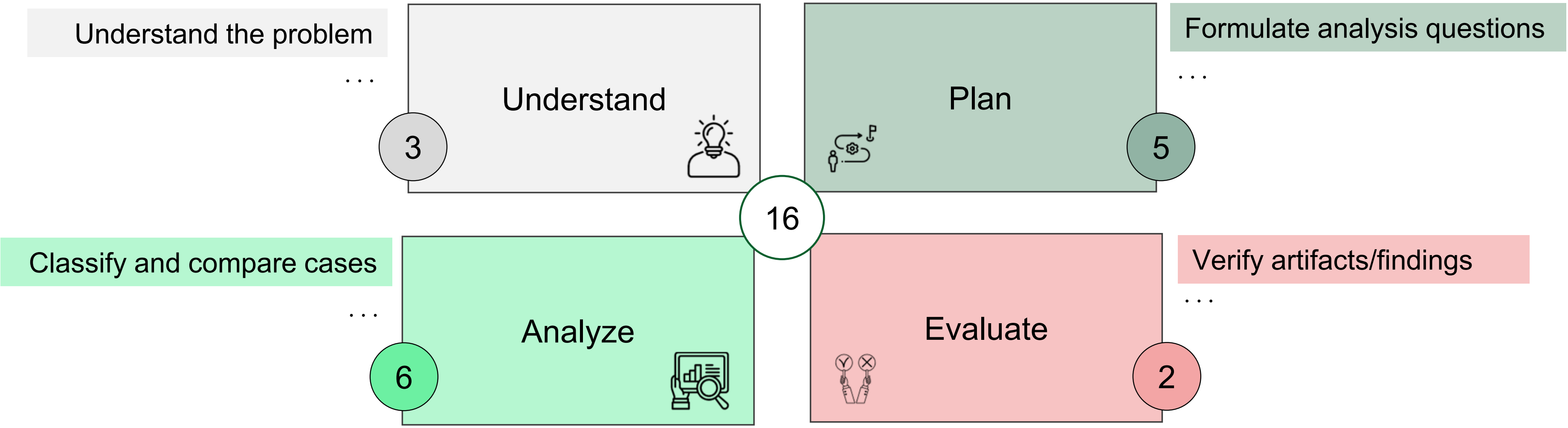


ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:16:37,1	00:16:59,8
P27	Statistics	Disco	00:16:37,1	00:16:59,8
…	…	…	…	…

Providing Context to User Interaction Logs

Usage of Common Strategies

Process mining strategies derived from the analysis of interview data.



Understand the problem

. . .

Understand

3

Plan

5

Formulate analysis questions

. . .

16

Classify and compare cases

. . .

Analyze

6

Evaluate

2

Verify artifacts/findings

. . .

Source: F. Zerbato, P. Soffer, B. Weber, Process Mining Practices: Evidence from Interviews.

Providing Context to User Interaction Logs

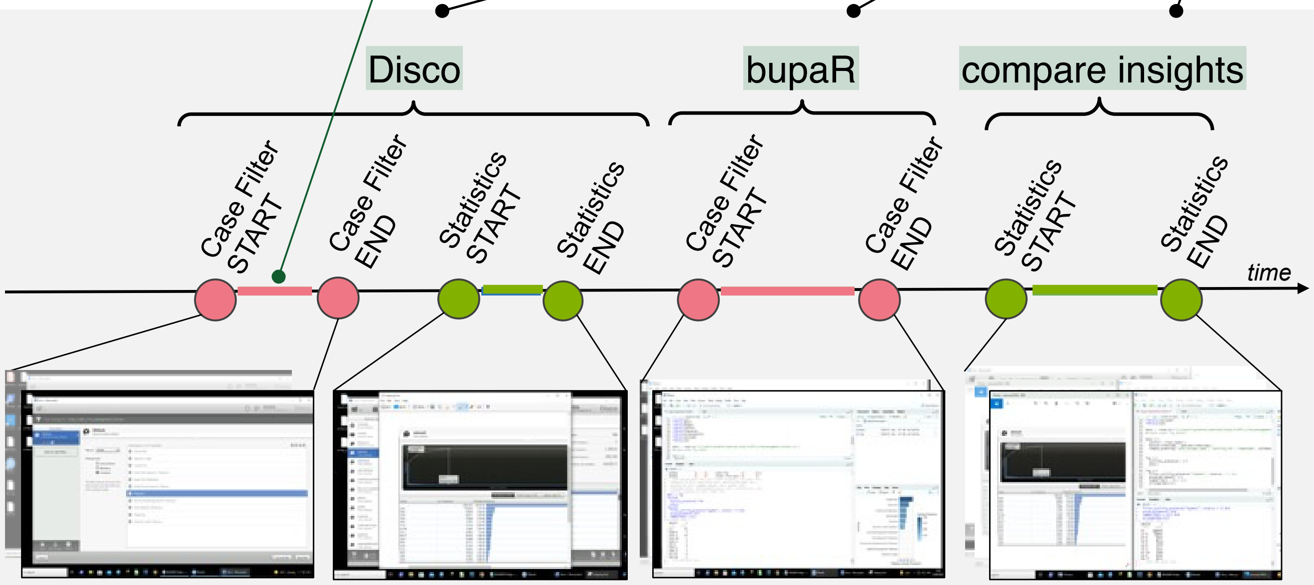
Strategy: Verify artifacts and findings

"Often I try to combine different tools to understand for sanity check if we have the same insights in different tools."

Interview Data

Think-Aloud Data

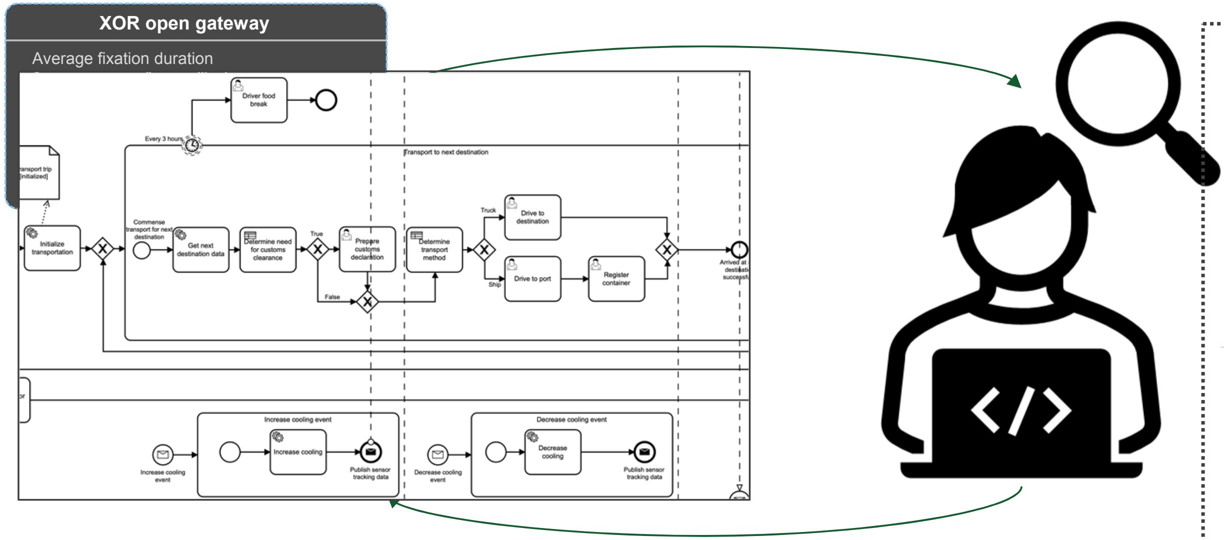
[00:09:38,3 – 00:11:09,9]
"I'm trying to filter the Payment activity to see all the cases that we don't have a payment. I've tried using the filter forbidden..."



ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:16:37,1	00:16:59,8
P27	Statistics	Disco	00:16:37,1	00:16:59,8
…	…	…	…	…

Associating a User’s Cognitive and Affective State With a Software Design Artifact

changes user’s cognitive and affective state



reads, creates, makes sense of, validates

Which parts of the model are perceived as difficult?

Neuro-physiological measures to continuously assess a user’s cognitive and affective state (e.g., cognitive load)

Eye-related measures

Skin-related measures

Heart-related measures

Brain-related measures

Source: Amine Abbad-Andaloussi, Thierry Sorg, Barbara Weber: Estimating Developers' Cognitive Load at a Fine-grained Level Using Eye-tracking Measures

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery and Exploration

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences

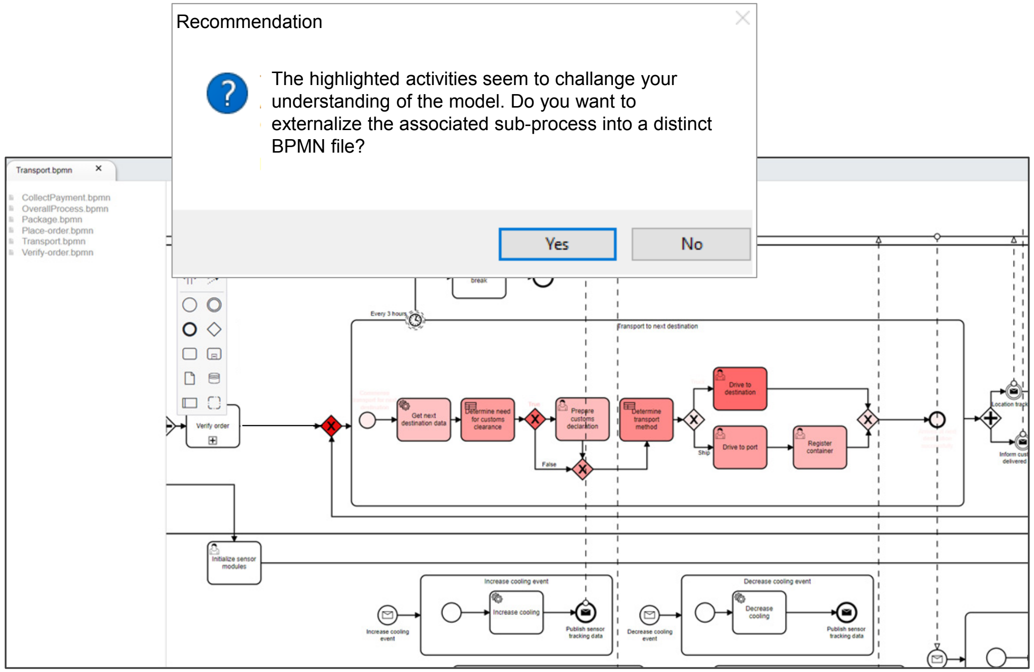
Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns

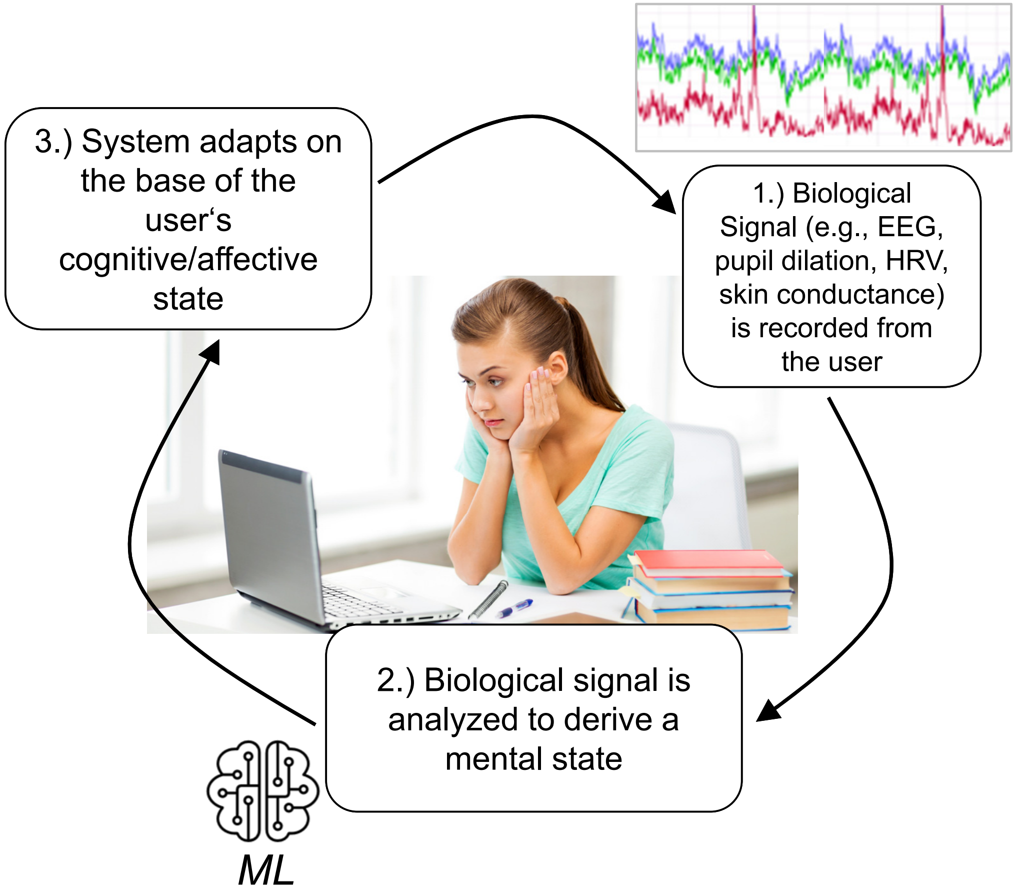
Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Biofeedback and Neuro-adaptive Software Systems



Recommendation

The highlighted activities seem to challange your understanding of the model. Do you want to externalize the associated sub-process into a distinct BPMN file?



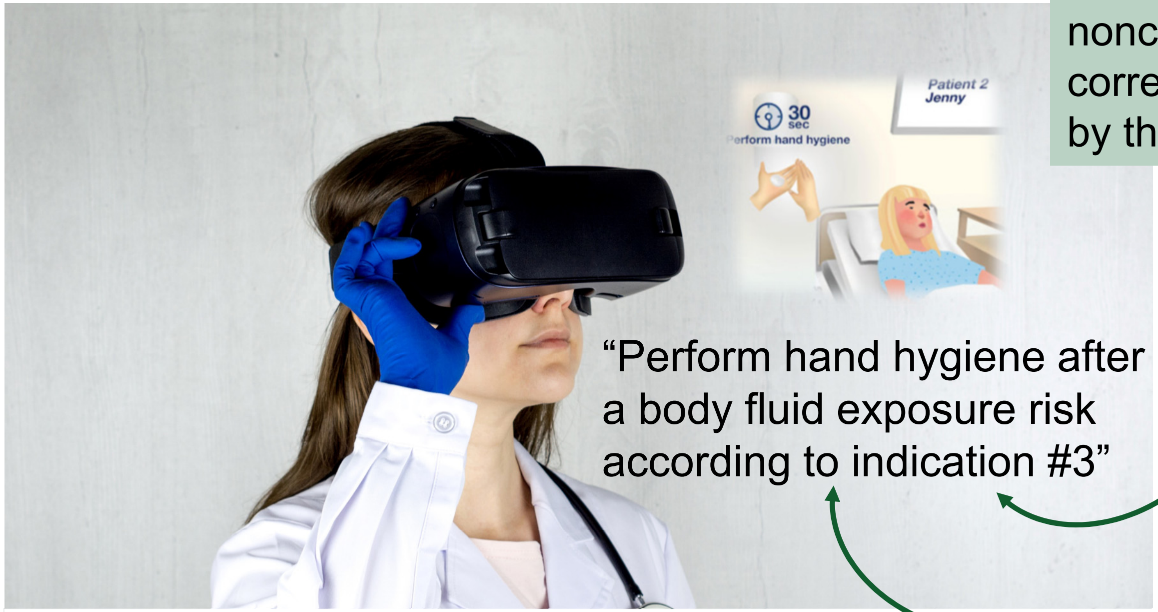
3.) System adapts on the base of the user‘s cognitive/affective state

1.) Biological Signal (e.g., EEG, pupil dilation, HRV, skin conductance) is recorded from the user

2.) Biological signal is analyzed to derive a mental state

Interpretable Feedback

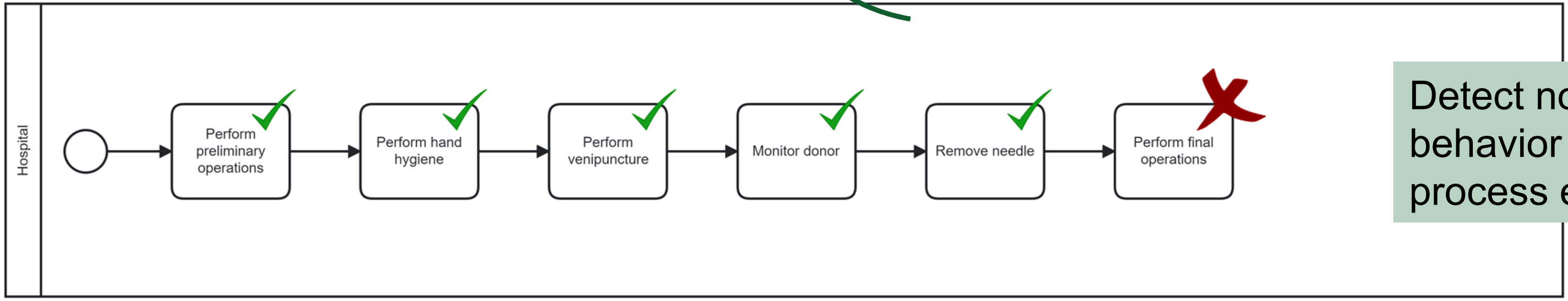
Inform the agent about nonconformance and indicate the correct course of action prescribed by the process description



“Perform hand hygiene after a body fluid exposure risk according to indication #3”

WHO guidelines on drawing blood:

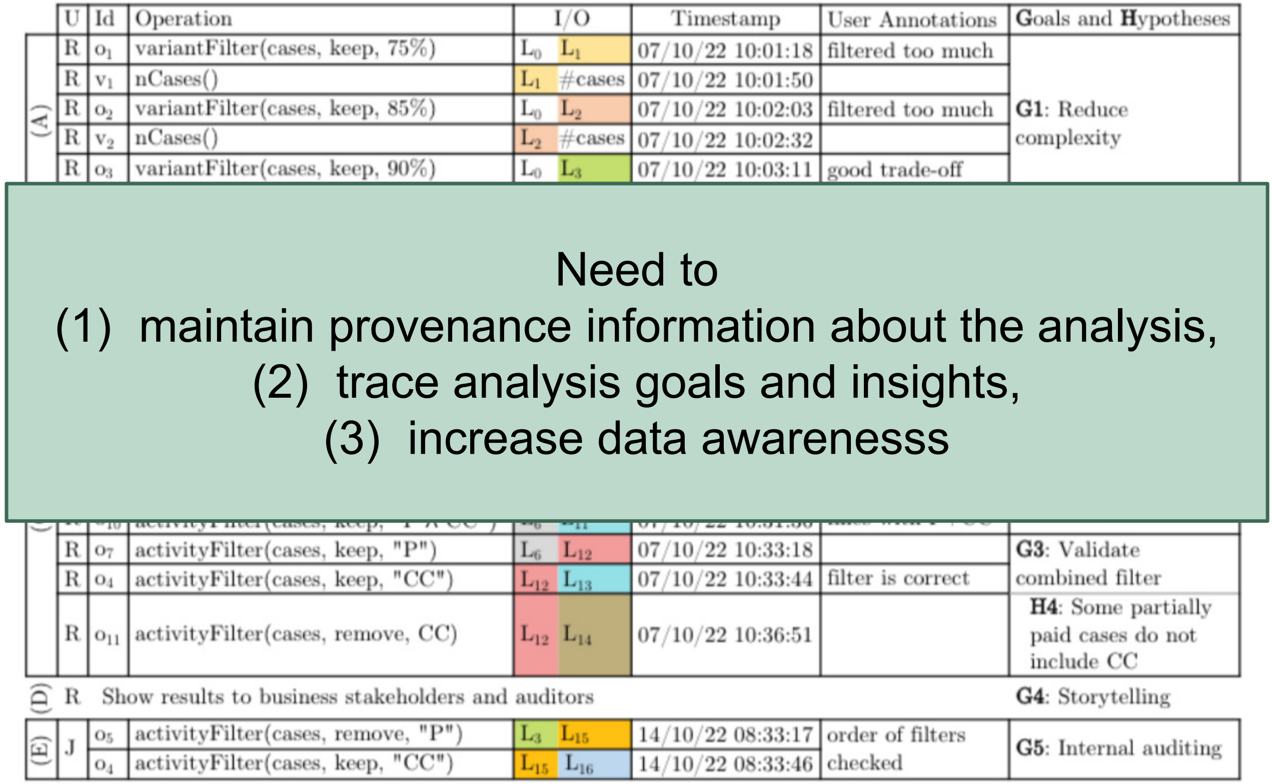
best practices in phlebotomy



Detect nonconforming behavior during process execution

Data-driven Tool Development

Example of a recorded process mining analysis



	U	Id	Operation	I/O		Timestamp	User Annotations	**G**oals and **H**ypotheses
(A)	R	o_1	variantFilter(cases, keep, 75%)	L_0	L_1	07/10/22 10:01:18	filtered too much	**G1**: Reduce complexity
	R	v_1	nCases()	L_1	#cases	07/10/22 10:01:50		
	R	o_2	variantFilter(cases, keep, 85%)	L_0	L_2	07/10/22 10:02:03	filtered too much	
	R	v_2	nCases()	L_2	#cases	07/10/22 10:02:32		
	R	o_3	variantFilter(cases, keep, 90%)	L_0	L_3	07/10/22 10:03:11	good trade-off	
	R	o_7	activityFilter(cases, keep, "P")	L_6	L_{12}	07/10/22 10:33:18		**G3**: Validate combined filter
	R	o_4	activityFilter(cases, keep, "CC")	L_{12}	L_{13}	07/10/22 10:33:44	filter is correct	
	R	o_{11}	activityFilter(cases, remove, CC)	L_{12}	L_{14}	07/10/22 10:36:51		**H4**: Some partially paid cases do not include CC
(D)	R	Show results to business stakeholders and auditors						**G4**: Storytelling
(E)	J	o_5	activityFilter(cases, remove, "P")	L_3	L_{15}	14/10/22 08:33:17	order of filters checked	**G5**: Internal auditing
		o_4	activityFilter(cases, keep, "CC")	L_{15}	L_{16}	14/10/22 08:33:46		

Need to
(1) maintain provenance information about the analysis,
(2) trace analysis goals and insights,
(3) increase data awarenesss

Supporting Provenance and Data Awareness in Exploratory Process Mining

Francesca Zerbato[0000−0001−7797−4602]1, Andrea Burattin[0000−0002−0837−0183]2,
Hagen Völzer[0000−0003−3547−3847]1, Paul Nelson Becker2, Elia Boscaini2, and
Barbara Weber[0000−0002−6004−4860]1

1 University of St. Gallen, St. Gallen, Switzerland
{francesca.zerbato|hagen.voelzer|barbara.weber}@unisg.ch
2 Technical University of Denmark, Kgs. Lyngby, Denmark
{s194702|s194720}@student.dtu.dk, andbur@dtu.dk

Abstract. Like other analytic fields, process mining is complex and knowledge-intensive and, thus, requires the substantial involvement of human analysts. The analysis process unfolds into many steps, producing multiple results and artifacts that analysts need to validate, reproduce and potentially reuse. We propose a system supporting the validation, reproducibility, and reuse of analysis results via analytic provenance and data awareness. This aims at increasing the transparency and rigor of exploratory process mining analysis as a basis for its stepwise maturation. We outline the purpose of the system, describe the problems it addresses, derive requirements and propose a design satisfying these requirements. We then demonstrate the feasibility of the central aspects of the design.

Keywords: Process Mining · Exploratory Analysis · System Requirements and Design · Analytic Provenance · Data Awareness · User Support

1 Introduction

Process mining comprises methods to analyze event data generated in information systems during the execution of business processes. Process mining is quickly growing in adoption, and so is its business impact [9].

Like other data science disciplines, process mining requires the substantial involvement of humans, e.g., process analysts, to obtain insights from raw event data [7]. Analysts often freely explore the data with the available tools to gain a basic understanding of what it represents, investigate different scenarios, and create hypotheses. Hypotheses can then be tested using best practices, but more exploration is required if the test fails or the results are inconclusive [19]. Each insight that emerges during the analysis informs which subsequent analysis steps are chosen. On the one hand, the choices made during the analysis yield many possible reasonable results that need to be assessed. On the other hand, such choices might give rise to potential inconsistencies in the analysis process [14].

Due to its knowledge-intensive character and emergent course of action, an exploratory analysis includes many manual and error-prone steps that are often

Source: Francesca Zerbato, Andrea Burattin, Hagen Völzer, Paul Nelson Becker, Elia Boscaini, Barbara Weber: *Supporting Provenance and Data Awareness in Exploratory Process Mining.*

Summary

- Consider leveraging digital trace data beyond traditional business processes
- Carefully planning data collection pays off!
- Going beyond traditional business processes offers great opportunities but brings challenges in terms of process observability, event correlation, and event abstraction

Acknowlegements

ProMiSE: Process Mining Support for End-Users

FNSNF
FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE
Grant No.: 200021 197032

Universität St.Gallen

University of Haifa

DTU

ProAmbitIon: Online Process Conformance Checking with Ambiguities Driven by the Internet of Things

FNSNF
FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE
SPIRIT project 208497

Universität St.Gallen

Tecnológico de Monterrey

Kantonsspital St.Gallen

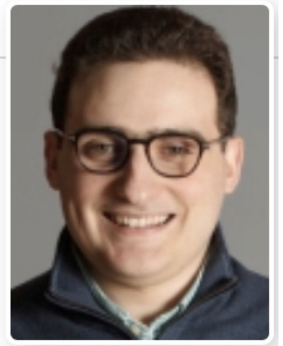
DTU

Supporting Software Maintenance With Psycho-physiological Measures and Artifact Metrics

Universität St.Gallen

IPF Fellowship
Grant No.: 1031574

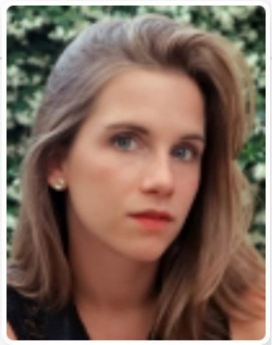
Thank You to My Team



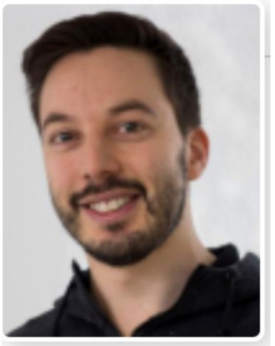
Dr. Amine Abbad Andaloussi
IPF Postdoctoral Fellow

Prof. Dr. Ronny Seiger
Assistant Professor

Thierry Sorg
PhD student



Dr. Francesca Zerbato
Senior Researcher



Dr. Marco Franceschetti
Senior Researcher

Dr. Hagen Völzer
Scientific Project Manager

Lisa Zimmermann
PhD student

References

- Amine Abbad Andaloussi, Thierry Sorg, Barbara Weber: *Estimating developers' cognitive load at a fine-grained level using eye-tracking measures.* ICPC 2022: 111-121
- Amine Abbad Andaloussi, Francesca Zerbato, Andrea Burattin, Tijs Slaats, Thomas T. Hildebrandt, Barbara Weber: *Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud.* Softw. Syst. Model. 20(5): 1437-1464 (2021)
- Thomas H. Davenport and Andrew Spanyi. *What Process Mining Is, and Why Companies Should Do It*. https://hbr.org/2019/04/what-process-mining-is-and-why-companies-should-do-it. Harvard Business Review, 201
- Mangler, Juergen, Joscha Grüger, Lukas Malburg, Matthias Ehrendorfer, Yannis Bertrand, Janik-Vasily Benzin, Stefanie Rinderle-Ma, Estefania Serral Asensio, and Ralph Bergmann. 2023. *DataStream XES Extension: Embedding IoT Sensor Data into Extensible Event Stream Logs.* Future Internet 15, no. 3: 109. https://doi.org/10.3390/fi15030109
- Seiger, Ronny, Marco Franceschetti, and Barbara Weber. 2023. *An Interactive Method for Detection of Process Activity Executions from IoT Data.* Future Internet 15, no. 2: 77. https://doi.org/10.3390/fi15020077
- R. Seiger, L. Malburg, B. Weber, R. Bergmann, *Integrating process management and event processing in smart factories: A systems architecture and use cases.* Journal of Manufacturing Systems, Volume 63, April 2022, Pages 575-592.
- Wil M. P. van der Aalst, Josep Carmona: *Process Mining Handbook*. Lecture Notes in Business Information Processing 448, Springer 2022, ISBN 978-3-031-08847-6.
- van der Aalst, W. et al. (2012). *Process Mining Manifesto.* In: Daniel, F., Barkaoui, K., Dustdar, S. (eds) Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, vol 99. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28108-2_19
- vom Brocke, Jan and van der Aalst, Wil MP and Grisold, Thomas and Kremser, Waldemar and Mendling, Jan and Pentland, Brian and Recker, Jan and Recker, Jan and Roeglinger, Maximilian and Rosemann, Michael and Weber, Barbara, *Process Science: The Interdisciplinary Study of Continuous Change* (September 3, 2021). Available at SSRN: https://ssrn.com/abstract=3916817 or http://dx.doi.org/10.2139/ssrn.3916817
- Francesca Zerbato, Andrea Burattin, Hagen Völzer, Paul Nelson Becker, Elia Boscaini, Barbara Weber: *Supporting Provenance and Data Awareness in Exploratory Process Mining.* Accepted for CAiSE.
- Francesca Zerbato, Pnina Soffer, Barbara Weber: *Process Mining Practices: Evidence from Interviews.* BPM 2022: 268-285

Summary

- **Digital traces** come in many flavors as do the underlying processes
- It pays off to **carefully plan data collection** including the selection of data sources and to collect data to ensure that the different process elements can be linked with the collected data
- The extent to which **process knoweldge** is available largely influences event log generation as well as subsequent analysis
- Digital traces can be leveraged to **discover** so far unknown unknowns, to **test** known knowns and **monitor** known unknowns
- Huge potential of **multi-modal data** and **contextualization of events** to support the identification of root causes.
- Digital traces can be used for **interpretable feedback,** the development of **adaptive systems** and are an important source for **data-driven tool development**

Process Science Activities

Phase	Goal	Exemplary Methods
Discovery	Capturing and describing processes	Techniques such as processes mining, to create **representations of processes** using **digital event data**; **event-based architectures** to organize **data collection** as well as computational methods to analyze the data and to **identify patterns in processes**
Explanation	Understanding, why, how and when a process unfolds	Methods supporting sense-making around processes in a specific context, e.g., qualitative empirical research to study the **context in which a pattern is situated**. Leads to propositions or entire theories on **cause effect relationships embedded in a situational context**
Intervention	Intervening and shaping the process into desired directions	Methods to **develop and evaluate interventions** to processes. Applying, e.g., design-oriented research (aka engineering research), developing interventions based on explanatory research and evaluating effects of such interventions in process event data.

vom Brocke et al., Process Science: The Interdisciplinary Study of Continuous Change

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences, Create Augmented Representations

Conformance Checking

Process Monitoring

Linking Data Sources and Contextualizing Events and Patterns
(Supporting the identification of root-causes)

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Universität St.Gallen

Process Science in Action

Selection of Data Sources, Data Collection, and Event Log Generation

Process Discovery
(Exploring the unknown unknown)

Create „Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences, Create Augmented Representations

Conformance Checking
(Testing the known known)

Process Monitoring
(Monitoring the known unknown)

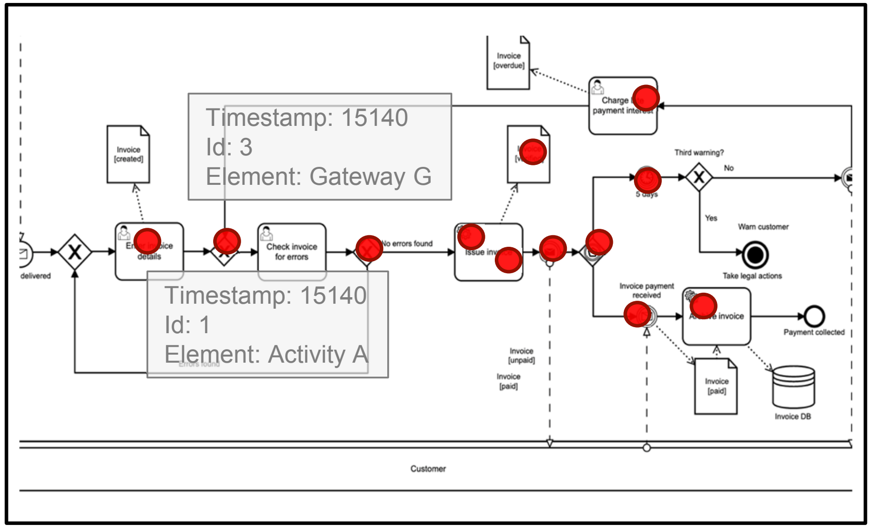
Linking Data Sources and Contextualizing Events and Patterns
(Supporting the identification of root-causes)

Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems
Data-driven Tool Development

Automated Mapping of Attentional Processes to Software Design Artifacts

Eye-tracker

Data collection

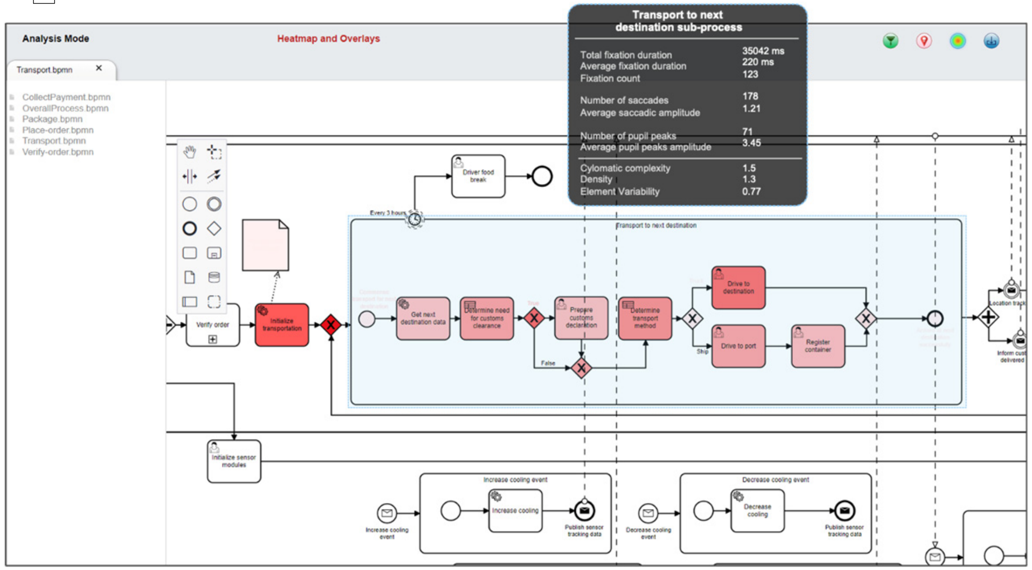


Timestamp: 15140
Id: 3
Element: Gateway G

Timestamp: 15140
Id: 1
Element: Activity A

Automated mapping of gazes to process model elements; each element is considered as Area of Interest (AOI)

Analysis



Transport to next destination sub-process

Total fixation duration 35042 ms
Average fixation duration 220 ms
Fixation count 123

Number of saccades 178
Average saccadic amplitude 1.21

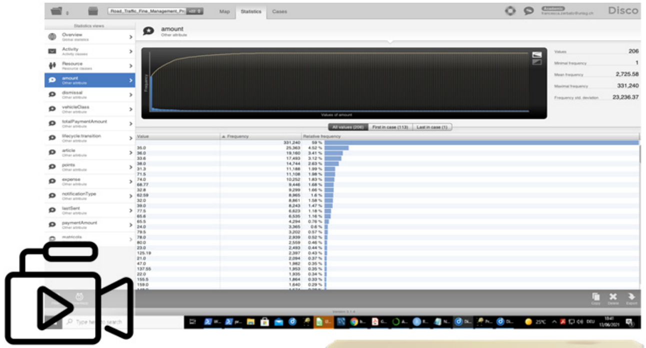
Number of pupil peaks 71
Average pupil peaks amplitude 3.45

Cylomatic complexity 1.5
Density 1.3
Element Variability 0.77

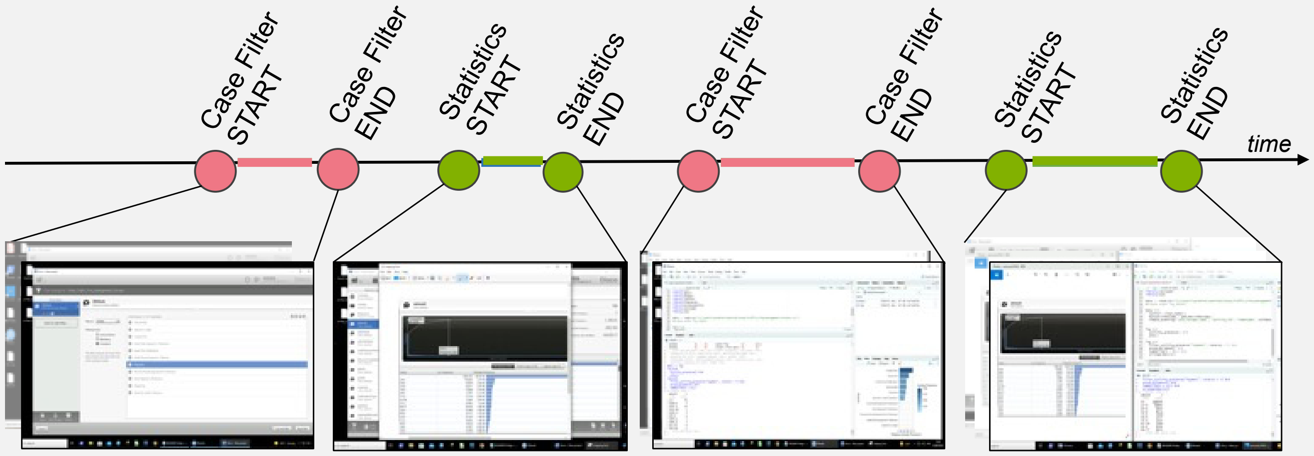
Instantaneous calculation of AOI-based measures and generation of heatmaps *(without the need to manually define AOIs)*

Source: Prototype developed by Amine Abbad Andaloussi

Creating User Interaction Logs From Screen Recordings

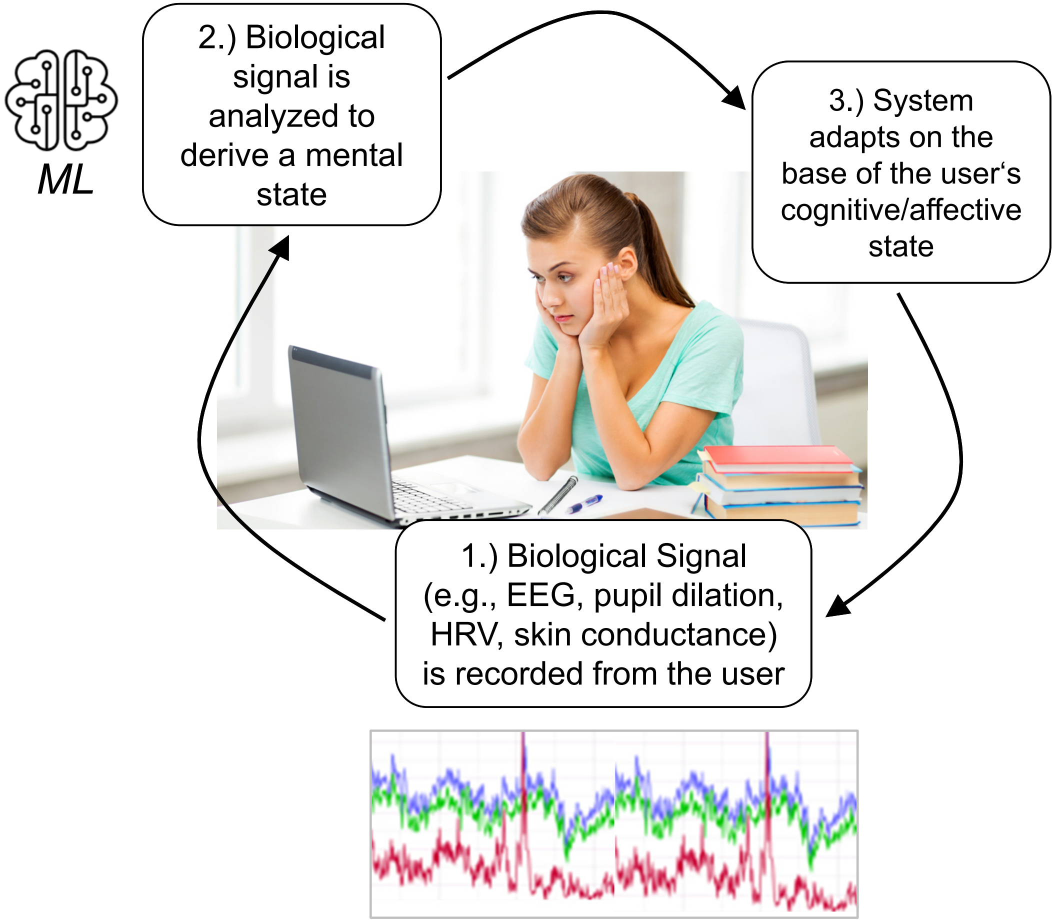


Screen Recording



ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:11:46,1	00:12:34,3
P27	Statistics	Disco	00:11:46,1	00:12:34,3
…	…	…	…	…

Biofeedback and Neuro-adaptive Software Systems



2.) Biological signal is analyzed to derive a mental state

3.) System adapts on the base of the user‘s cognitive/affective state

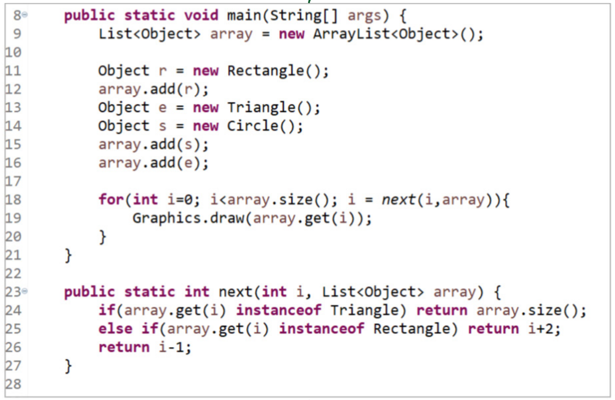
1.) Biological Signal (e.g., EEG, pupil dilation, HRV, skin conductance) is recorded from the user

Highlights of the mentally demanding parts of code to faciliate code review

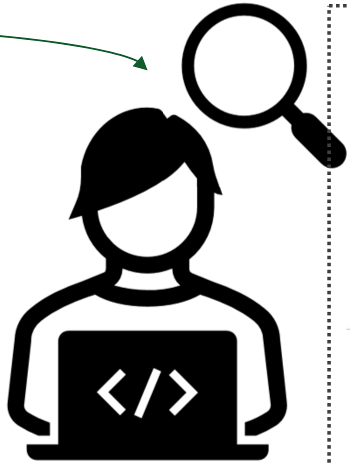
```
    public static ApplicationController instance() {
        return controller;
    }

    /**
     * Private class constructor. Access the application controller through the
     * {@link #instance()} method.
     */
    private ApplicationController() {
        // creates the master configuration
        configuration = UIConfiguration.master();

        // creates the panels
        mainPage = new MainPage(configuration.getChild(MainPage.
                class.getCanonicalName()));
        loadProcessPage = new LoadProcessPage(configuration.
                getChild(LoadProcessPage.class
                .getCanonicalName()));
        waitingPage = new WaitingPage(configuration.
                getChild(WaitingPage.class.getCanonicalName()));

        // creates the main frame
        mainFrame = new MainFrame(this);
        mainFrame.addPage(mainPage);
        mainFrame.addPage(loadProcessPage);
        mainFrame.addPage(waitingPage);

        // creates children controllers
        logsController = new LogController(this);
        modelController = new ProcessController(this);
        consoleController = new ConsoleController(this);

        // initialization logging
        Logger.instance().debug("Application started!");
        Logger.instance().debug("You have " + CPUUtils.
                CPUAvailable() + " CPU(s) available");
    }
```

Associating a User's Cognitive and Affective State With a Software Design Artifact

changes user's cognitive and affective state



```java
    public static void main(String[] args) {
        List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object> array) {
        if(array.get(i) instanceof Triangle) return array.size();
        else if(array.get(i) instanceof Rectangle) return i+2;
        return i-1;
    }
```



reads, creates, makes sense of, validates

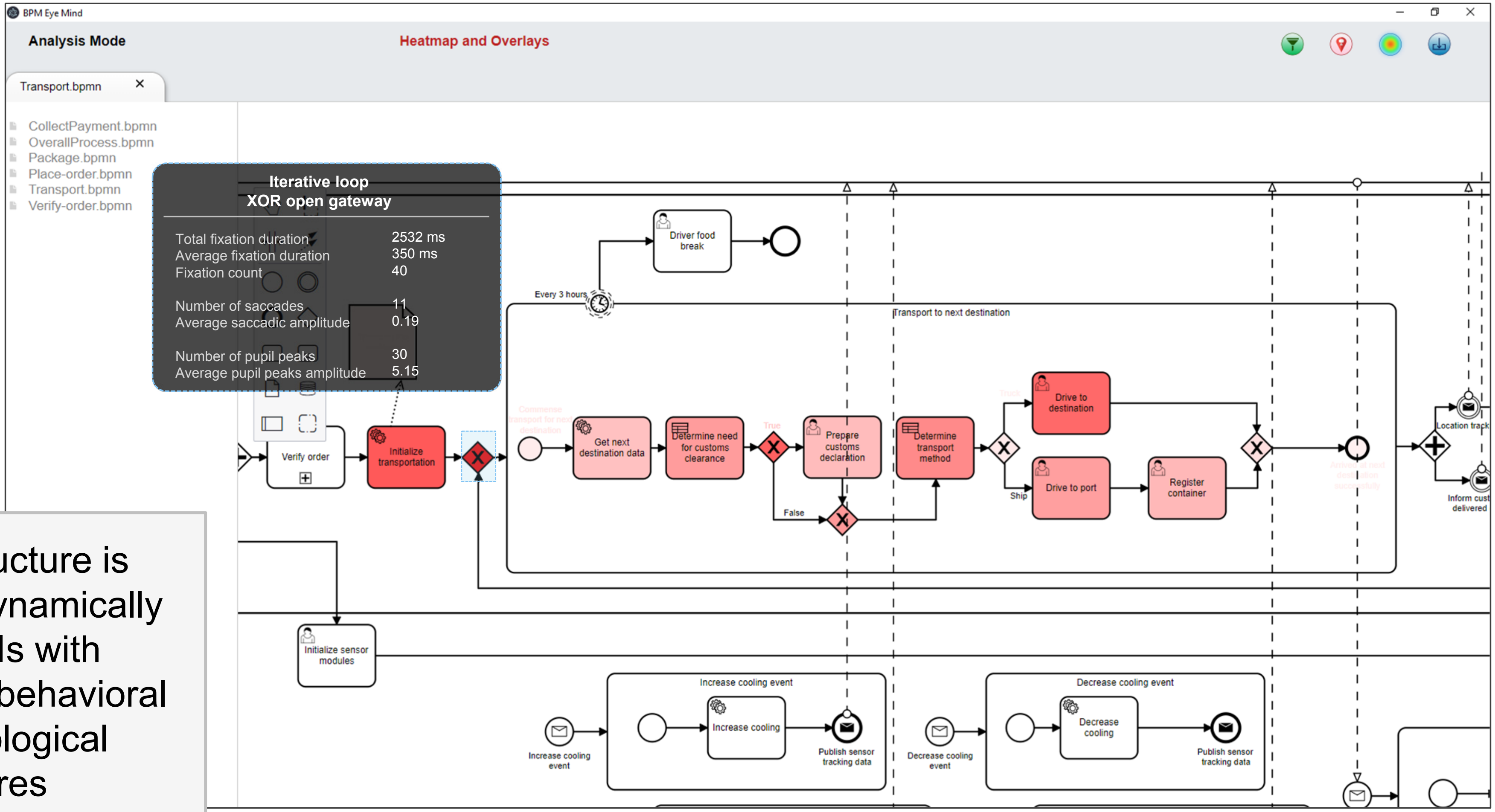
Which parts of the code are perceived as difficult by the developer?

Neuro-physiological measures to continuously assess a user's cognitive and affective state (e.g., cognitive load)

- Eye-related measures
- Skin-related measures
- Heart-related measures
- Brain-related measures

Source: Amine Abbad-Andaloussi, Thierry Sorg, Barbara Weber: Estimating Developers' Cognitive Load at a Fine-grained Level Using Eye-tracking Measures

Eye Mind: Process Model Augmented with Eye-tracking Metrics



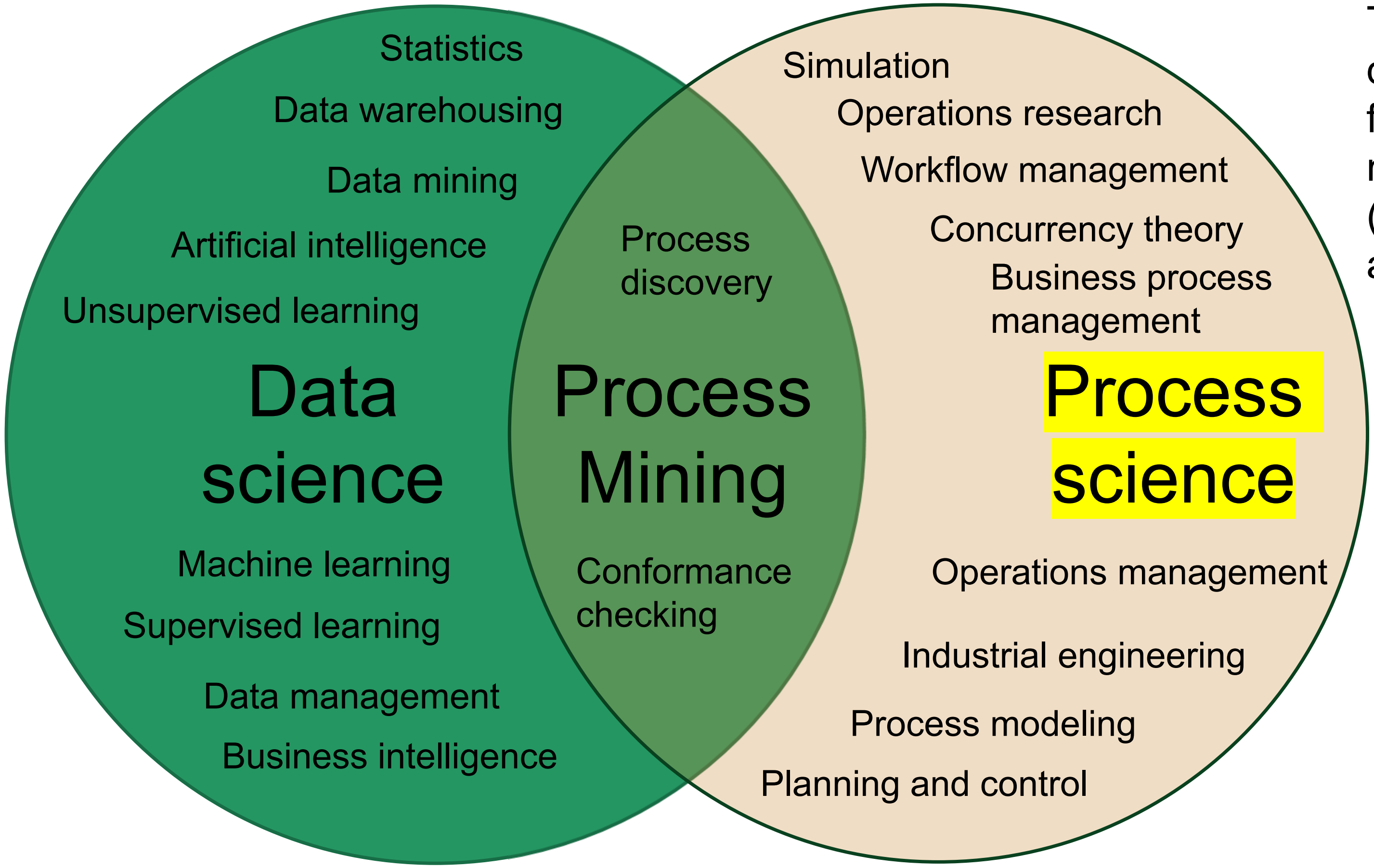
Iterative loop
XOR open gateway

Total fixation duration	2532 ms
Average fixation duration	350 ms
Fixation count	40
Number of saccades	11
Average saccadic amplitude	0.19
Number of pupil peaks	30
Average pupil peaks amplitude	5.15

Process structure is exploited to dynamically create AOIs with projections of behavioral and physiological measures

Source: Prototype developed by Amine Abbad Andaloussi

Process Mining: At the Intersection of Data and Process Science



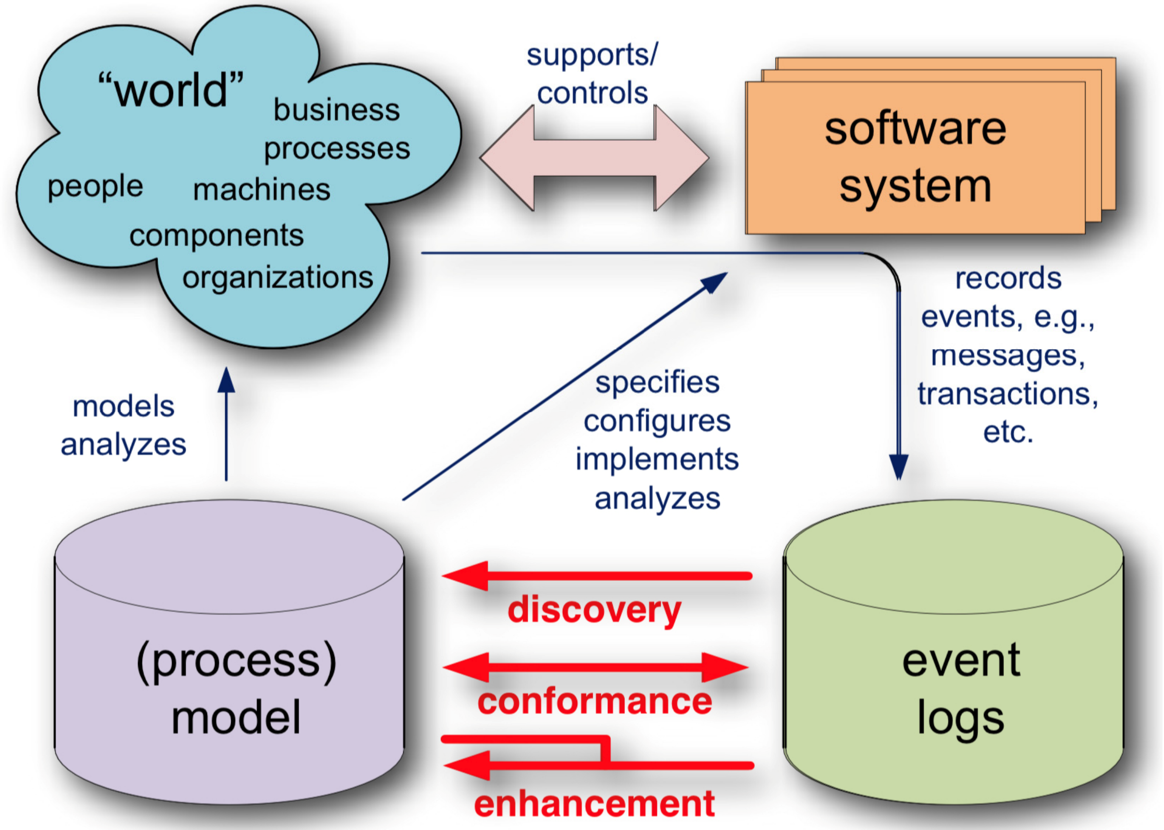
Traditionally, not data-driven and focusses on modeling (languages) and automation.

Traditionally, not process-centric and focusses on specific tasks or decisions

Source: van der Aalst & Carmona: Process Mining Handbook

Universität St.Gallen

Process Mining: The Big Picture



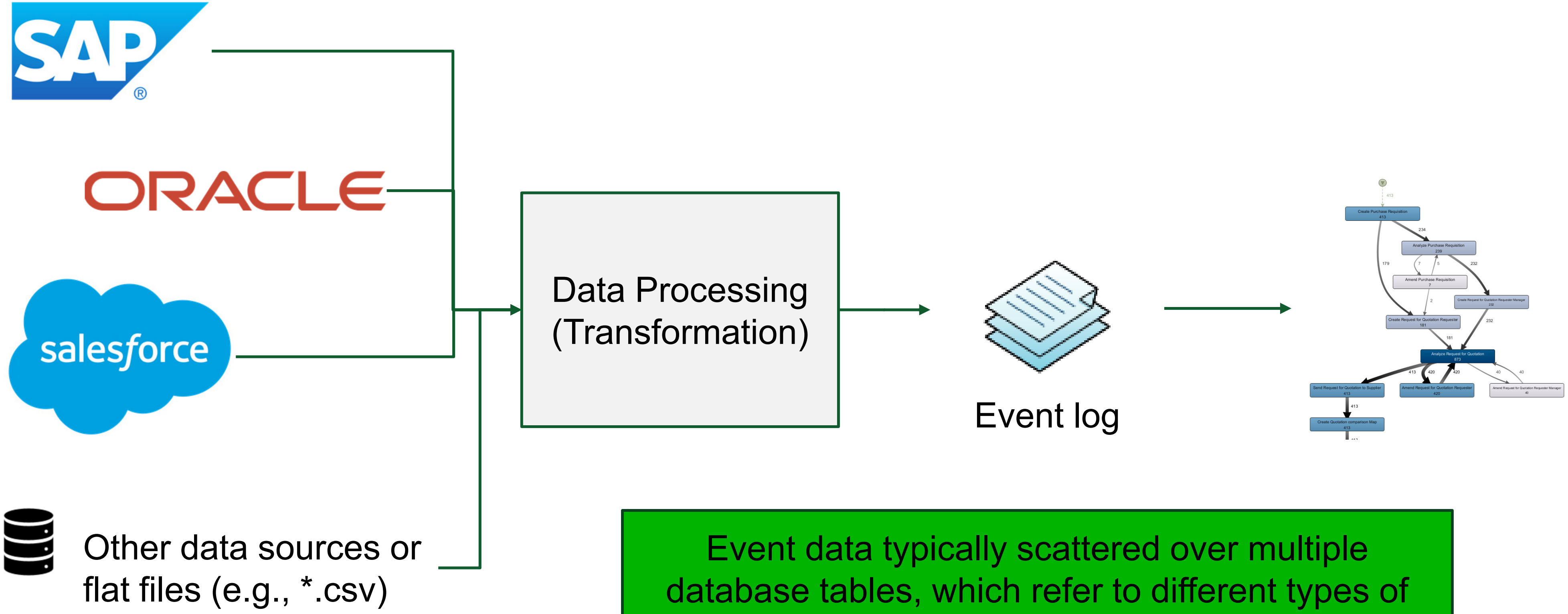
3 main types of process mining:

- discovery,
- conformance and
- enhancement

Source: van der Aalst et al.: Process Mining Manifesto

Universität St.Gallen

ERP and CRM Systems: Common Data Sources for Process Mining



Data Processing (Transformation)

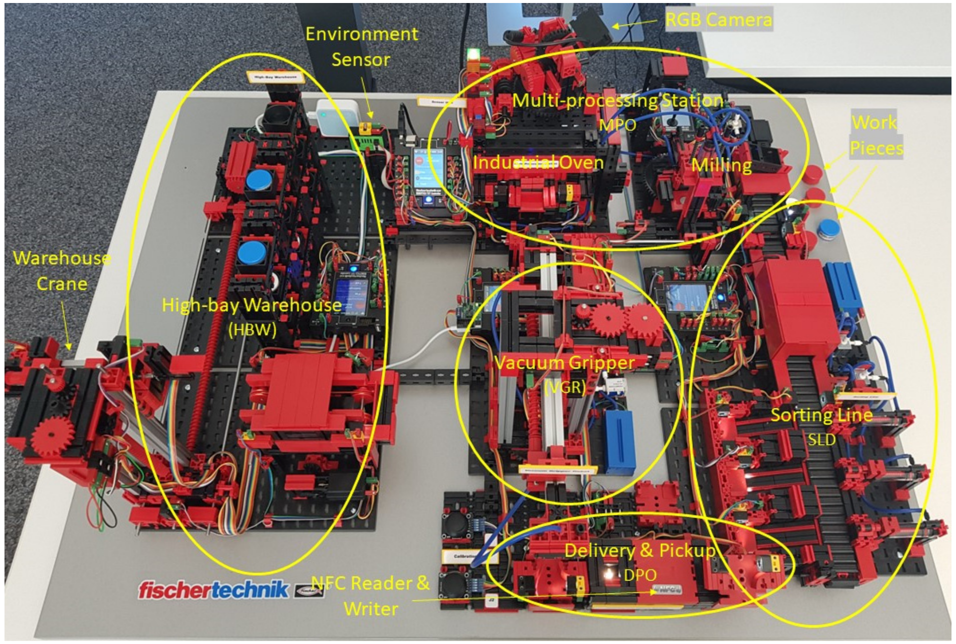
Event log

Other data sources or flat files (e.g., *.csv)

Event data typically scattered over multiple database tables, which refer to different types of objects.

Internet of Things: An Emerging Event Source

IoT is an increasingly important event source in areas like security, manufacturing, healthcare and transport



Smart Factory @ ICS

Sensors:

- Switches
- Light barriers
- Color sensors
- Environment
- Camera
- NFC

Actuators:

- Motors
- Compressors
- Valves

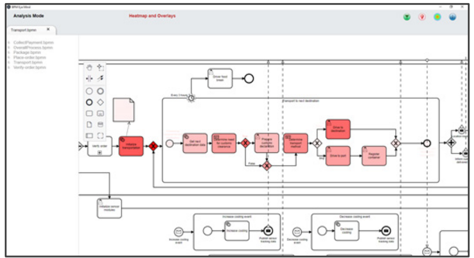
Example event from factory:

Topic: FTFactory/HBW_1

{ "id": "FTFactory/HBW_1", "timestamp": "2020-12-11 13:35:35.50", "i1_light_barrier_interrupted": false, "i2_light_barrier_interrupted": true, "i3_light_barrier_interrupted": true, "i4_light_barrier_interrupted": false, "i5_position_switch_pressed": true, "i6_position_switch_pressed": true, "i7_position_switch_pressed": false, "i8_position_switch_pressed": true, "m1_speed": 0, "m2_speed": 0, "m3_speed": 0, "m4_speed": 0, "current_state": "ready", "current_task": "", "current_task_elapsed_seconds_since_start": 0, "current_sub_task": "", "failure_label": "", "current_pos_x": 0, "current_pos_y": 0, "target_pos_x": 0, "target_pos_y": 0, "amount_of_stored_workpieces": 0}

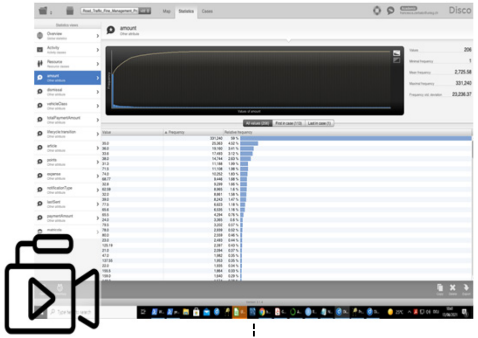
Event Producers

Process Observability Largely Differs

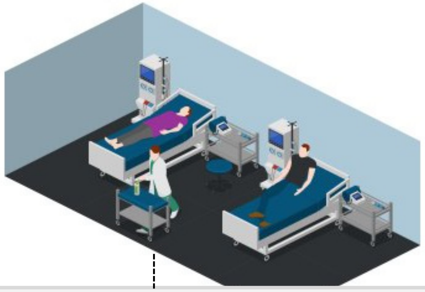


Navigation, scrolling and zooming during model comprehension (depending on tool); large parts of the process occur in the **reader's mind**

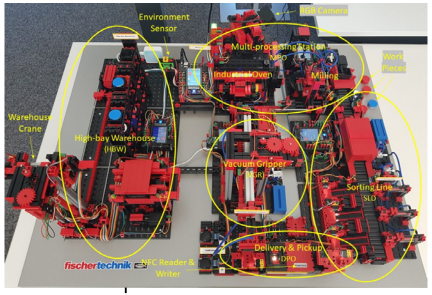
Tool interactions during analysis (depending on the tool); large parts of the process occur in the **analyst's mind**

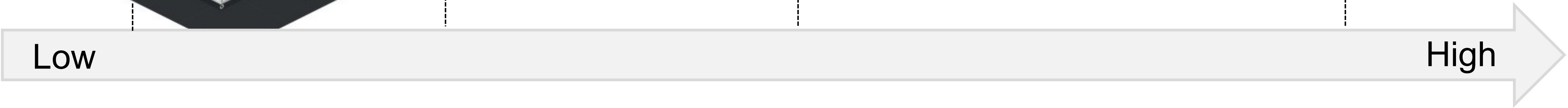


Process is largely manual; most parts performed **outside of any IT system**

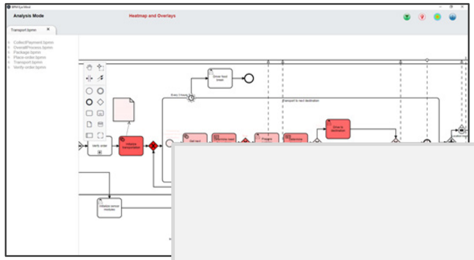


Equipped with **sensors** and **actuators**





Process Observability Largely Differs

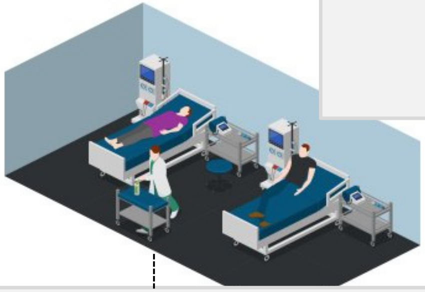


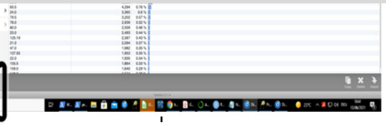
Navigation, scrolling and zooming during model comprehension (depending on tool): large parts of the process occur in

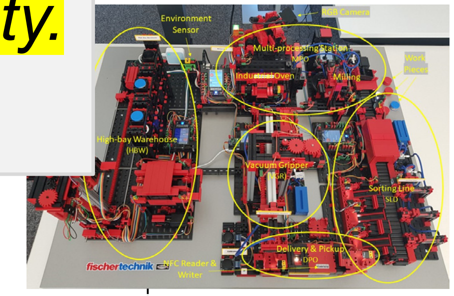
Usage of sensors and additional forms of data collection to increase process observability. Incompleteness and low fidelity of data can be sources of ambiguity.

Process is largely m
most parts perfor
outside of any IT s

uipped with **sensors**
d **actuators**







Low

High

Observability

A Challenge for Human-centered Work Procesess

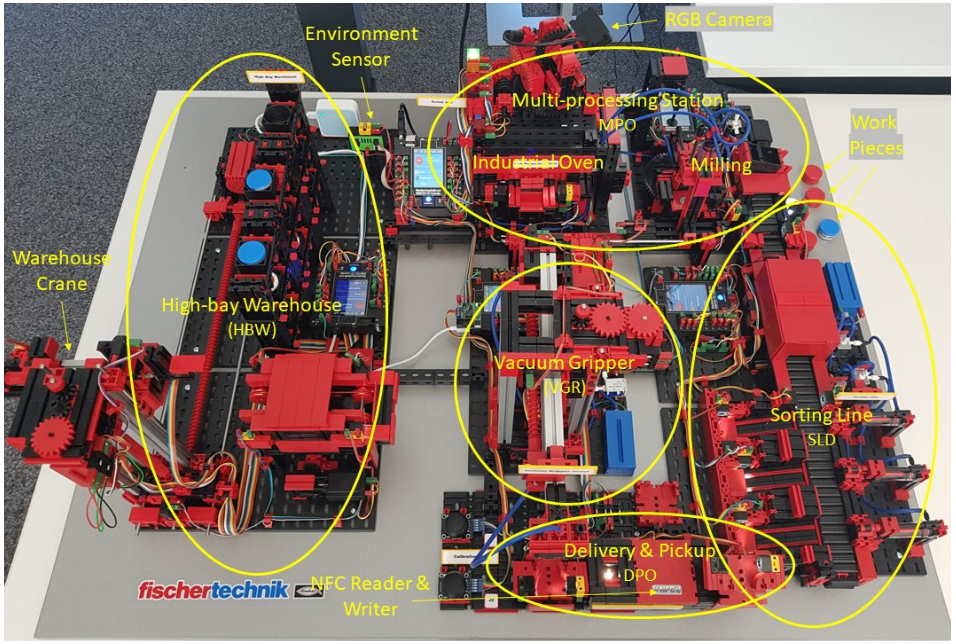
- Characteristics of human-centered work processes
    - Often include manual steps which do not leave traces in any IT system
    - Highly flexible resulting into numerous variants
    - Steps that only happen in the minds of users

- Observability of process steps as pre-condition for event log generation
    - Tracking of interactions (with digital or physical objects)
    - Recording of verbal utterances
    - Video recordings
    - Measurement of of brain and autonomous nervous system activity

    - Not always obvious what to log

Internet of Things: An Emerging Event Source

IoT is an increasingly important event source in areas like security, manufacturing, healthcare and transport



Smart Factory @ ICS

Sensors:

- Switches
- Light barriers
- Color sensors
- Environment
- Camera
- NFC

Actuators:

- Motors
- Compressors
- Valves

Example event from factory:
Topic: FTFactory/HBW_1

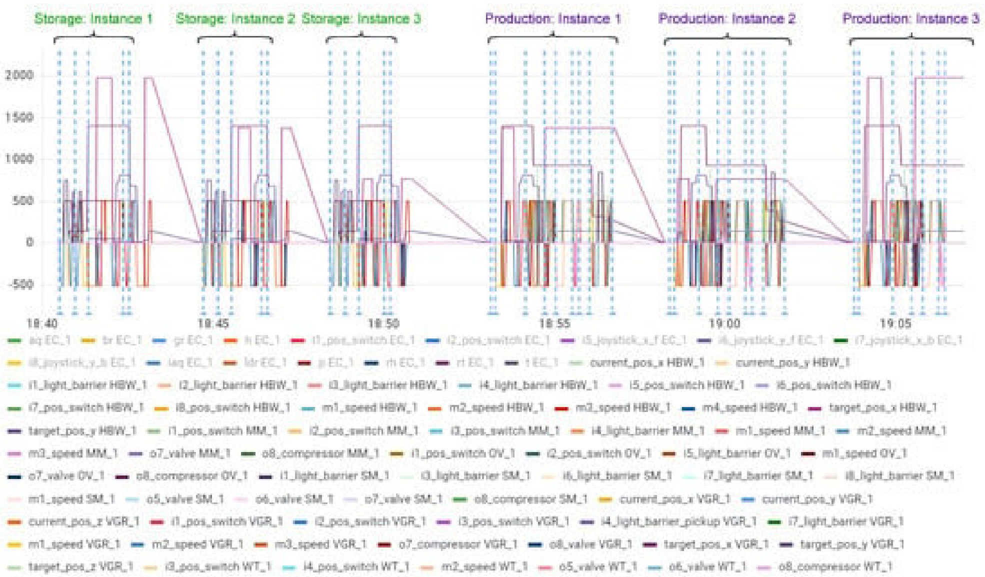
{ "id": "FTFactory/HBW_1", "timestamp": "2020-12-11 13:35:35.50", "i1_light_barrier_interrupted": false, "i2_light_barrier_interrupted": true, "i3_light_barrier_interrupted": true
ed": 0,
"m2_speed": 0, "m3_speed": 0, "m4_speed": 0,
"current_state": "ready", "current_task": "",
"current_task_elapsed_seconds_since_start": 0,
"current_sub_task": "", "failure_label": "",
"current_pos_x": 0, "current_pos_y": 0,
"target_pos_x": 0, "target_pos_y": 0,
"amount_of_stored_workpieces": 0}

Requires bridging the abstraction gap

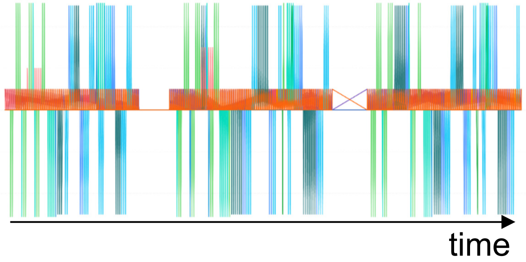
Event Producers

Universität St.Gallen

Process Activity Detection from Sensors



Event sequences of all sensors and actuators of smart factory



Bottom up

Process Awareness of IoT Data

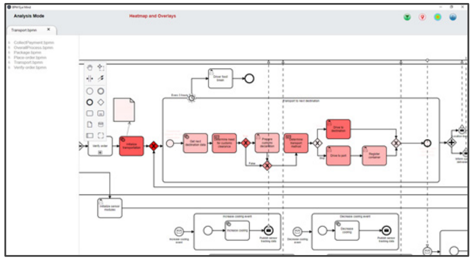
- Starting point is a set of IoT data from sensors and actuators of CPS components
- Contextualization of sensor events in the context of process executions, i.e., association of sensor events with concrete activity executions within a specific process instance

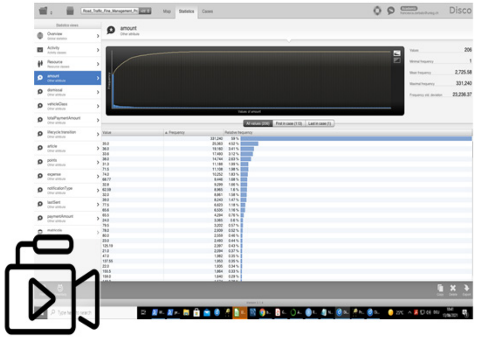
Source: An Interactive Method for Detection of Process Activity Executions from IoT Data

Universität St.Gallen

Availability of Process Knowledge

Process and **activities** largely **unknown;** high flexibility and variability



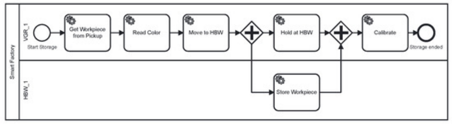


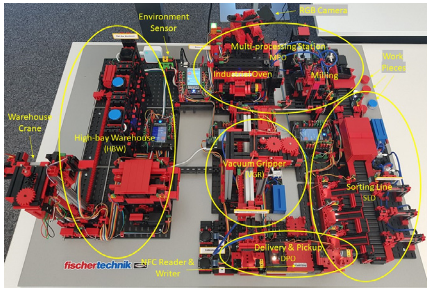
Guidelines including **process steps;** indication for hand hygiene (**business rules**); some flexibility and variability

WHO guidelines on drawing blood:

best practices in phlebotomy

Process and **activities known**, but not explicit; end-to-end visibility lacking; repetitive and well structured



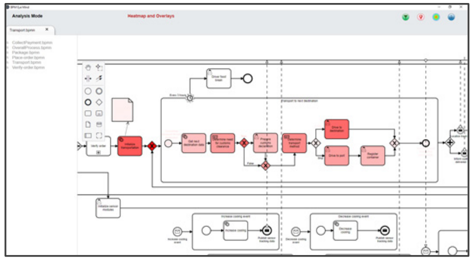


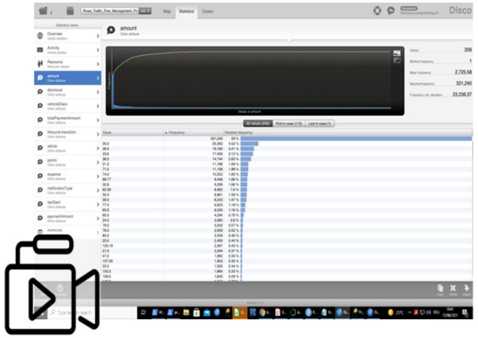
Low → High

Universität St.Gallen

Availability of Process Knowledge

Process and **activities** largely **unknown;** high flexibility and variability



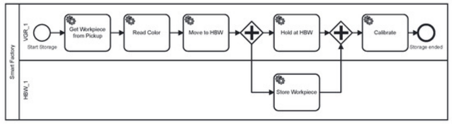


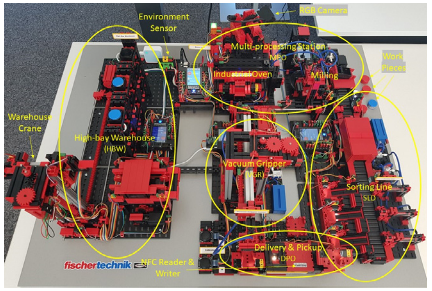
Guidelines including **process steps;** indication for hand hygiene (**business rules**); some flexibility and variability

WHO guidelines on drawing blood:

best practices in phlebotomy

Process and **activities known**, but not explicit; end-to-end visibility lacking; repetitive and well structured

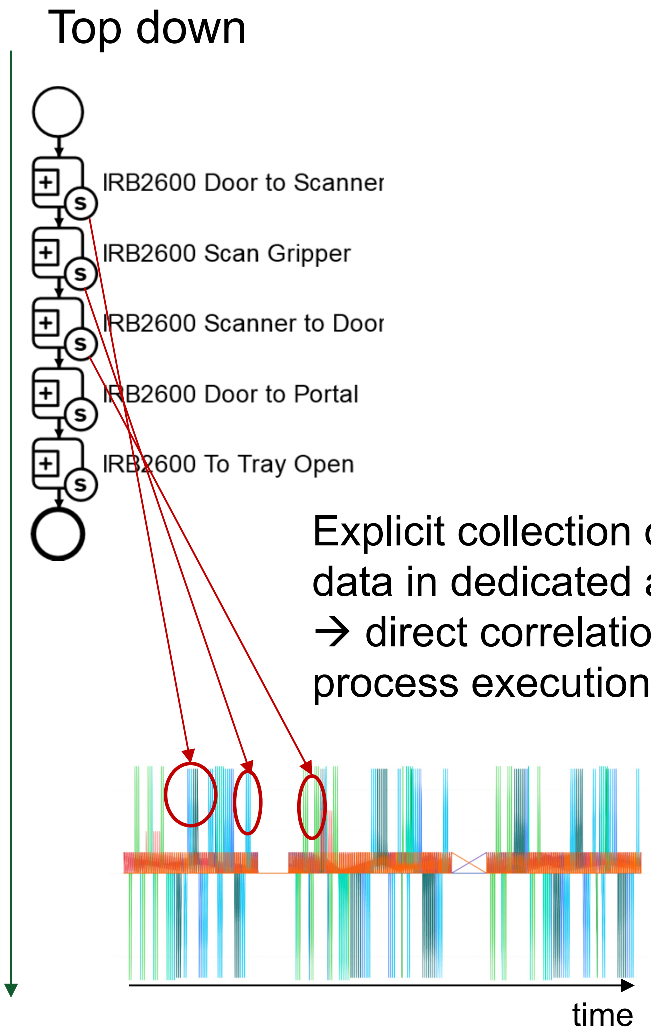




Low → High

Universität St.Gallen

Process-driven Execution and Collection of IoT Data in Context



Explicit collection of sensor data in dedicated activities → direct correlation with process execution

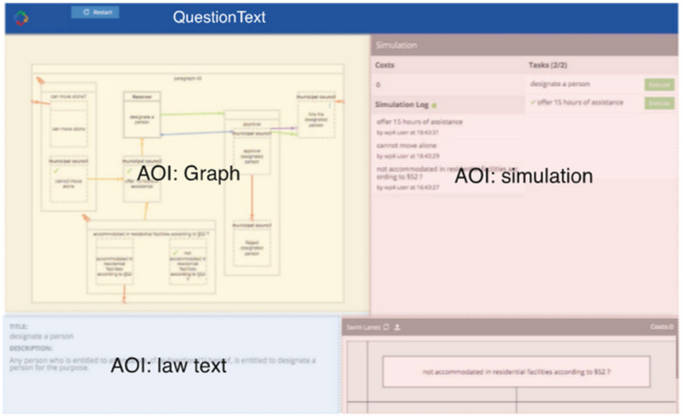
- **IoT data** is collected during process execution and gets **embedded in the broader process context**
- This results in **IoT-enriched event logs** which associate sensor data with the corresponding process execution events
- Allows to calculate IoT-based metrics for different process elements and create augmented process representations

Source: DataStream XES Extension: Embedding IoT Sensor Data into Extensible Event Stream Logs

Tracking Humans Engaging with Software Design Artifacts

Tracking Humans Engaging with Static Software Design Artifacts

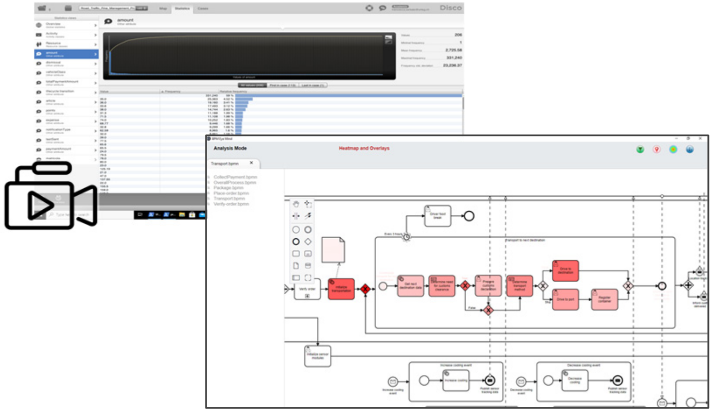
Examples: Fixed screen, images



Modeling specialists versus domain experts using hybrid artifacts in the context of different comprehension tasks

Tracking Humans Engaging with Interactive Software Design Artifacts

Examples: Videos, large software design artifact which require zooming and scrolling, software products with multiple files

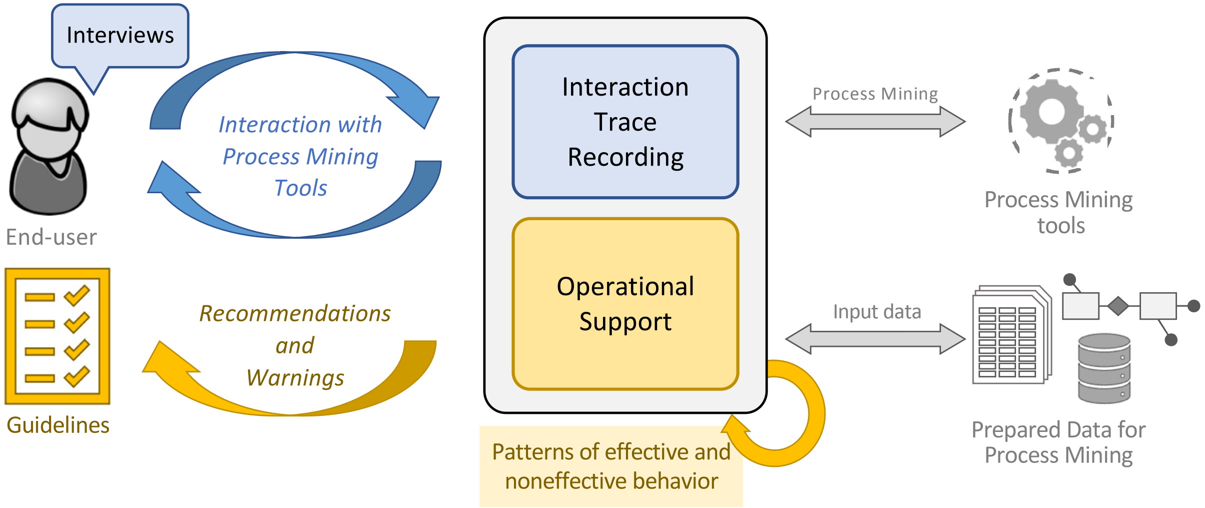


ProMiSE: Process Mining Support for End-Users

Central goal is to gain a comprehensive understanding of the "**process of process mining**"

FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE
Grant No.: 200021 197032

- Concrete outcomes:
  - frequent patterns of effective and noneffective behavior
  - analysis profiles
  - common analysis strategies
  - typical challenges



→ To develop **methodological guidance** and **operational support** to assist **novice analysts** during the analysis effectively

L. Zimmermann, F. Zerbato. and B. Weber: Process Mining Challenges Perceived by Analysts: An Interview Study. BPMDS/EMMSAD@CAiSE 2022: 3-17.

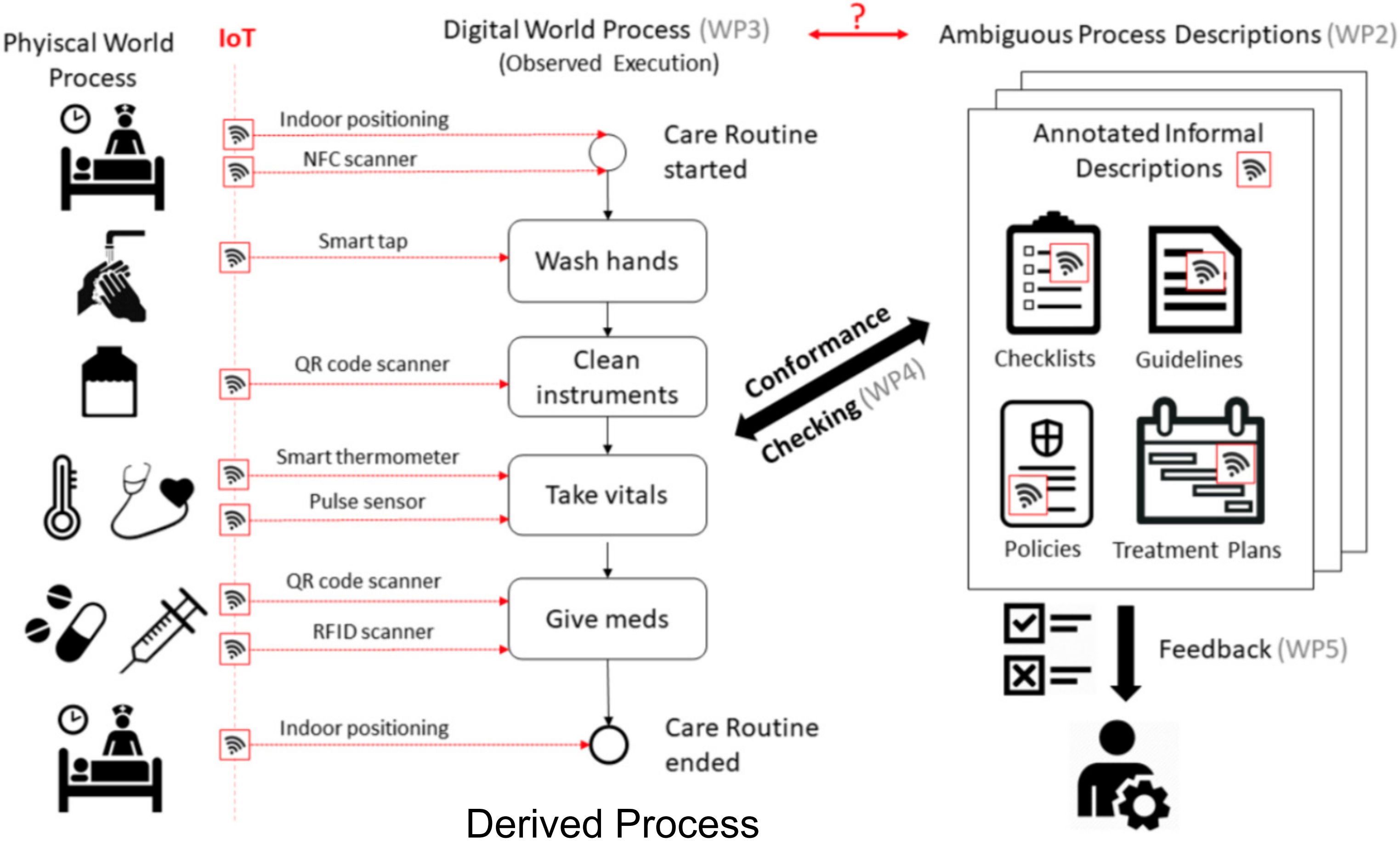
Towards Online Conformance Checking with Ambiguities Driven by the Internet of Things

Fonds national suisse de la recherche scientifique

SPIRIT project 208497

Two application domains

- Industry 4.0 and Industrial IoT
- Health care: clinical guidelines related to hygene

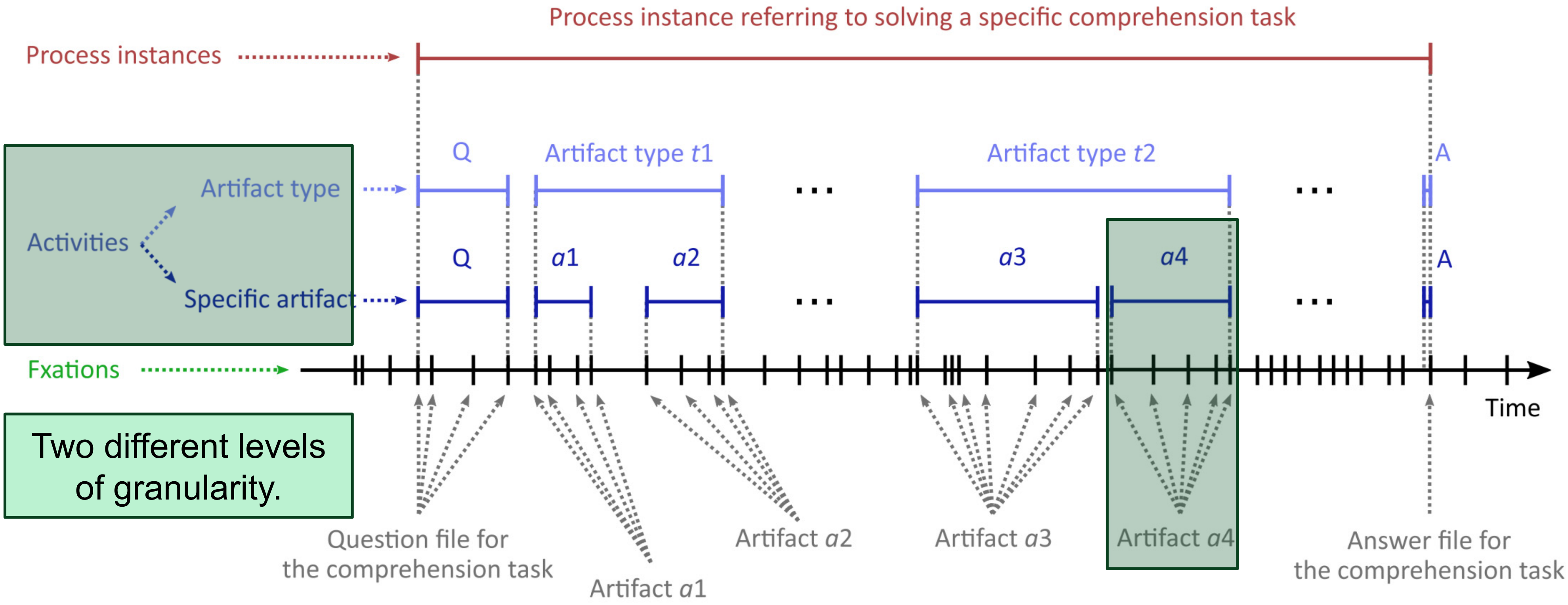


IoT Event Streams

Derived Process Event Streams

Event abstraction: contiguous fixations referring to the same artifact (or artifact type respectively) are grouped in an activity.

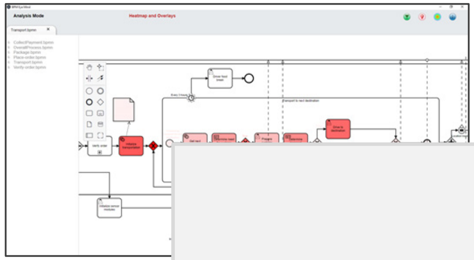
Method for Mining Reading Patterns from Eye-tracking Data



Two different levels of granularity.

Source: Mining reading patterns from eye-tracking data: method and demonstration

Process Observability Largely Differs

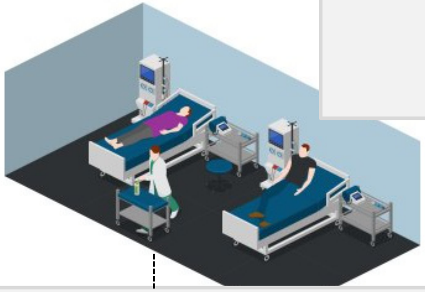


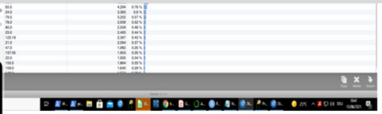
Navigation, scrolling and zooming during model comprehension (depending on tool): large parts of the process occur in

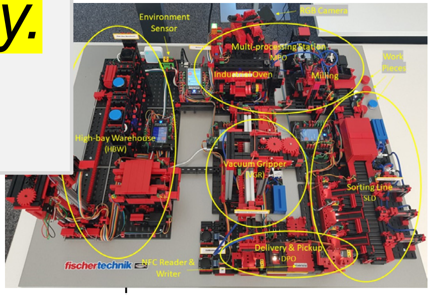
Usage of sensors and additional forms of data collection to increase process observability. Incompleteness and low fidelity of data can be sources of ambiguity.

Process is largely m
most parts perfor
outside of any IT s

uipped with **sensors**
d **actuators**



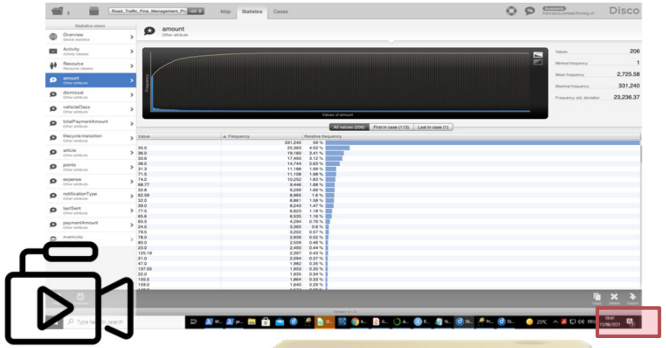




Low

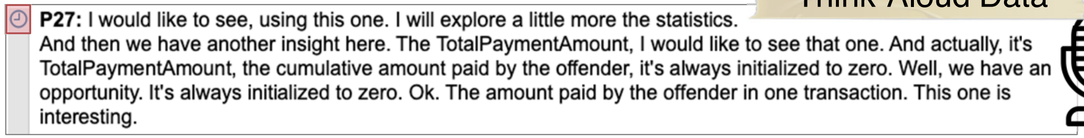
High

Creating User Interaction Logs



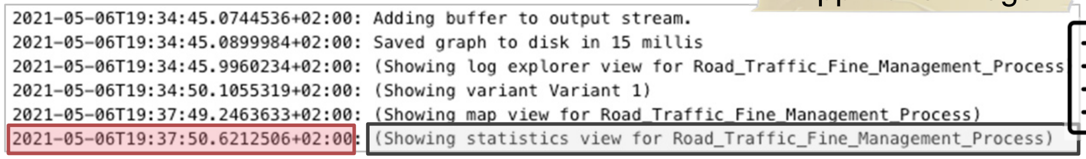
Screen Recording

Think-Aloud Data

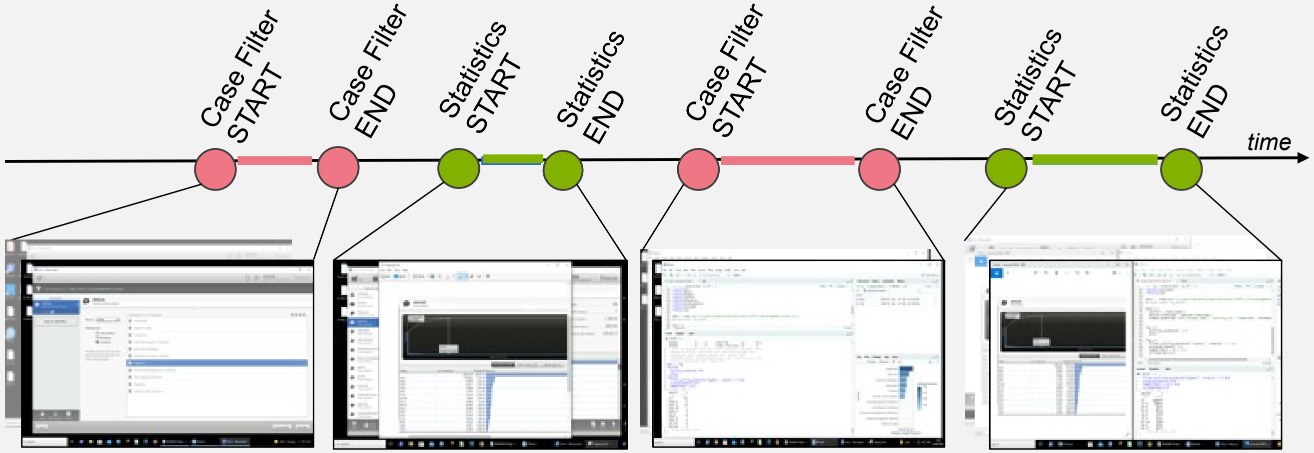


P27: I would like to see, using this one. I will explore a little more the statistics. And then we have another insight here. The TotalPaymentAmount, I would like to see that one. And actually, it's TotalPaymentAmount, the cumulative amount paid by the offender, it's always initialized to zero. Well, we have an opportunity. It's always initialized to zero. Ok. The amount paid by the offender in one transaction. This one is interesting.

Application Logs



```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```

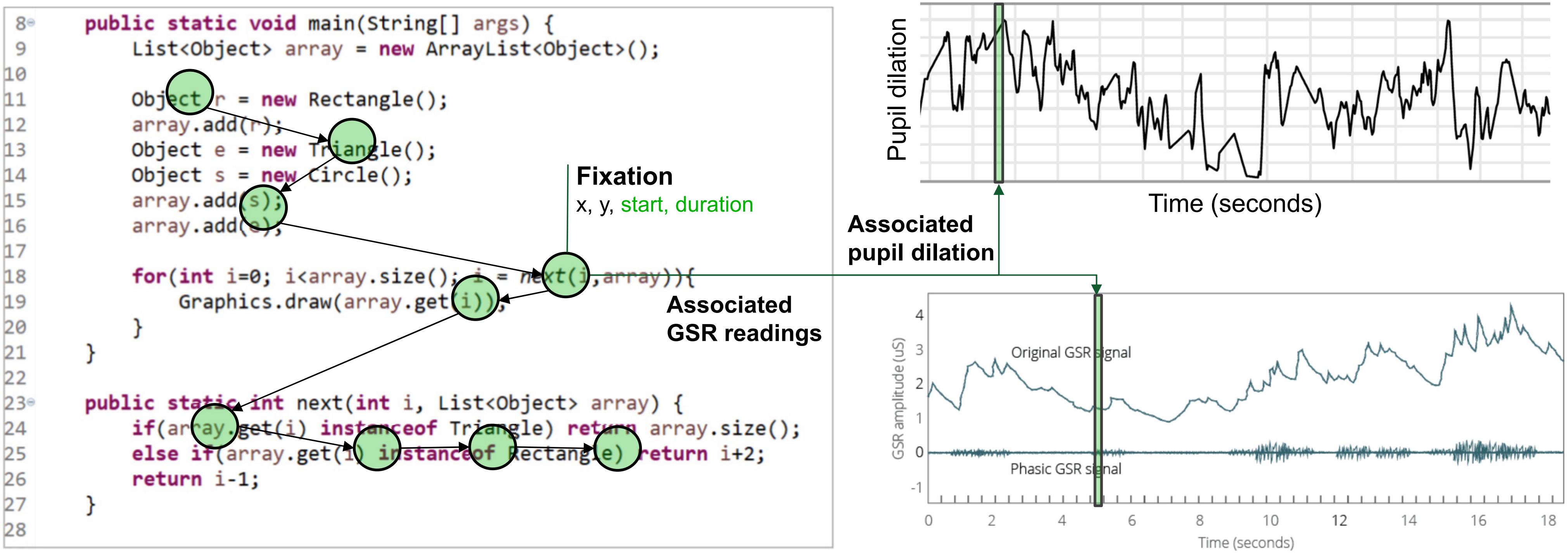


ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:11:46,1	00:12:34,3
P27	Statistics	Disco	00:11:46,1	00:12:34,3
…	…	…	…	…

Temporal Association of Events

Providing Context to Fixations

Challenge: Synchronization of events



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Fixation
x, y, start, duration

Associated pupil dilation

Associated GSR readings

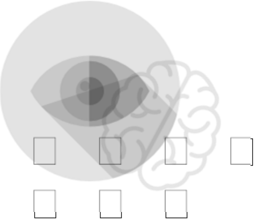
Sequence of fixations over time

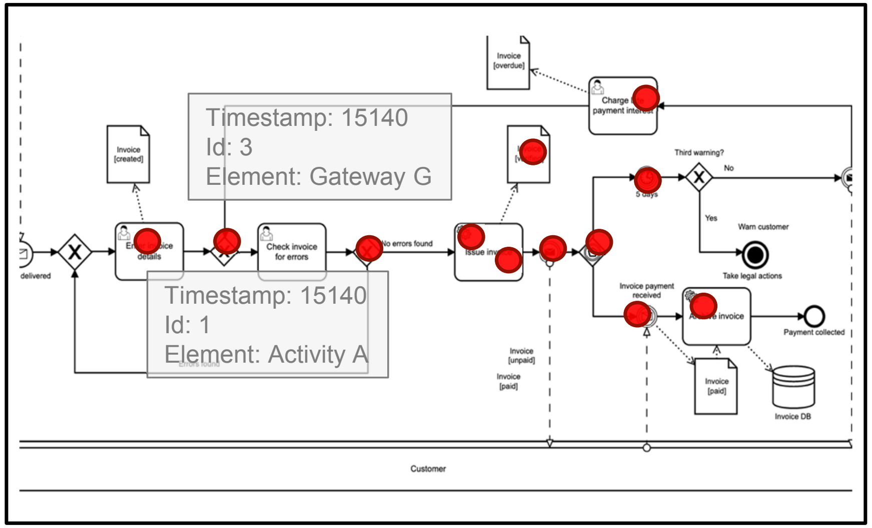
Challenge: Signals differ in terms of latency (time between stimulus and reaction)

Automated Mapping to Software Design Artifacts: The Case of Process Models

Eye-tracker

Data collection

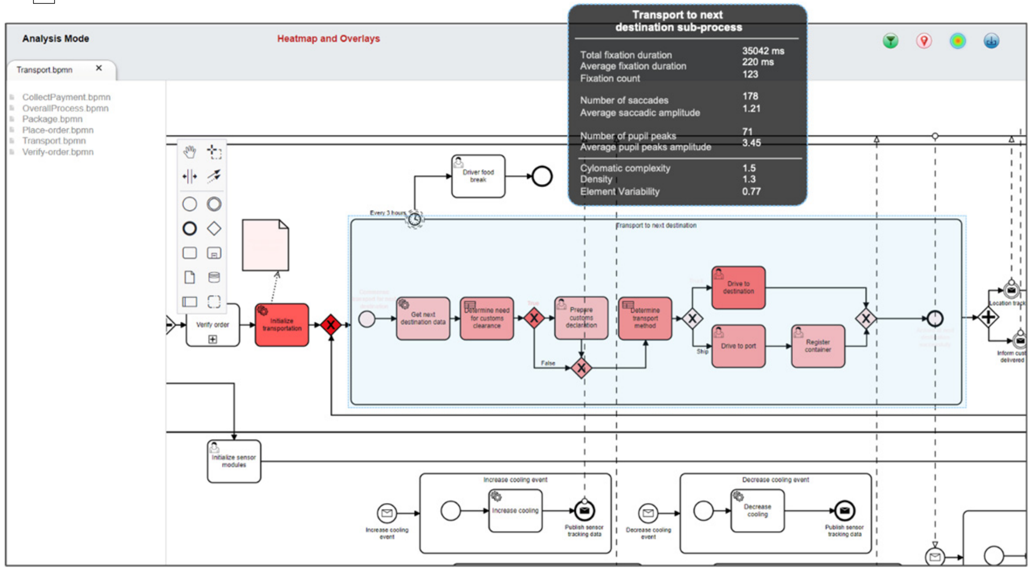




Timestamp: 15140
Id: 3
Element: Gateway G

Timestamp: 15140
Id: 1
Element: Activity A

Analysis



Transport to next destination sub-process

Total fixation duration 35042 ms
Average fixation duration 220 ms
Fixation count 123

Number of saccades 178
Average saccadic amplitude 1.21

Number of pupil peaks 71
Average pupil peaks amplitude 3.45

Cylomatic complexity 1.5
Density 1.3
Element Variability 0.77

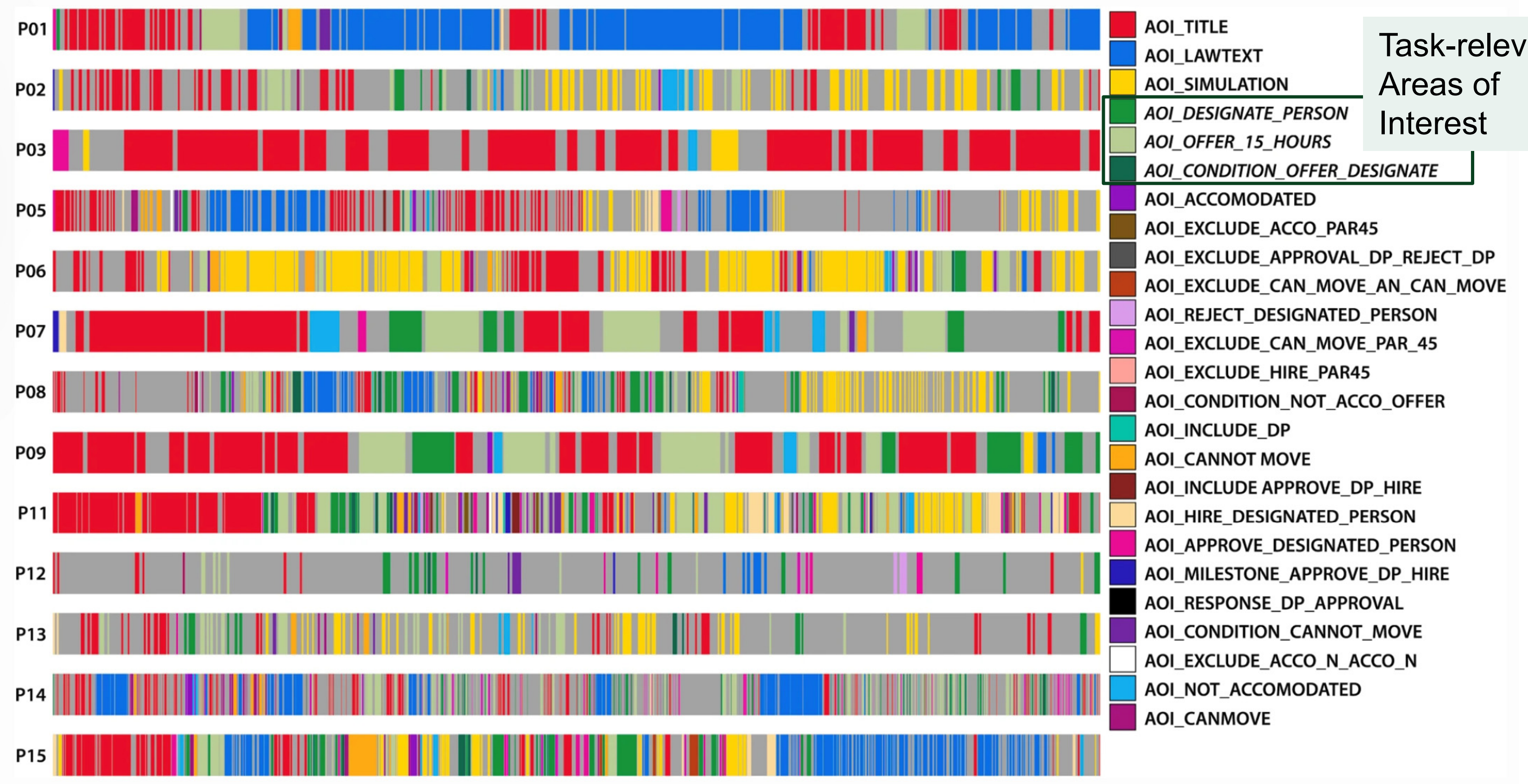
Automated mapping of gazes to process model elements; each element is considered as AOI

Instantanious calculation of AOI-based measures and generation of heatmaps *(without the need to manually define AOIs)*

Source: Prototype developed by Amine Abbad Andaloussi

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.

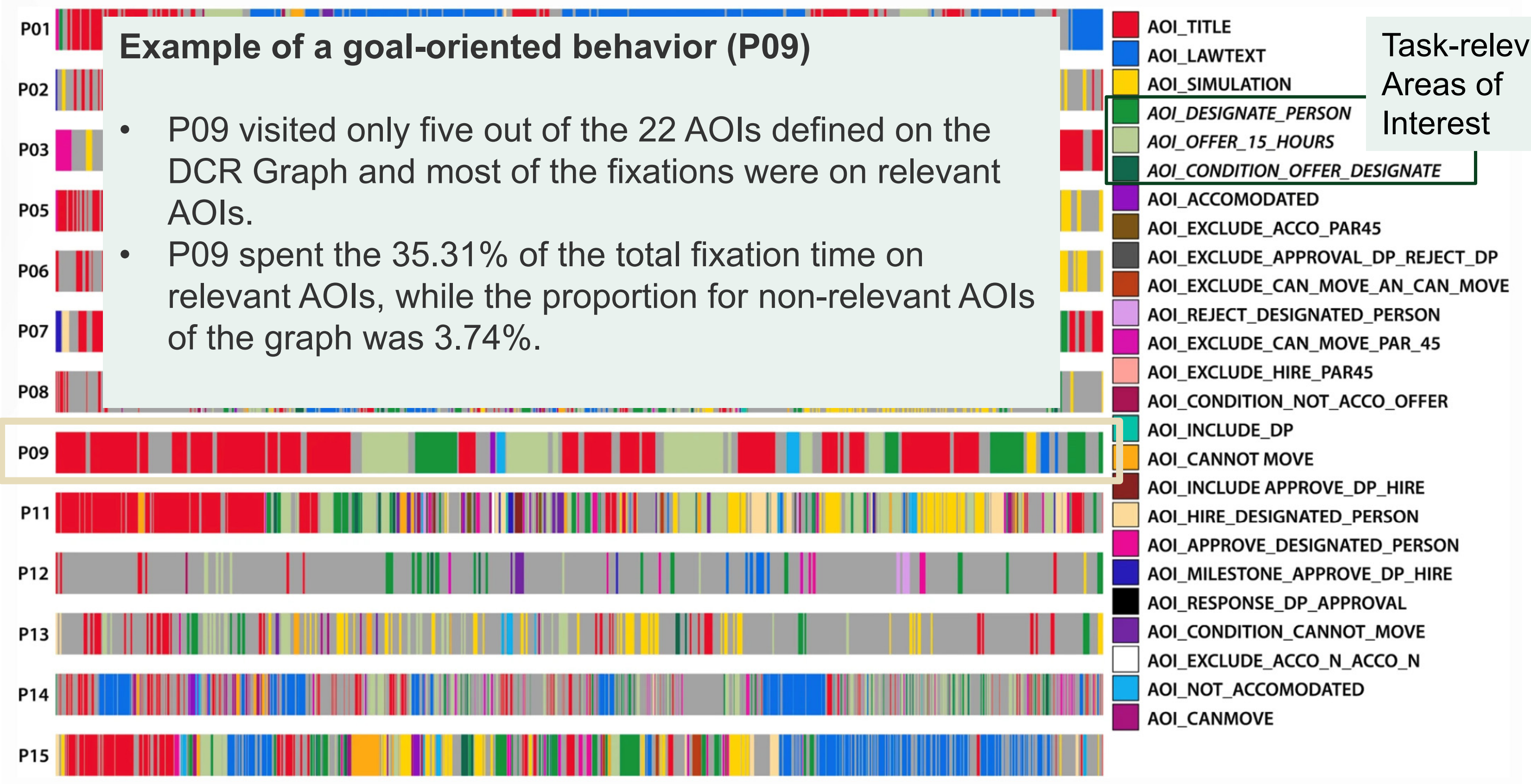


Task-relevant Areas of Interest

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.



Example of a goal-oriented behavior (P09)

- P09 visited only five out of the 22 AOIs defined on the DCR Graph and most of the fixations were on relevant AOIs.
- P09 spent the 35.31% of the total fixation time on relevant AOIs, while the proportion for non-relevant AOIs of the graph was 3.74%.

Task-relevant Areas of Interest

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.

Task-relevant Areas of Interest

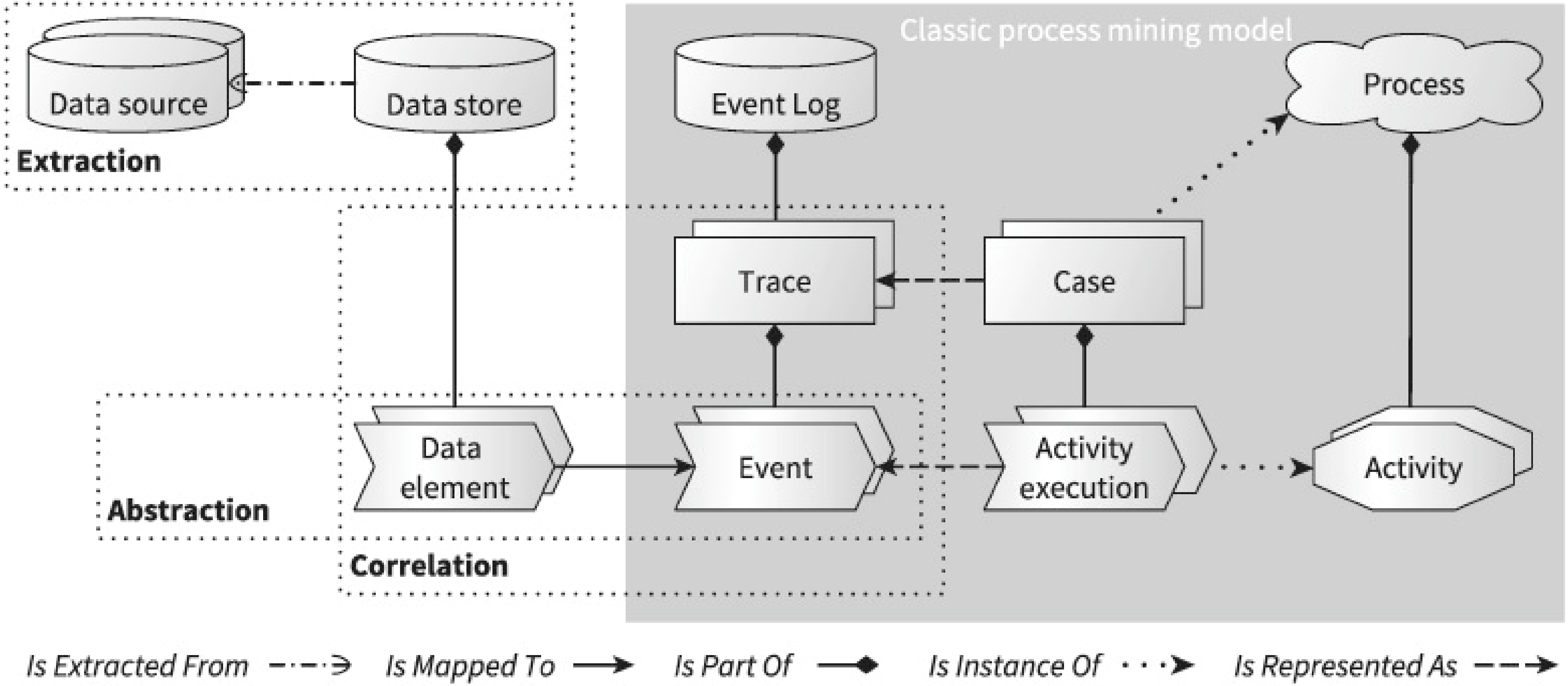
Example of an exploratory strategy (P14)

- P14 visited 16 out of 22 AOIs during the execution of the task. A large amount of these fixations were on non-relevant AOIs. Overall, P14 spent the 29.08% of the total fixation time fixating relevant AOIs in the graph, and the 29.33% of it fixating non-relevant ones.

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

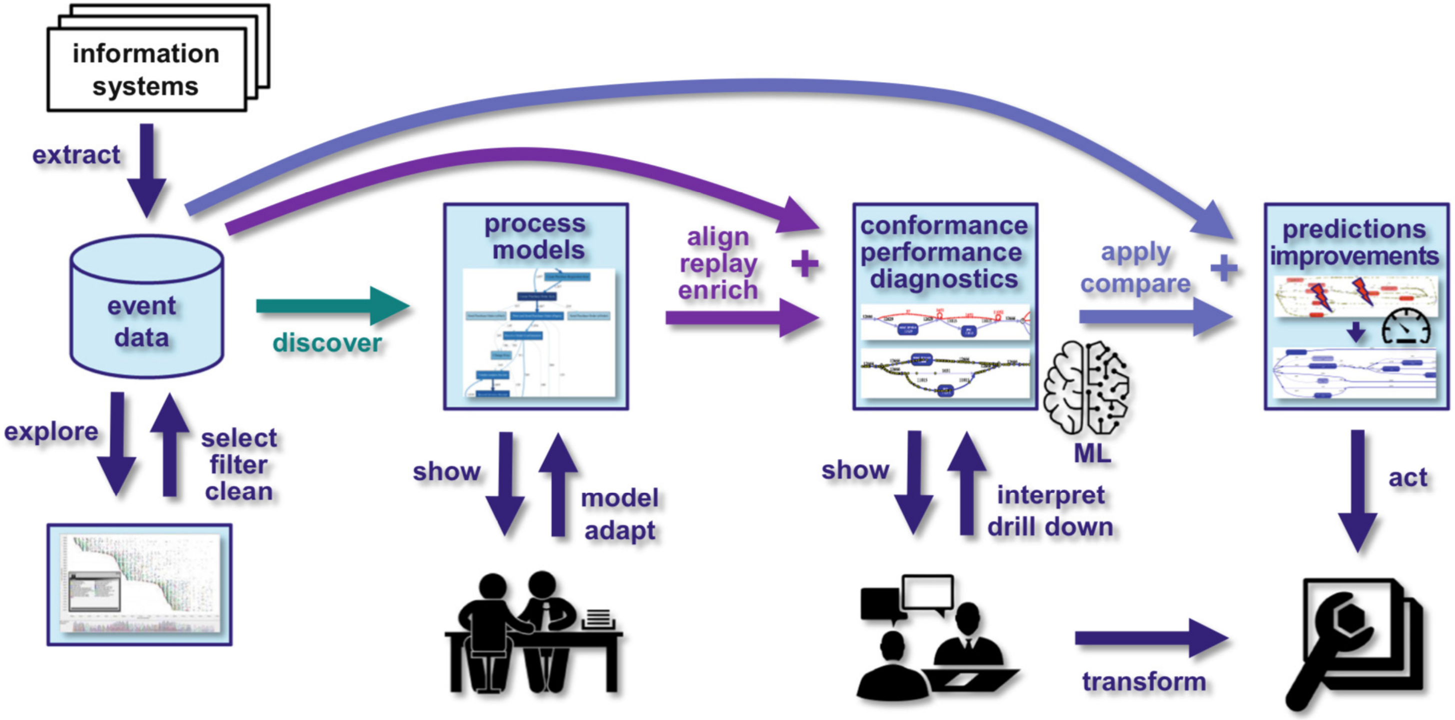
Event Log Preparation

Extraction, correlation, and abstraction of event data



Source: Extraction, correlation, and abstraction of event data for process mining

A 360° Overview of Process Mining

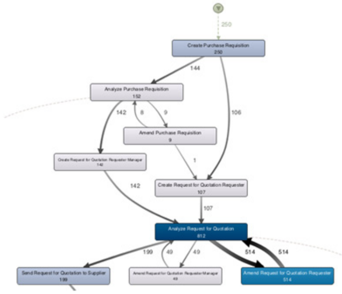


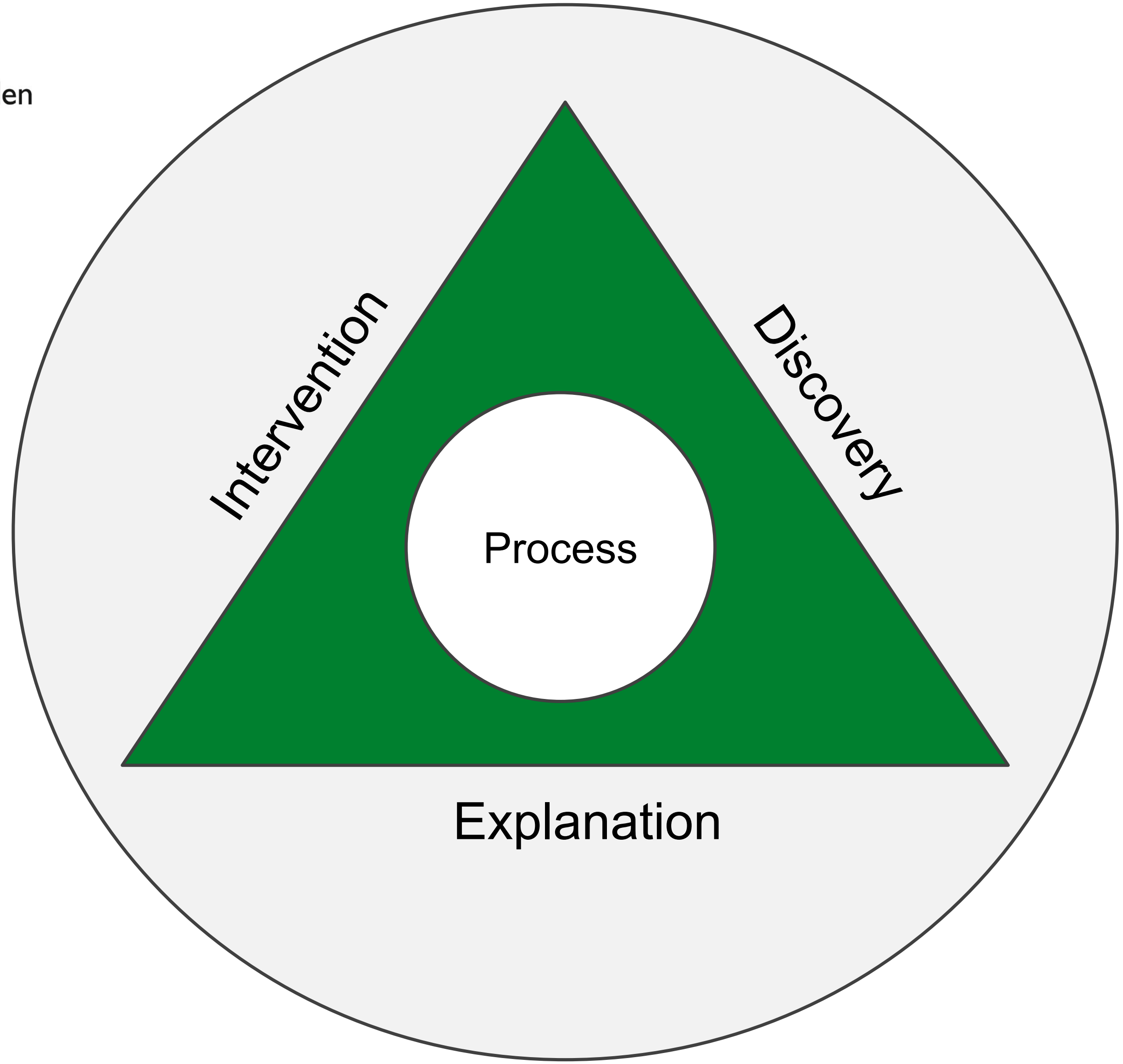
Source: Process Mining Handbook

Universität St.Gallen

Selected Examples

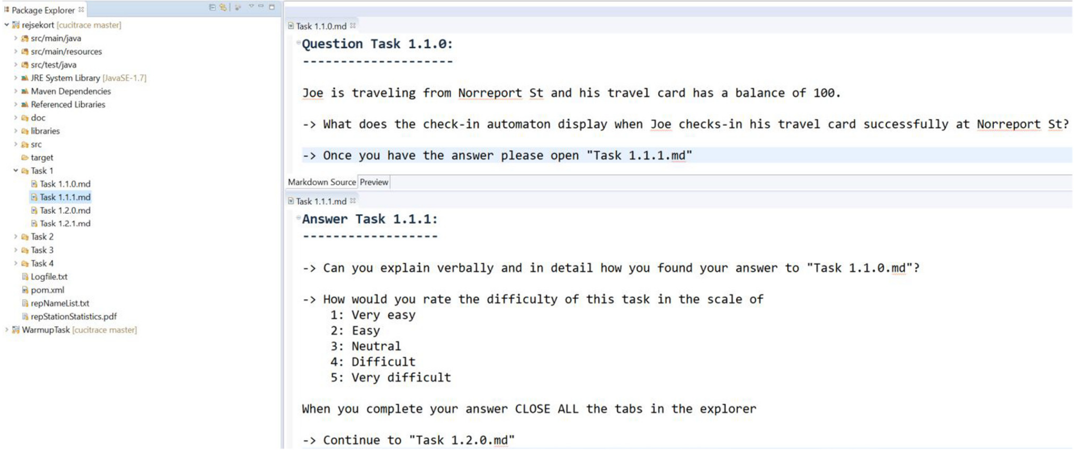
Process of Process Mining





Cognitive Processes

IoT-enabled business processes

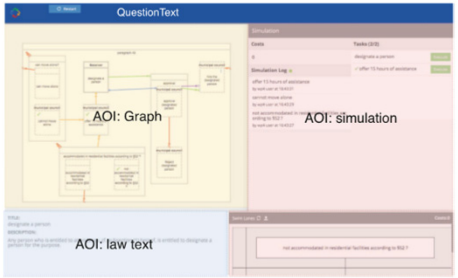


```
Question Task 1.1.0:
--------------------

Joe is traveling from Norreport St and his travel card has a balance of 100.

-> What does the check-in automaton display when Joe checks-in his travel card successfully at Norreport St?

-> Once you have the answer please open "Task 1.1.1.md"
```

```
Answer Task 1.1.1:
------------------

-> Can you explain verbally and in detail how you found your answer to "Task 1.1.0.md"?

-> How would you rate the difficulty of this task in the scale of
    1: Very easy
    2: Easy
    3: Neutral
    4: Difficult
    5: Very difficult

When you complete your answer CLOSE ALL the tabs in the explorer

-> Continue to "Task 1.2.0.md"
```

State-of-the art

In terms of process mining this boils down to event abstraction



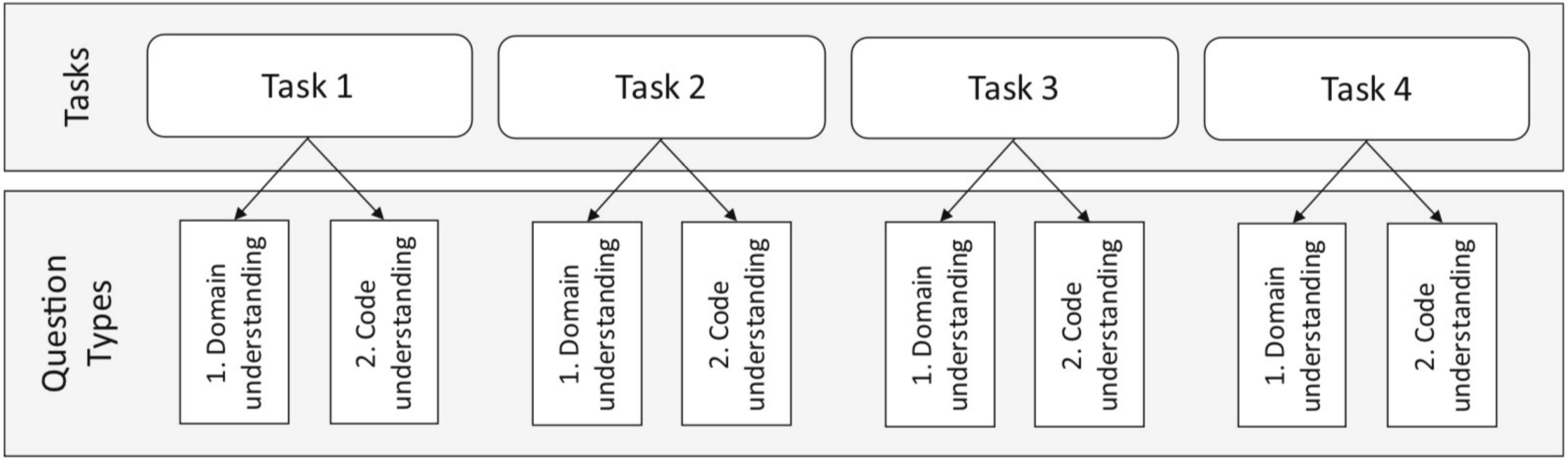
Requires the association of fixations to elements of the artifact

Static Artifacts

Dynamic Artifacts; critical to go beyond small examples and look at real system

The Process: Source Code Reading

Developers answering 8 questions related to 4 tasks which require them to use different code-related artifacts

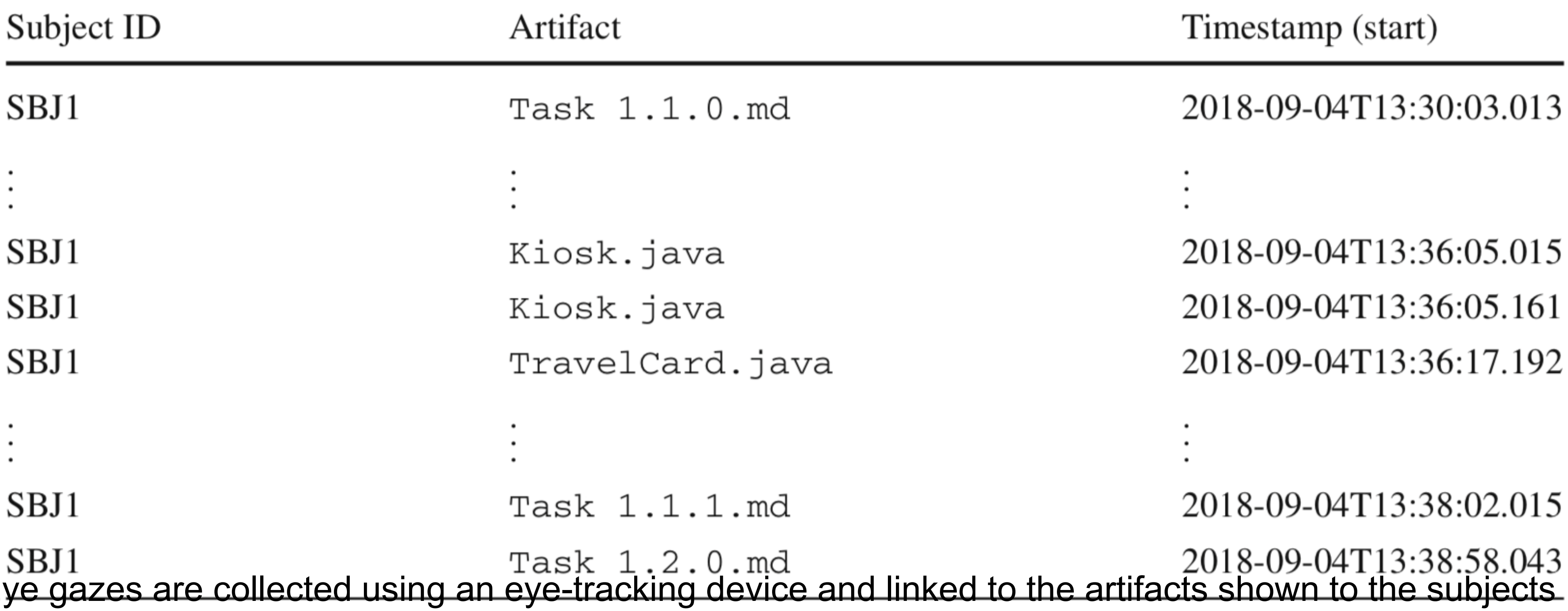


Source: Mining reading patterns from eye-tracking data: method and demonstration

Step 1: Collect Data

Event extraction: eye gazes from an eye-tracker linked to the artifact are used for event data extraction.

Collected Data (After Some Preparation)



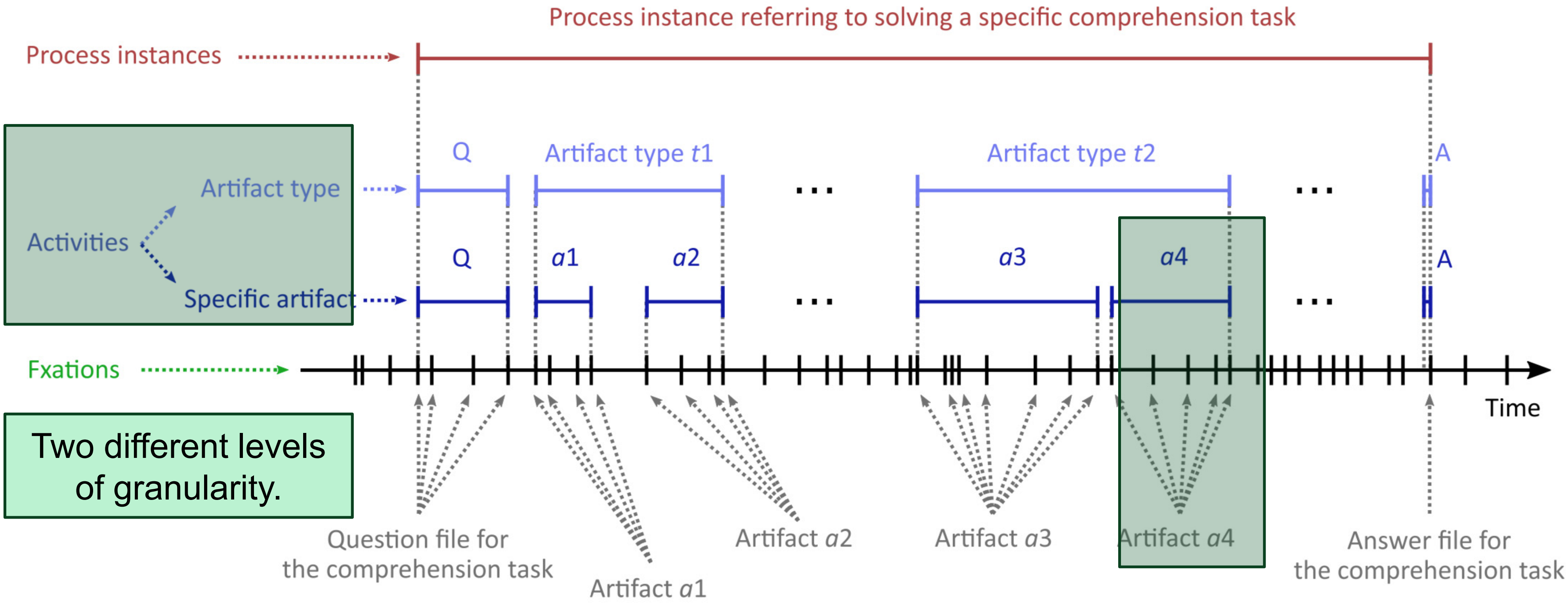
Subject ID	Artifact	Timestamp (start)
SBJ1	`Task 1.1.0.md`	2018-09-04T13:30:03.013
⋮	⋮	⋮
SBJ1	`Kiosk.java`	2018-09-04T13:36:05.015
SBJ1	`Kiosk.java`	2018-09-04T13:36:05.161
SBJ1	`TravelCard.java`	2018-09-04T13:36:17.192
⋮	⋮	⋮
SBJ1	`Task 1.1.1.md`	2018-09-04T13:38:02.015
SBJ1	`Task 1.2.0.md`	2018-09-04T13:38:58.043

Eye gazes are collected using an eye-tracking device and linked to the artifacts shown to the subjects

Source: Mining reading patterns from eye-tracking data: method and demonstration

Prep

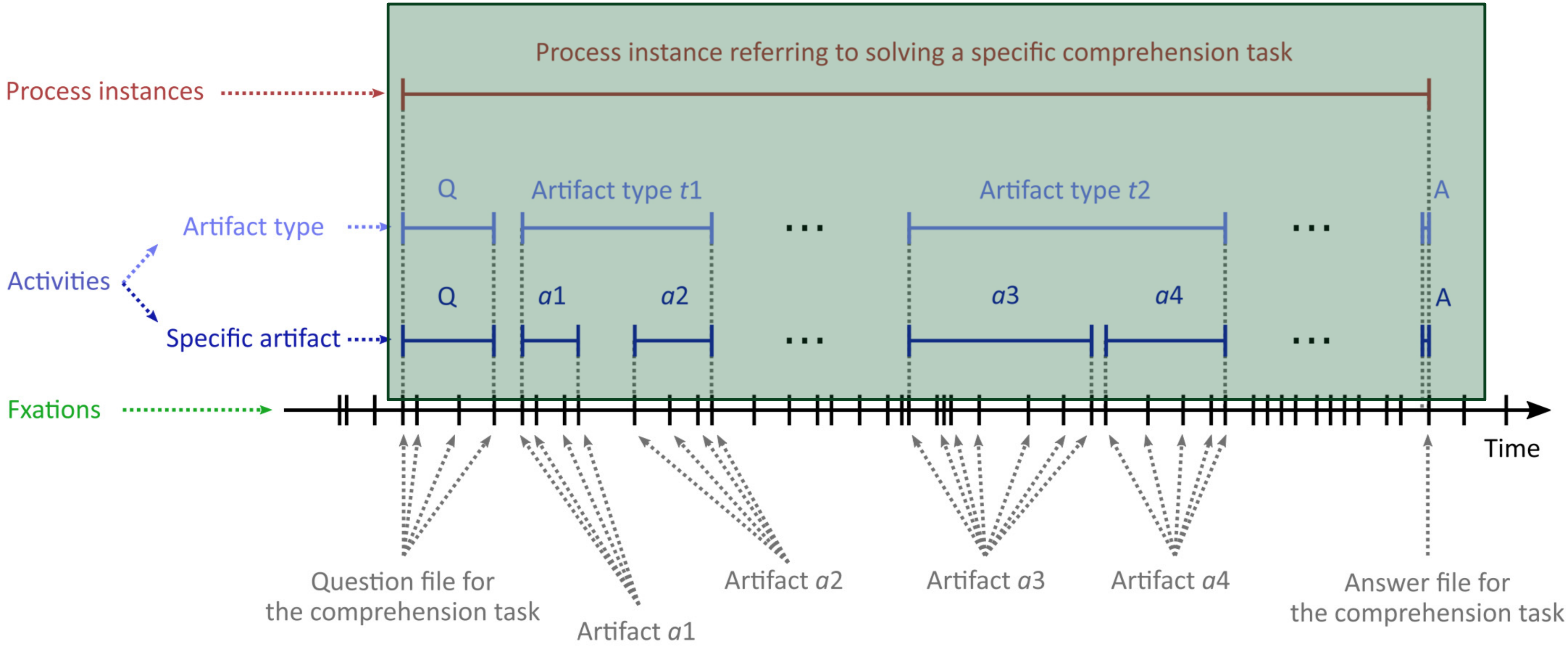
Event abstraction: contiguous fixations referring to the same artifact (or artifact type respectively) are grouped in an activity.



Two different levels of granularity.

Source: Mining reading patterns from eye-tracking data: method and demonstration

Event correlation: All fixation events between the first fixation on a question and the first fixation on the corresponding answer are considered to belong to the same process instance.



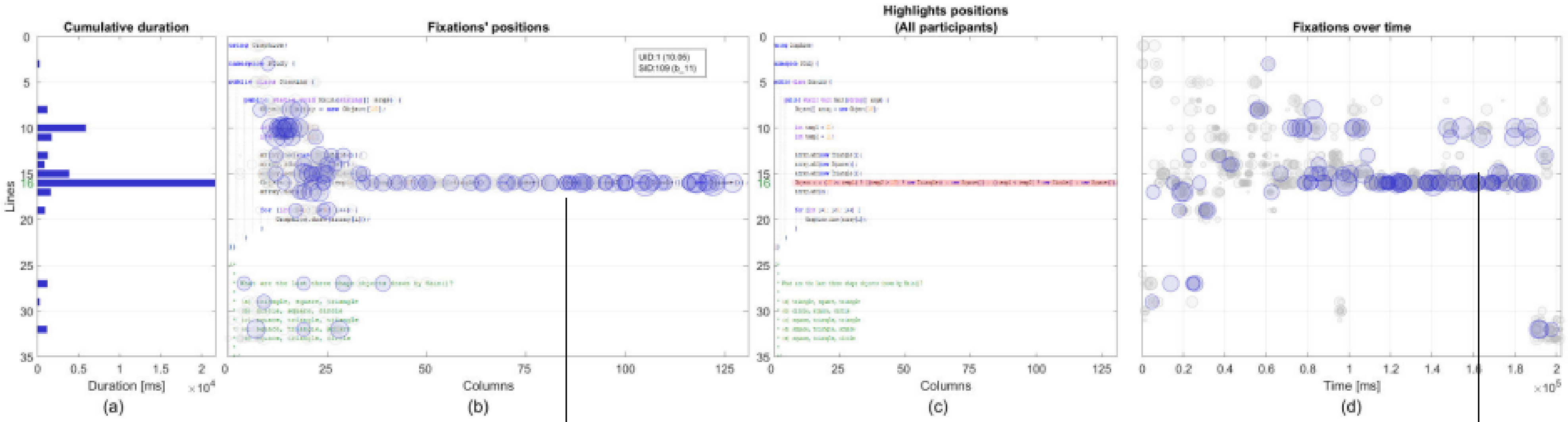
Source: Mining reading patterns from eye-tracking data: method and demonstration

Additional Steps in the Method

- Validate Data
  - Recording referring to subjects for which irregularities during data collection were observed were removed
- Partition Data
  - Map with aggregated behavior
  - Split data according to certain criteria (e.g., subject properties, task properties, answer properties)
- Mine the Reading Patterns
- Interpret the Results

Source: Mining reading patterns from eye-tracking data: method and demonstration

Eye-Movements During Source-code Reading



In (a) only long fixations (above 250ms) are considered. In (b) and (d), short fixations (below 250ms) are marked in gray while long fixations are marked in blue

Visual attention on line 16 particularly high.

Participant switches attention between line 16 and lines 10/11 (visual association)

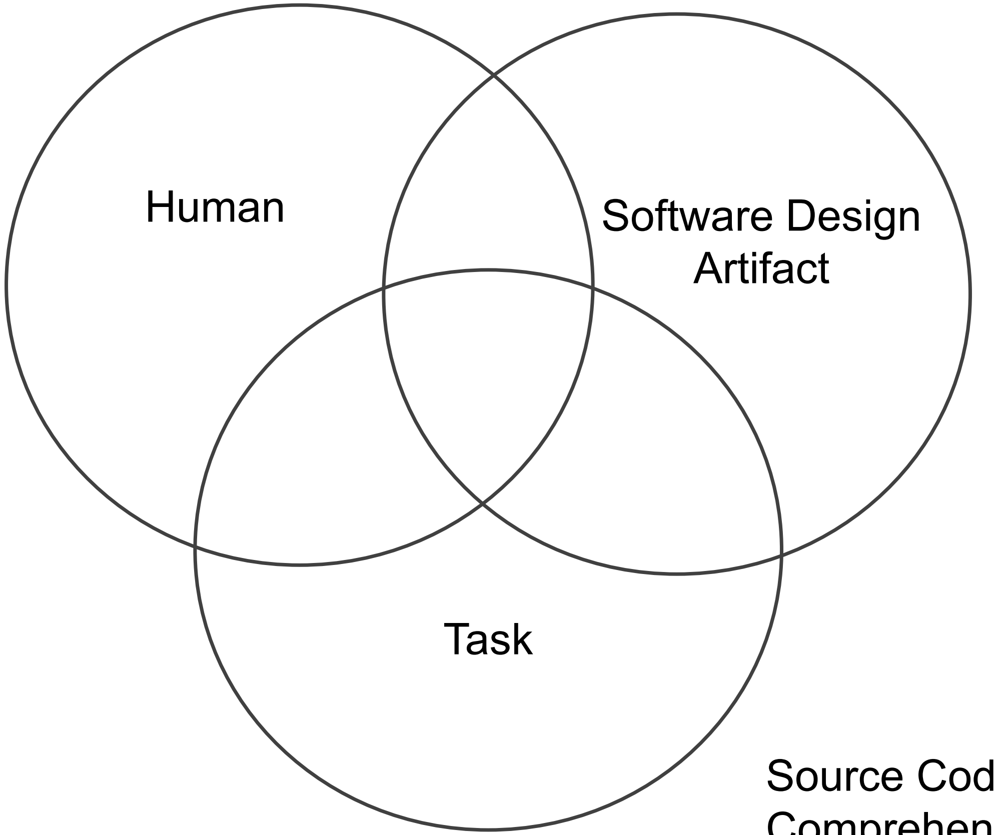
Source: Towards a Fine-grained Analysis of Cognitive Load During Program Comprehension

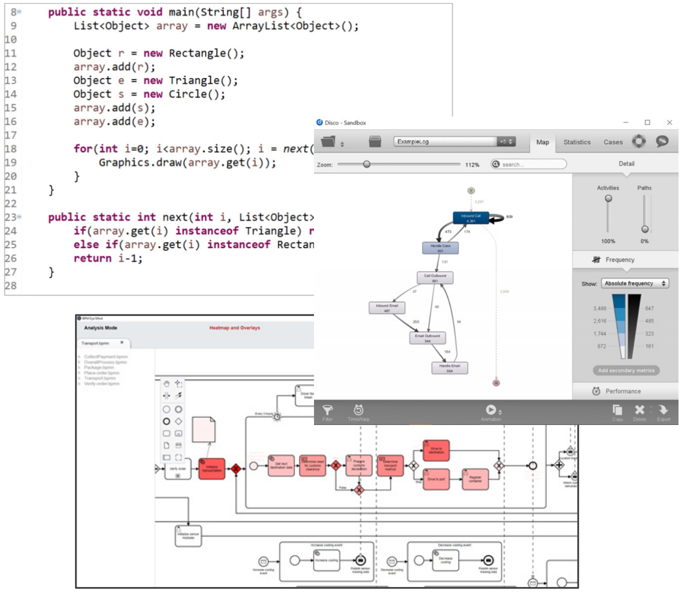
Universität St.Gallen

Developers

…

…

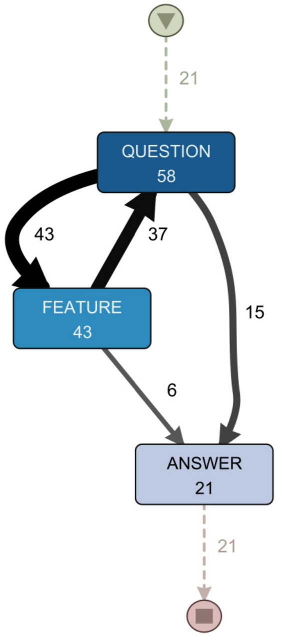




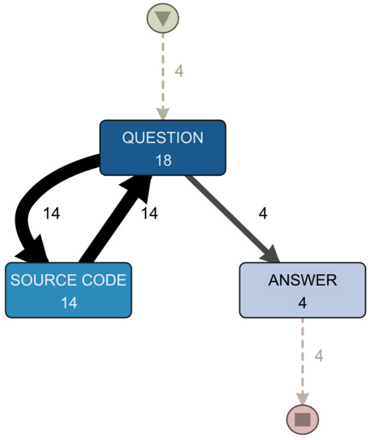
```
    public static void main(String[] args) {
        List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object>
        if(array.get(i) instanceof Triangle) r
        else if(array.get(i) instanceof Rectan
        return i-1;
    }

```

Source Code Comprehension

Artifact Types Used for Answering Domain Understanding Questions

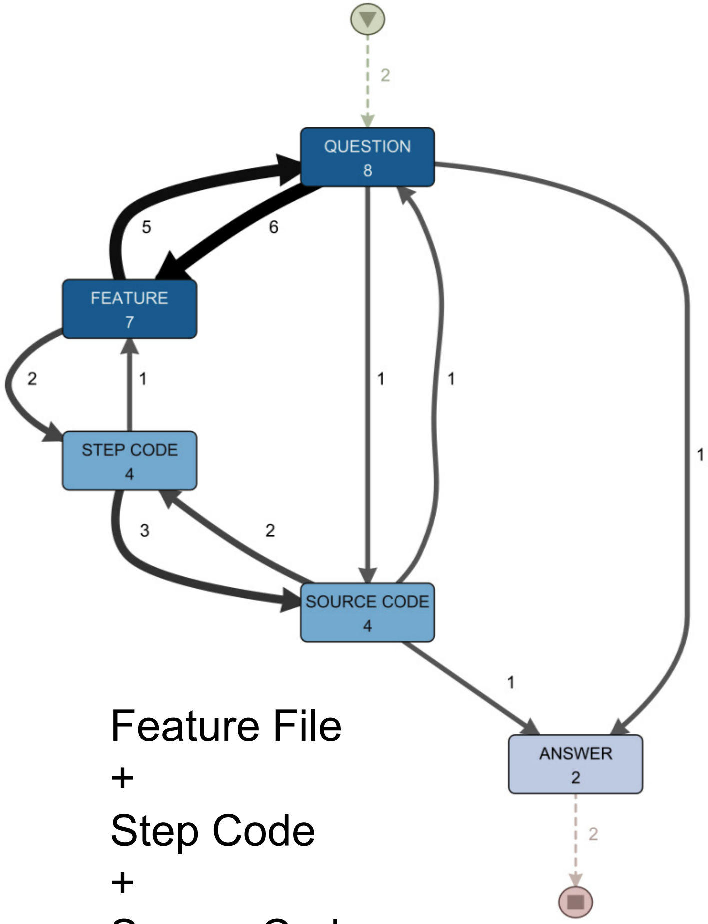


Feature File



Source Code

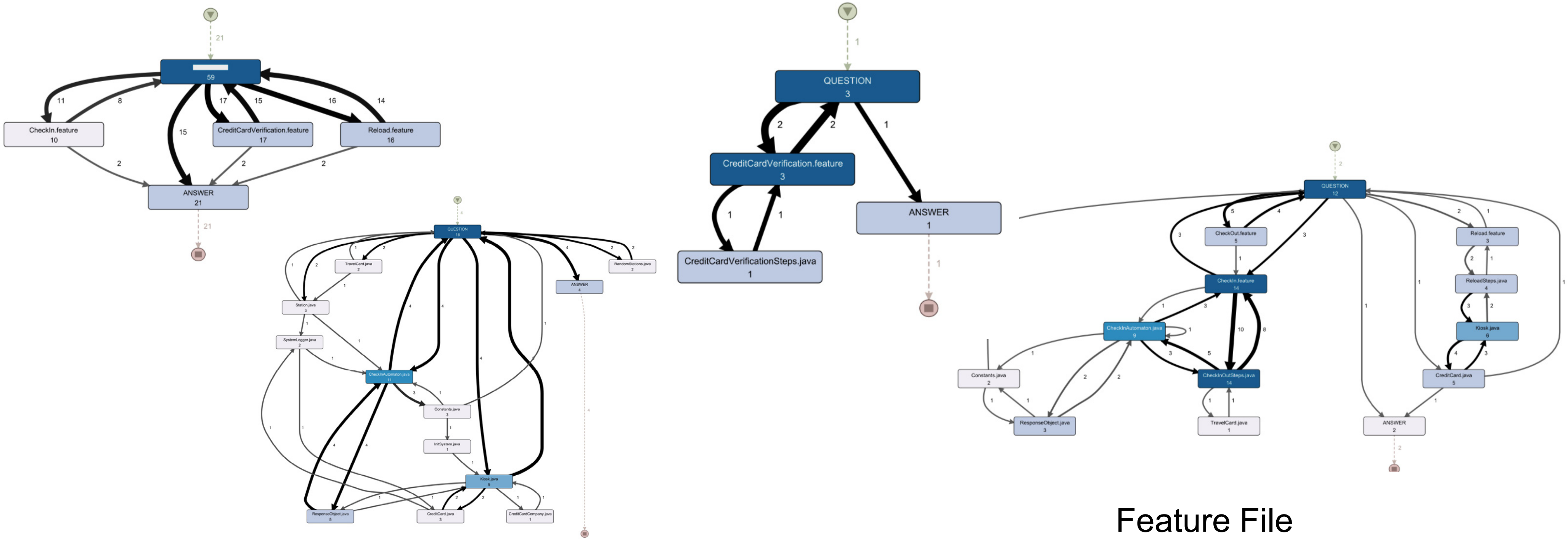
Feature File
+
Step Code



Feature File
+
Step Code
+
Source Code

Source: Mining reading patterns from eye-tracking data: method and demonstration

Files Used for Answering Domain Understanding Questions



Feature File

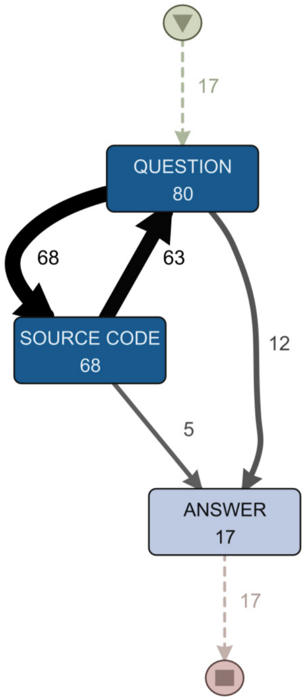
Source Code

Feature File
+
Step Code

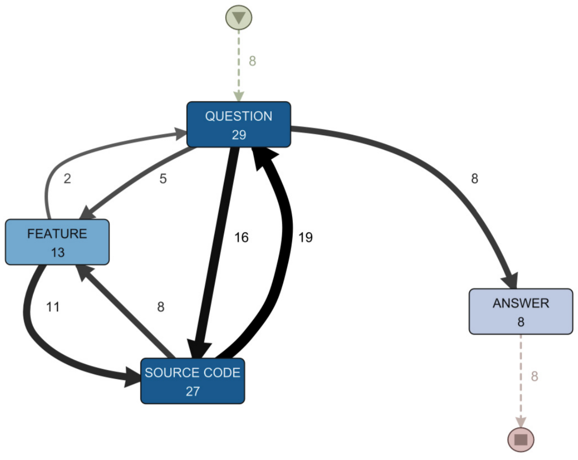
Feature File
+
Step Code
+
Source Code

Source: Mining reading patterns from eye-tracking data: method and demonstration

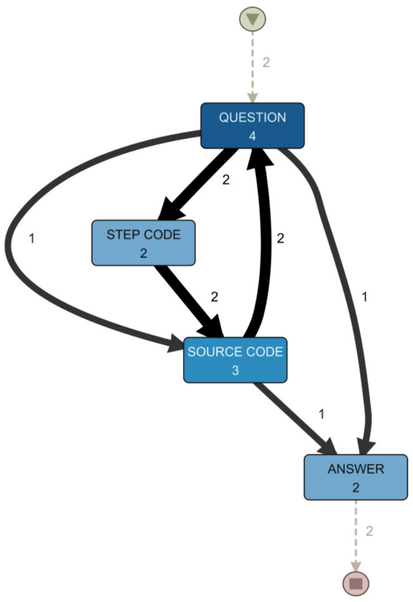
Artifact Types Used for Answering Code Understanding Questions



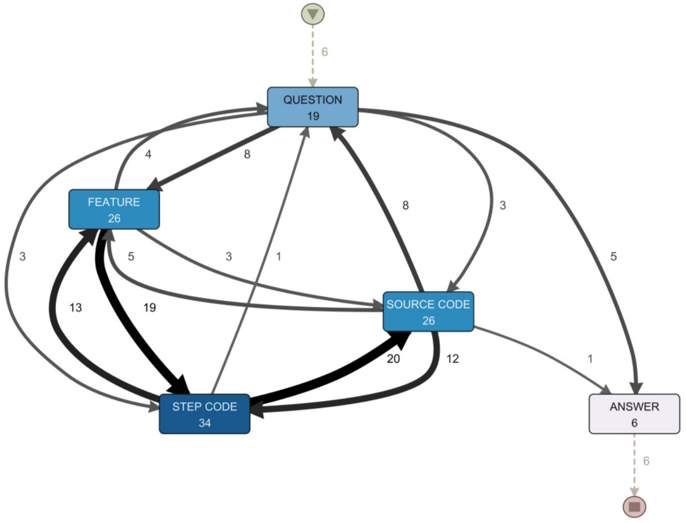
Source Code



Feature File
+
Step Code



Source Code
+
Step Code



Feature File
+
Step Code
+
Source Code

Source: Mining reading patterns from eye-tracking data: method and demonstration

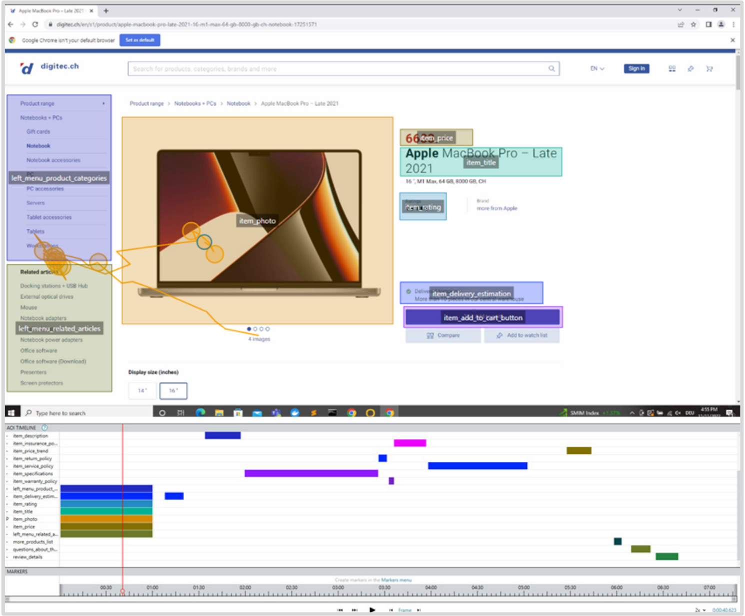
EyeMind and iTrace

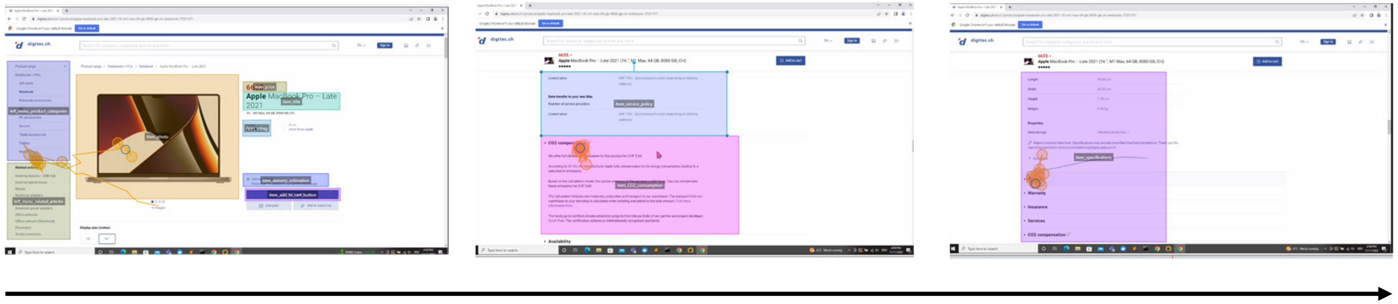
- Tool for Collecting and Analyzing
- Enriched Fixation Events

Manual Mapping of Dynamic AOIs

The challenge with dynamic stimuli

- Dynamic stimuli require to play the gaze video recording and manually asign/adjust areas of interest everytime the content of the screen is changed (e.g., with scrolling)
- Typically, manual and very time consuming process



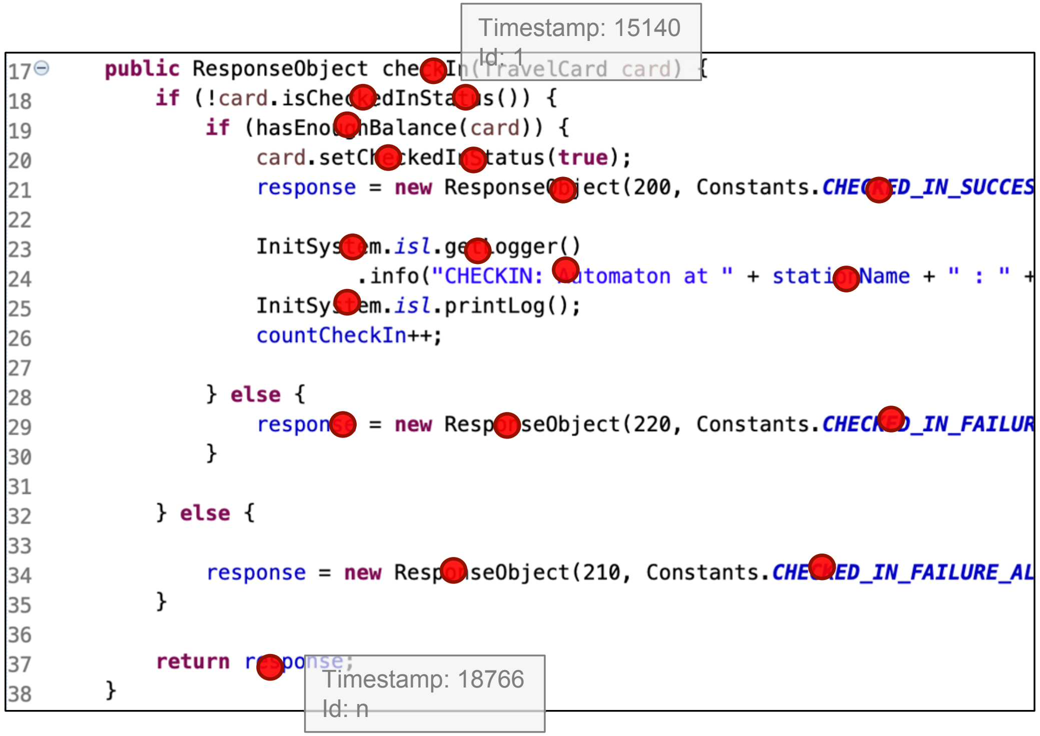


Scrolling can change the screen content within the interval of few seconds

Universität St.Gallen

Automated Mapping to Areas of Interest: The Case of Source-code

Eye-tracker



```
public ResponseObject checkIn(TravelCard card) {
    if (!card.isCheckedInStatus()) {
        if (hasEnoughBalance(card)) {
            card.setCheckedInStatus(true);
            response = new ResponseObject(200, Constants.CHECKED_IN_SUCCES

            InitSystem.isl.getLogger()
                    .info("CHECKIN: Automaton at " + stationName + " : " +
            InitSystem.isl.printLog();
            countCheckIn++;

        } else {
            response = new ResponseObject(220, Constants.CHECKED_IN_FAILUR
        }

    } else {

        response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_AL
    }

    return response;
}
```

Sample of user’s gazes on the source-code

iTrace

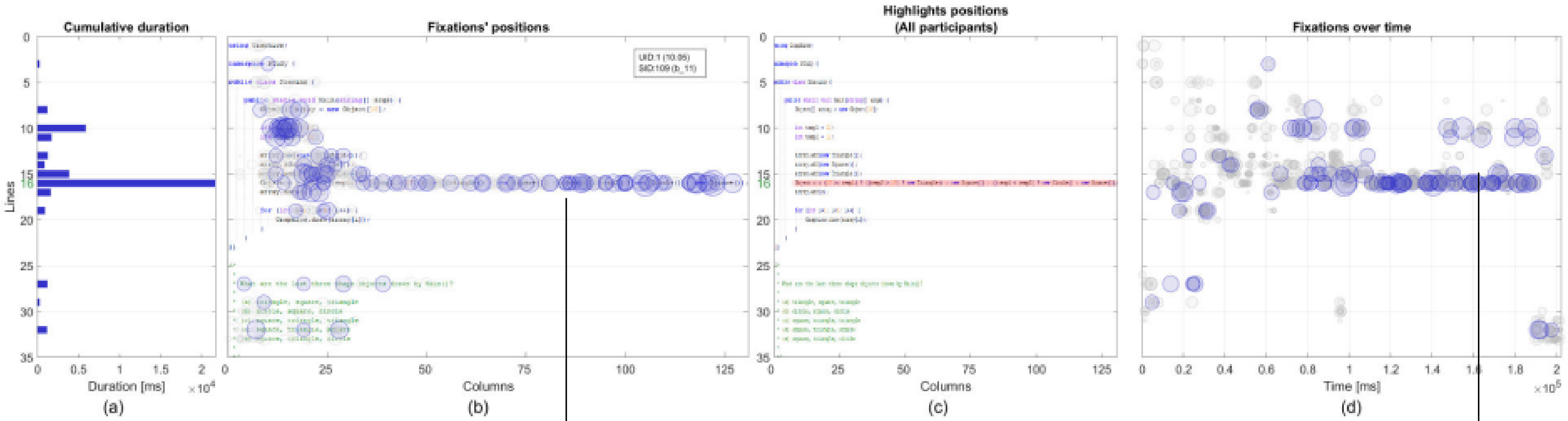
Automated mapping between gazes and areas of interest at the data collection

Gaze ID	Timestamp	Gaze X Pos.	Gaze Y Pos.	Line of code	Column of code
1	15140	1450	60	17	30
2	15148	1430	77	18	26
3	15156	1480	77	18	32
4	15164	1530	86	19	25
..	..	..	..	..	..

Sample of gaze file with automatically mapped AOIs

Source: https://www.i-trace.org

Eye-Movements During Source-code Reading



(a) (b) (c) (d)

In (a) only long fixations (above 250ms) are considered. In (b) and (d), short fixations (below 250ms) are marked in gray while long fixations are marked in blue

Visual attention on line 16 particularly high.

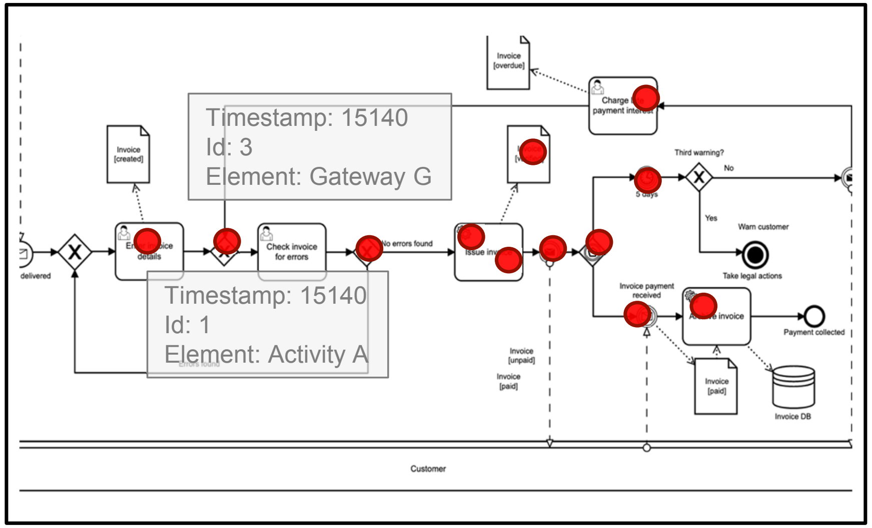
Participant switches attention between line 16 and lines 10/11 (visual association)

Source: Towards a Fine-grained Analysis of Cognitive Load During Program Comprehension

Automated Mapping to Areas of Interest: The Case of Process Models

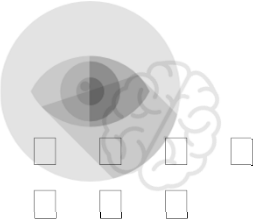
Eye-tracker

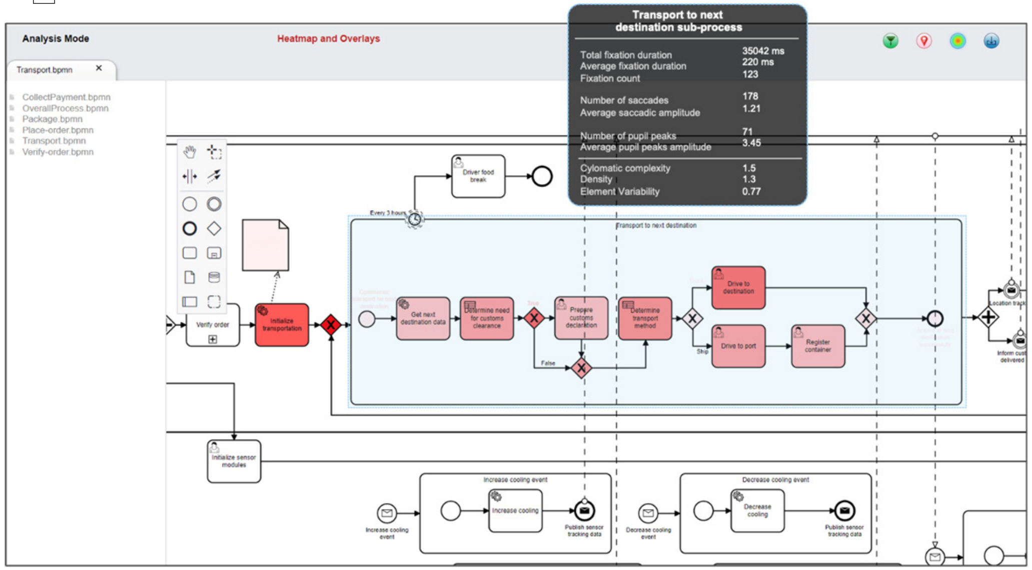
Data collection


Timestamp: 15140
Id: 3
Element: Gateway G

Timestamp: 15140
Id: 1
Element: Activity A

Analysis




Transport to next destination sub-process

Total fixation duration 35042 ms
Average fixation duration 220 ms
Fixation count 123

Number of saccades 178
Average saccadic amplitude 1.21

Number of pupil peaks 71
Average pupil peaks amplitude 3.45

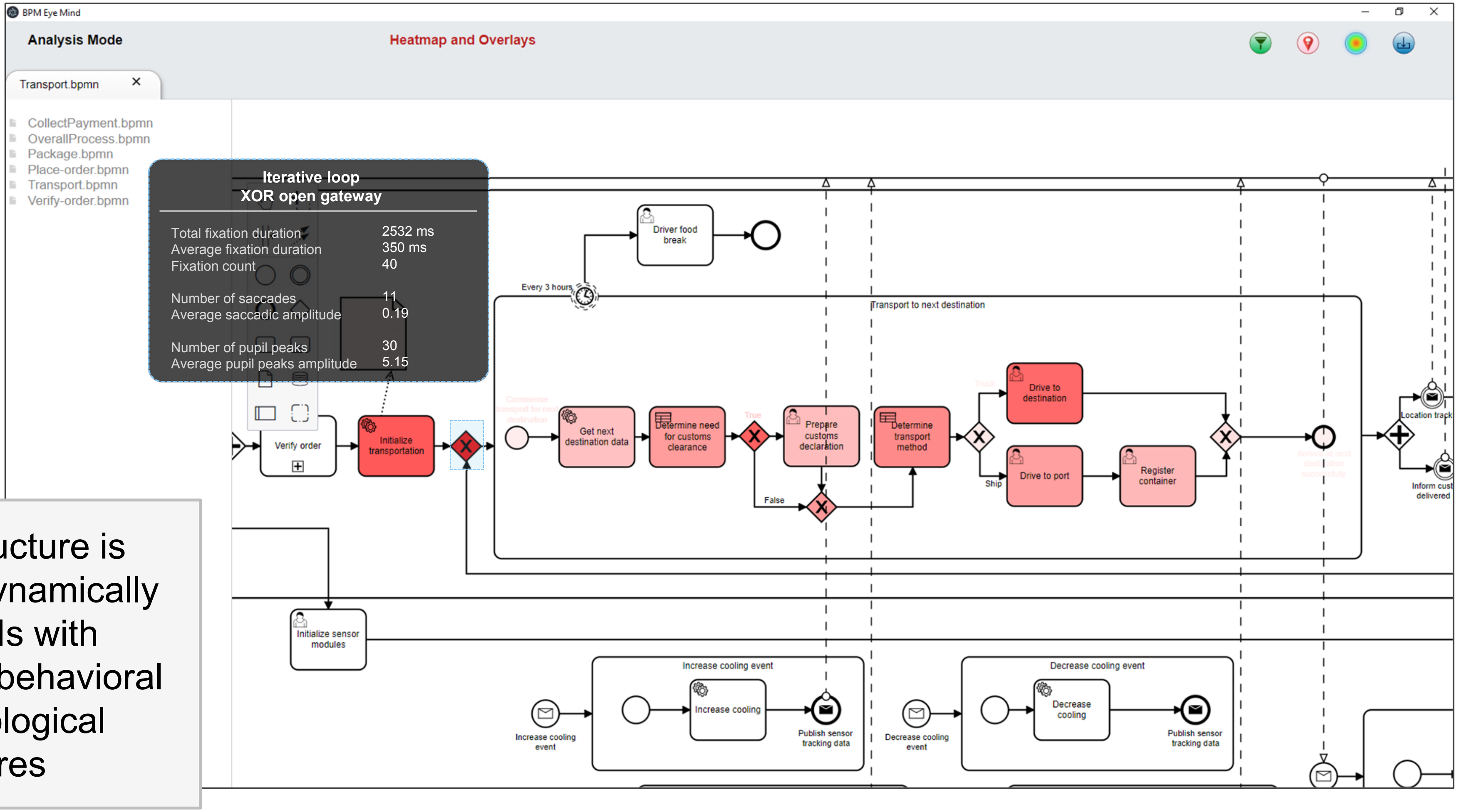
Cylomatic complexity 1.5
Density 1.3
Element Variability 0.77

Automated mapping of gazes to process model elements

Instantanious calculation of AOI-based measures and generation of heatmaps
(without the need to manually define AOIs)

Source: Prototype developed by Amine Abbad Andaloussi

Eye Mind: Process Model Augmented with Eye-tracking Metrics



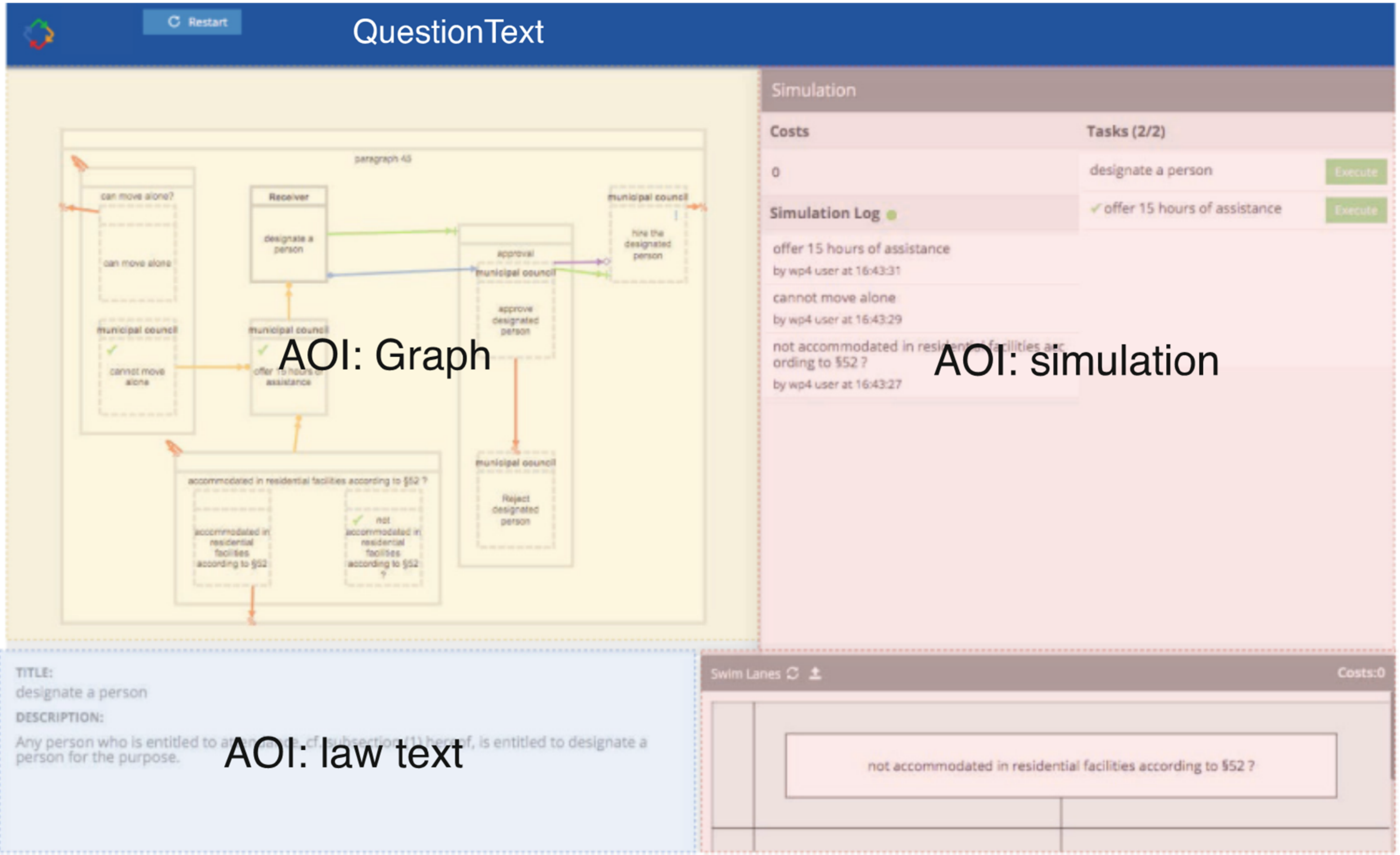
Iterative loop
XOR open gateway

Total fixation duration	2532 ms
Average fixation duration	350 ms
Fixation count	40
Number of saccades	11
Average saccadic amplitude	0.19
Number of pupil peaks	30
Average pupil peaks amplitude	5.15

Process structure is exploited to dynamically create AOIs with projections of behavioral and physiological measures

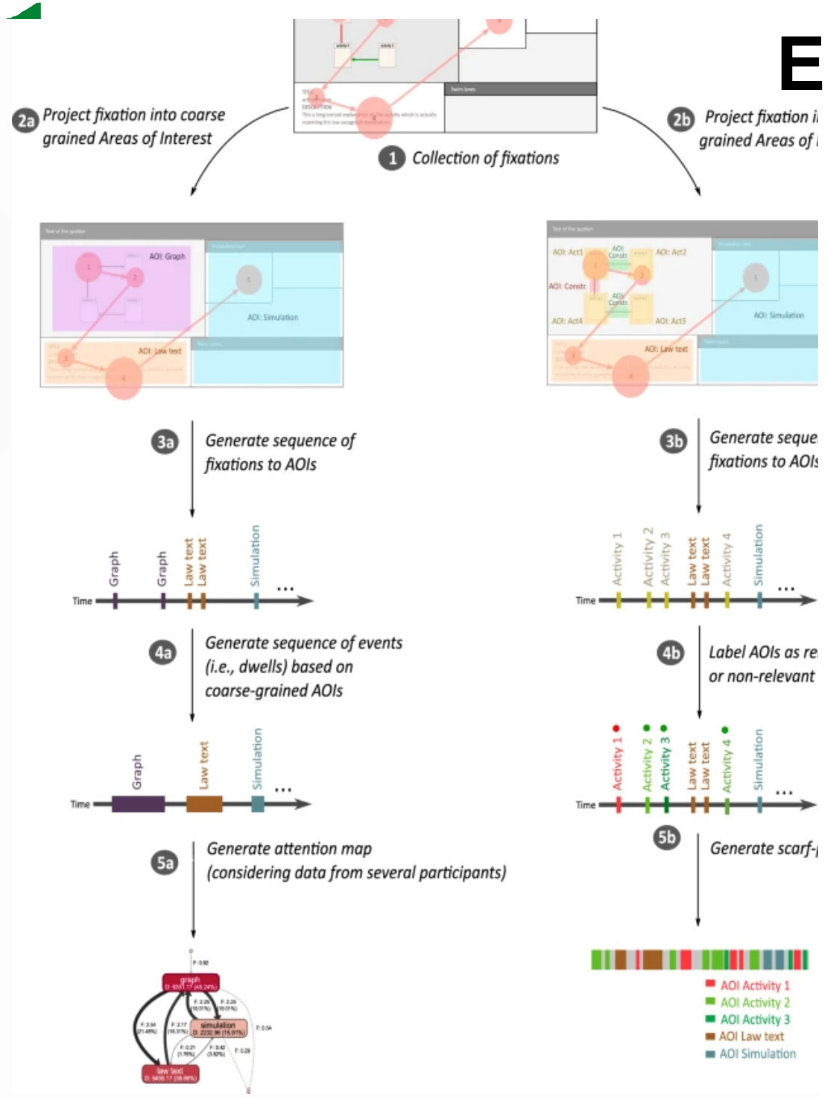
Example: Hybrid Process Models

Stimulus with 3 Areas of Interst



Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Example: Hybrid Process Models



- Screen divided into several static Areas of Interest (model, law text, and simulation)

- Compressed AOI Strings based on dwells

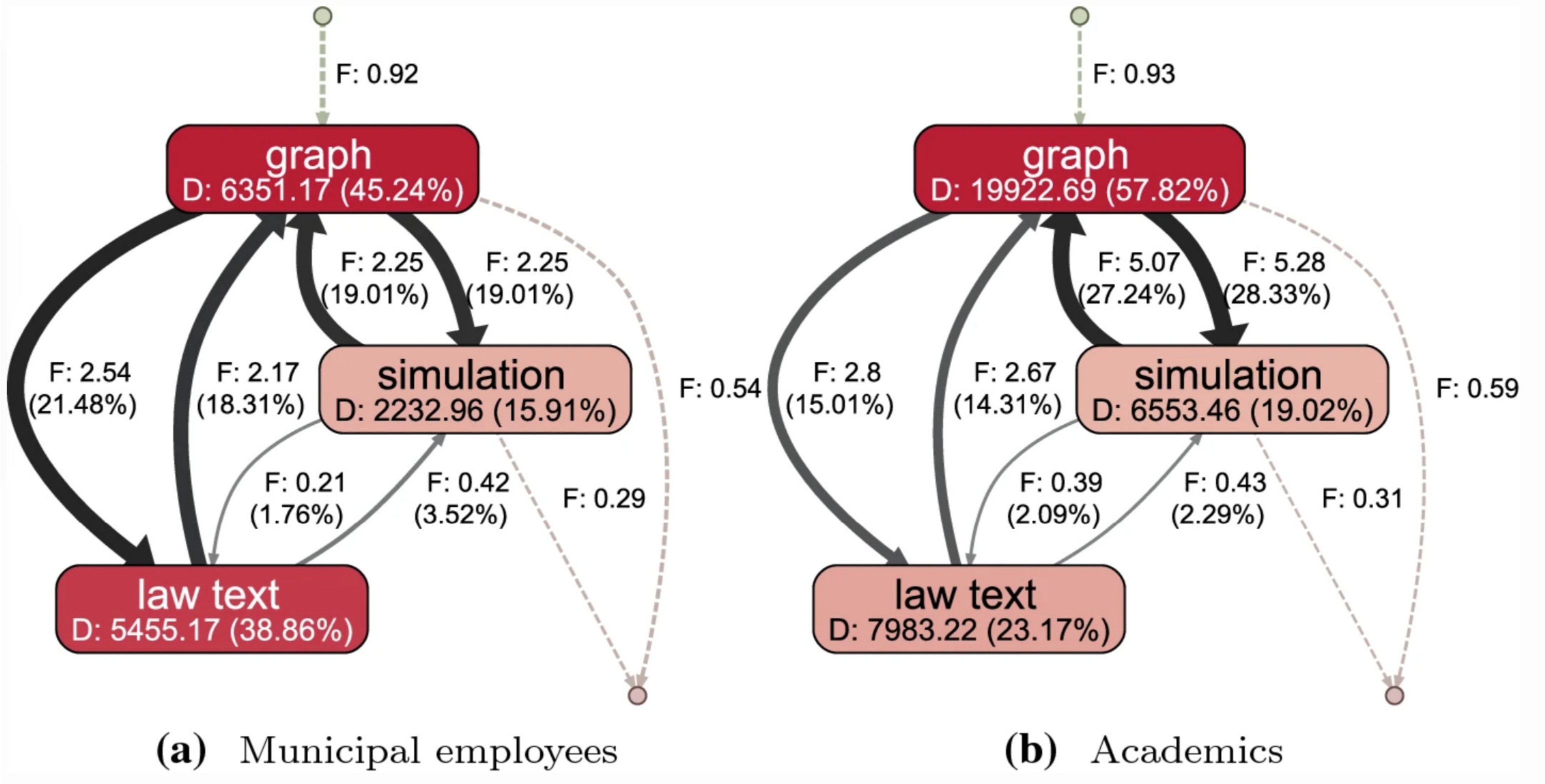
- Visualization of results as
  - Directly-Follow-Graphs (DFGs) using Process Mining
  - Scarf Plots

- Method triangulation: Think aloud

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Example: Hybrid Process Models

Visualization as Directly-Follow-Graph



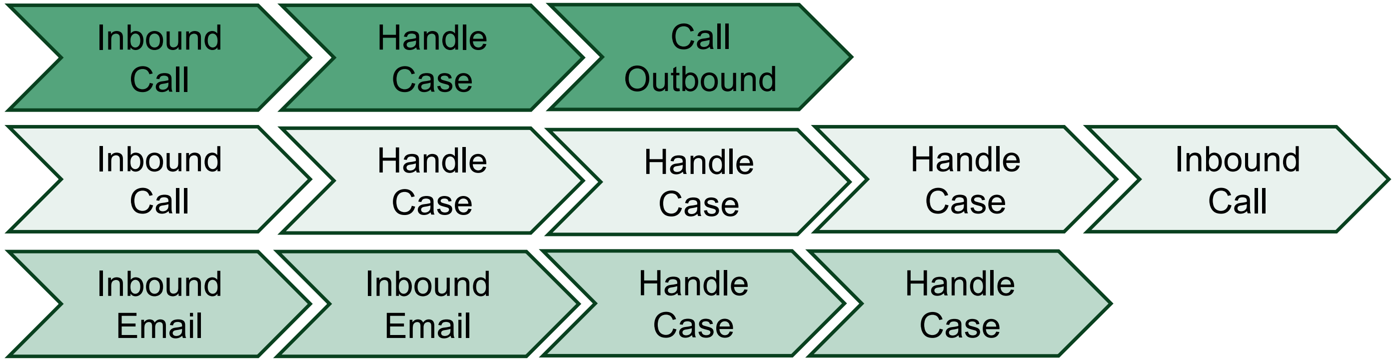
Attention maps in form of Directly-Follow-Graphs comparing the attention focus on different artifacts for municipal employees and academics. *D* is the mean fixation duration, and *F* is the mean transition frequency between two AOIs.

Going Beyond a Single Data Source

Linking Multiple Data Sources

	Start Date	End Date
	04.03.2010 07:35	04.03.2010 07:46
	04.03.2010 07:53	04.03.2010 07:55
	08.03.2010 11:16	08.03.2010 11:18
	09.03.2010 08:05	09.03.2010 08:10
	11.03.2010 10:30	11.03.2010 10:32
	11.03.2010 11:15	11.03.2010 11:19
d	11.03.2010 11:45	11.03.2010 11:52
il	14.03.2010 14:08	18.03.2010 08:04
	14.03.2010 17:53	14.03.2010 17:56
il	18.03.2010 08:06	18.03.2010 08:07
	18.03.2010 08:07	18.03.2010 08:08
	18.03.2010 08:09	18.03.2010 08:09

An event log contains traces
Each trace is a sequence of events



Inbound Call → Handle Case → Call Outbound

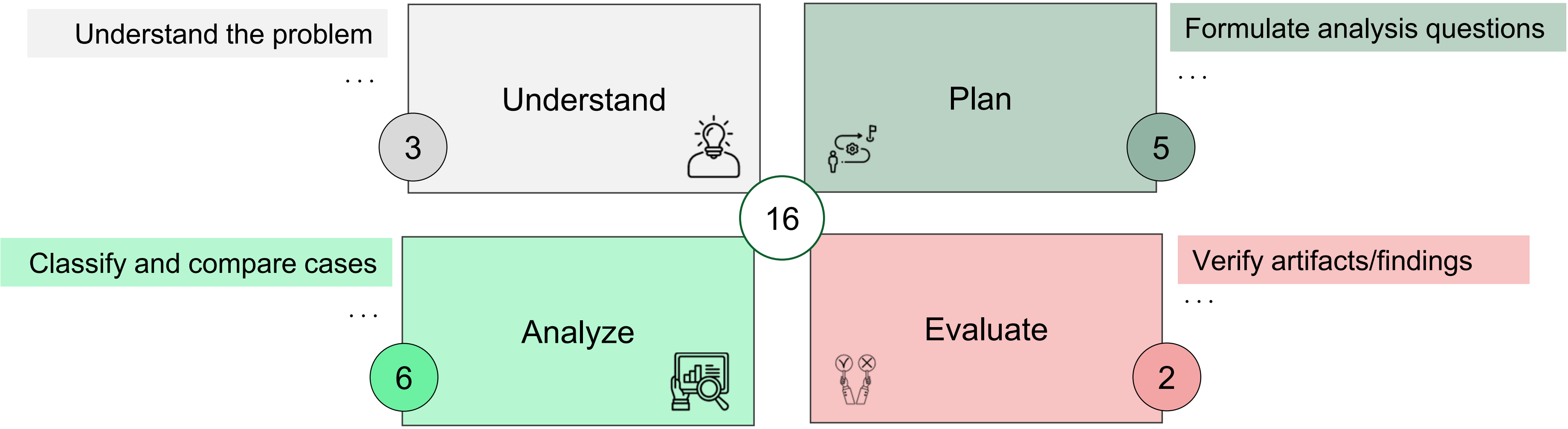
Inbound Call → Handle Case → Handle Case → Handle Case → Inbound Call

Inbound Email → Inbound Email → Handle Case → Handle Case

Process Mining Strategies

What are common strategies used in the analysis stage?

Process mining strategies derived from the analysis of interview data.



F. Zerbato, P. Soffer, B. Weber, Process Mining Practices: Evidence from Interviews. In International Conference on Business Process Management (BPM 22), *Accepted for publication.*

Providing Context to User Interaction Logs

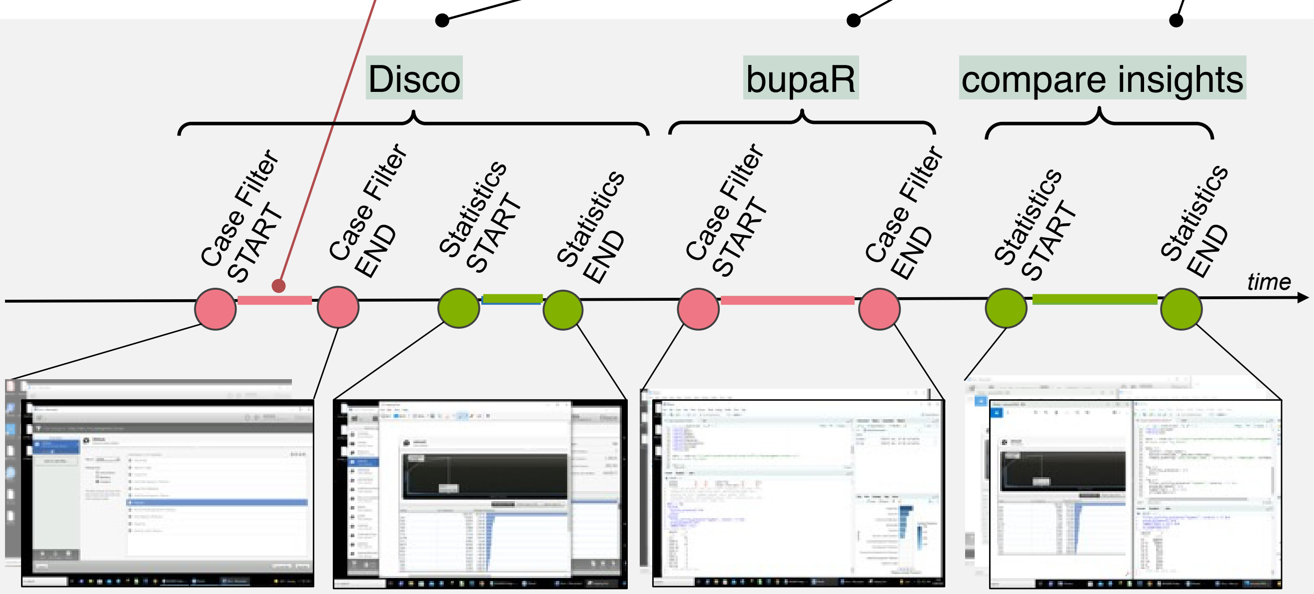
Strategy: Verify artifacts and findings

"Often I try to combine different tools to understand for sanity check if we have the same insights in different tools."

Interview Data

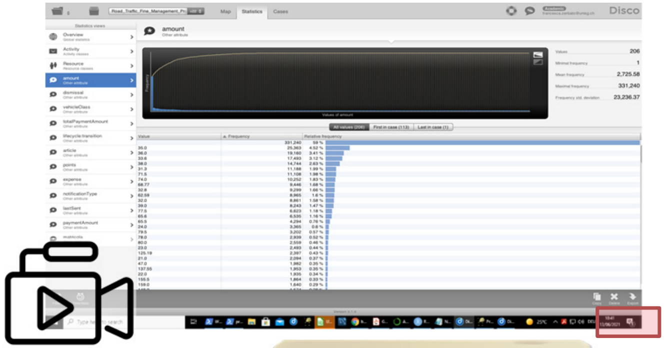
Think-Aloud Data

[00:09:38,3 – 00:11:09,9]
"I'm trying to filter the Payment activity to see all the cases that we don't have a payment. I've tried using the filter forbidden..."



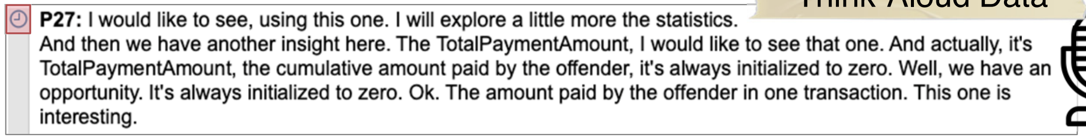
ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:11:46,1	00:12:34,3
P27	Statistics	Disco	00:11:46,1	00:12:34,3
…	…	…	…	…

Creating User Interaction Logs



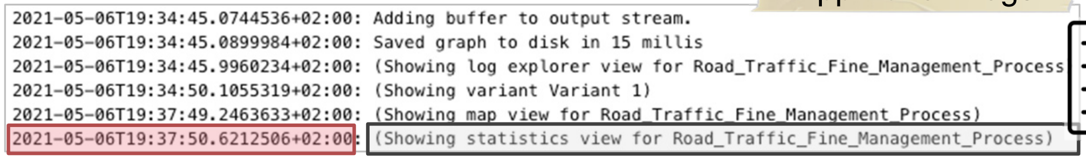
Screen Recording

Think-Aloud Data

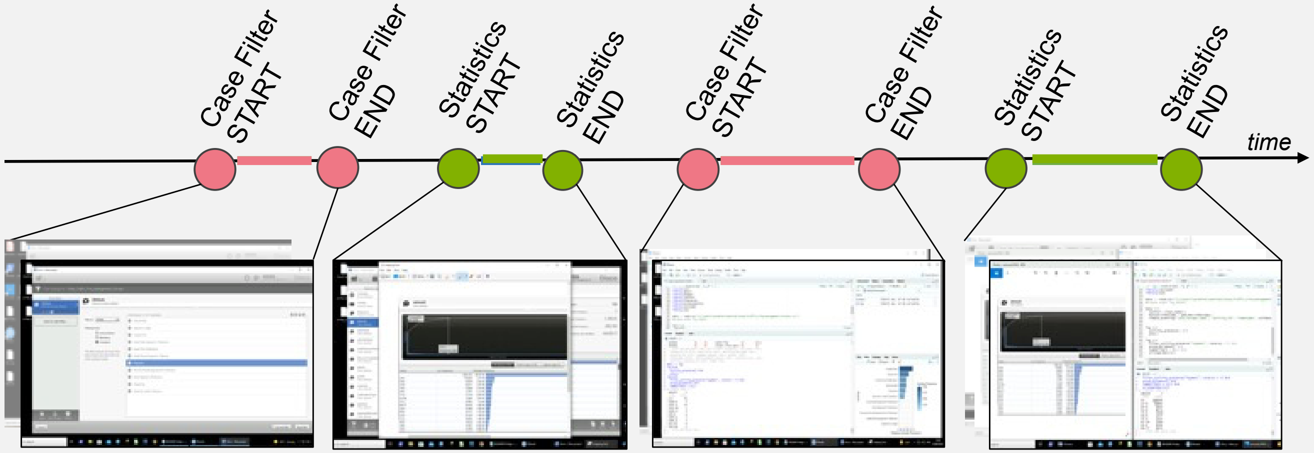


P27: I would like to see, using this one. I will explore a little more the statistics. And then we have another insight here. The TotalPaymentAmount, I would like to see that one. And actually, it's TotalPaymentAmount, the cumulative amount paid by the offender, it's always initialized to zero. Well, we have an opportunity. It's always initialized to zero. Ok. The amount paid by the offender in one transaction. This one is interesting.

Application Logs



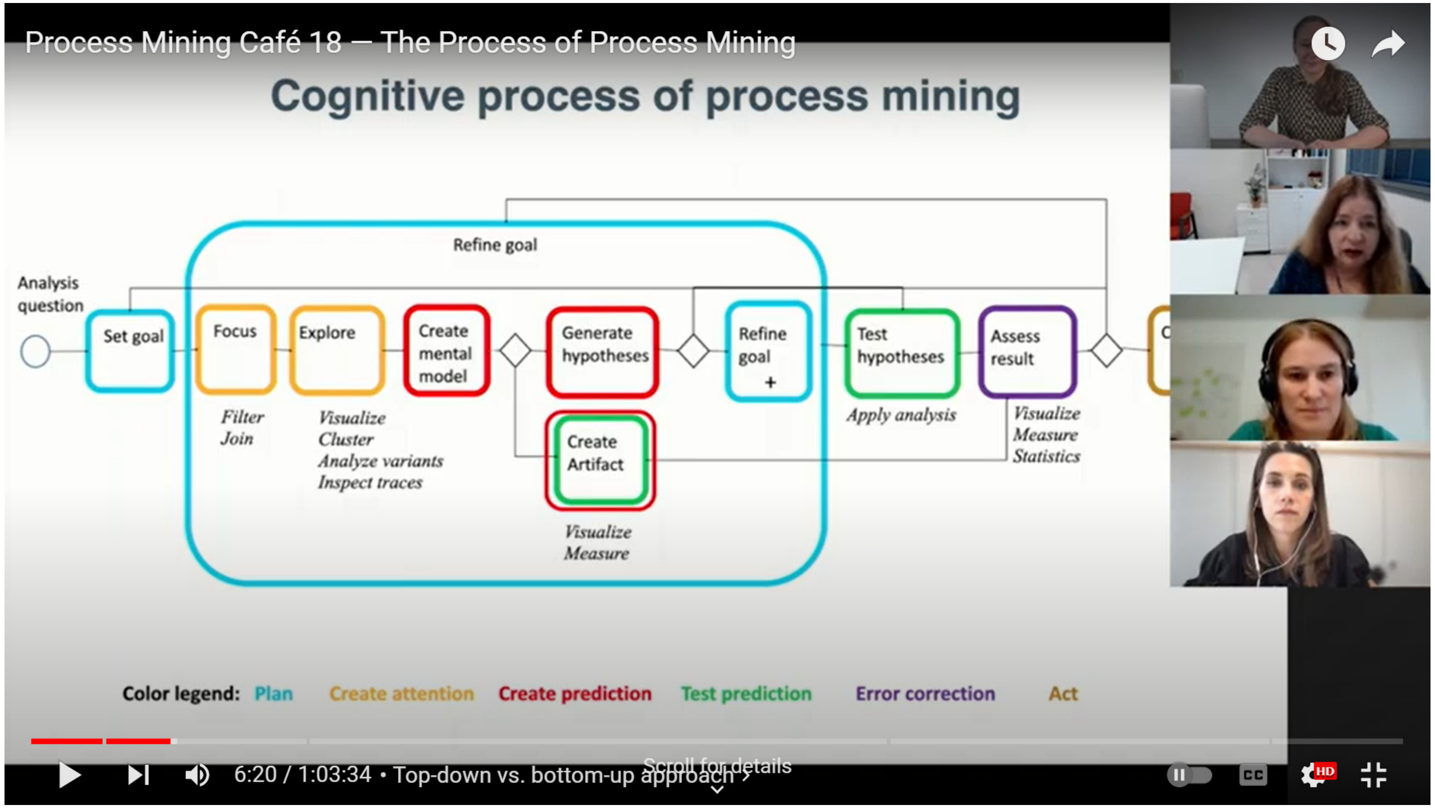
```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```



ID	Tool Function	Tool	Start	End
P27	PDF Reader	Acrobat Reader	00:04:50,3	00:06:17,4
…	…		…	…
P27	Case Filter	Disco	00:09:38,3	00:11:09,9
P27	Statistics	Disco	00:11:46,1	00:12:34,3
P27	Case Filter	bupaR	00:14:00,7	00:15:09,9
P27	Statistics	bupaR	00:11:46,1	00:12:34,3
P27	Statistics	Disco	00:11:46,1	00:12:34,3
…	…	…	…	…

Universität St.Gallen

Exploration and Hypotheses Testing are Core Components of Process Mining

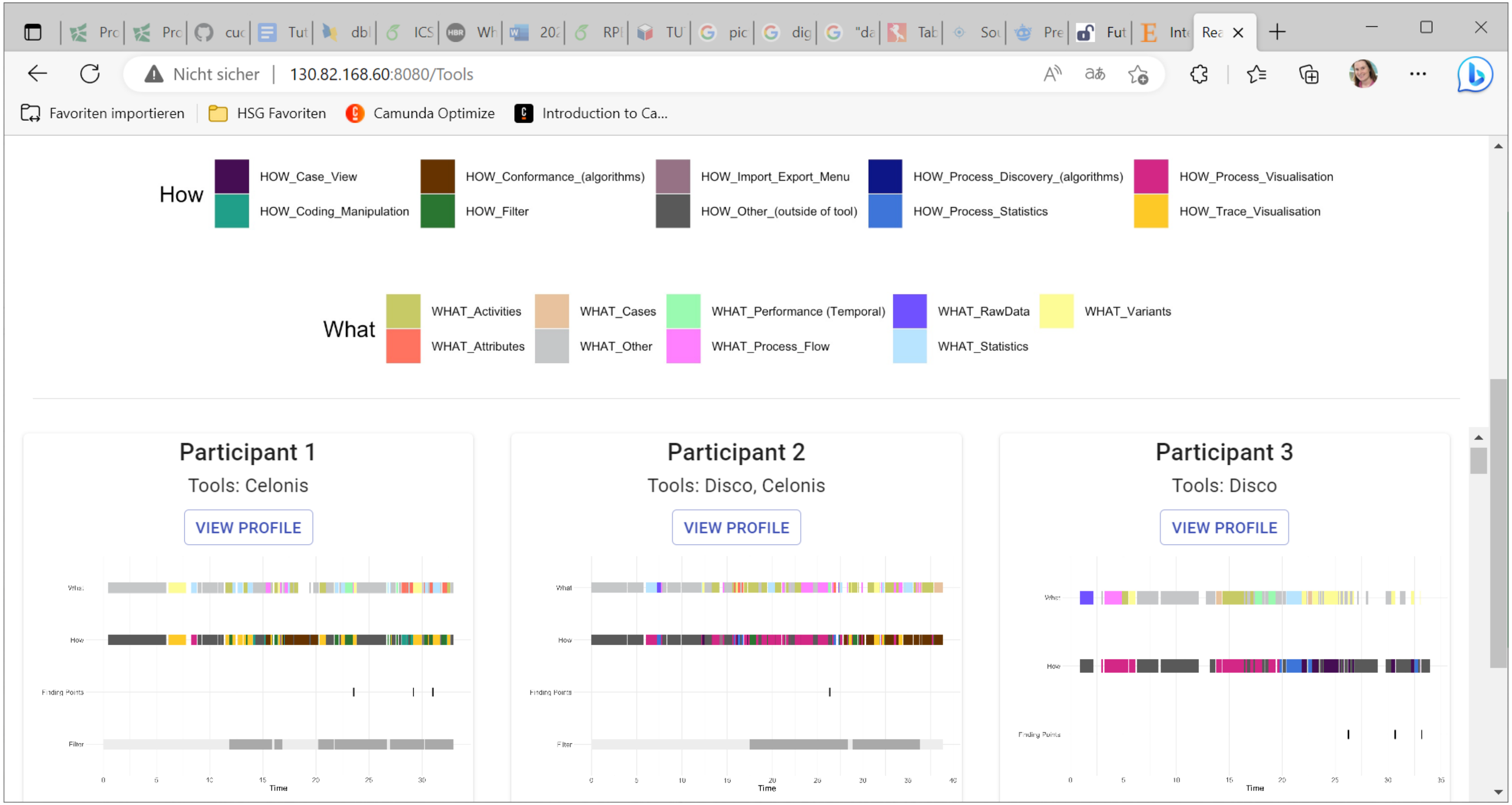


Process Mining Café 18 — The Process of Process Mining

Cognitive process of process mining

Work by Pnina Soffer and Her Team

Creation of Participant Profiles



How

- HOW_Case_View
- HOW_Coding_Manipulation
- HOW_Conformance_(algorithms)
- HOW_Filter
- HOW_Import_Export_Menu
- HOW_Other_(outside of tool)
- HOW_Process_Discovery_(algorithms)
- HOW_Process_Statistics
- HOW_Process_Visualisation
- HOW_Trace_Visualisation

What

- WHAT_Activities
- WHAT_Attributes
- WHAT_Cases
- WHAT_Other
- WHAT_Performance (Temporal)
- WHAT_Process_Flow
- WHAT_RawData
- WHAT_Statistics
- WHAT_Variants

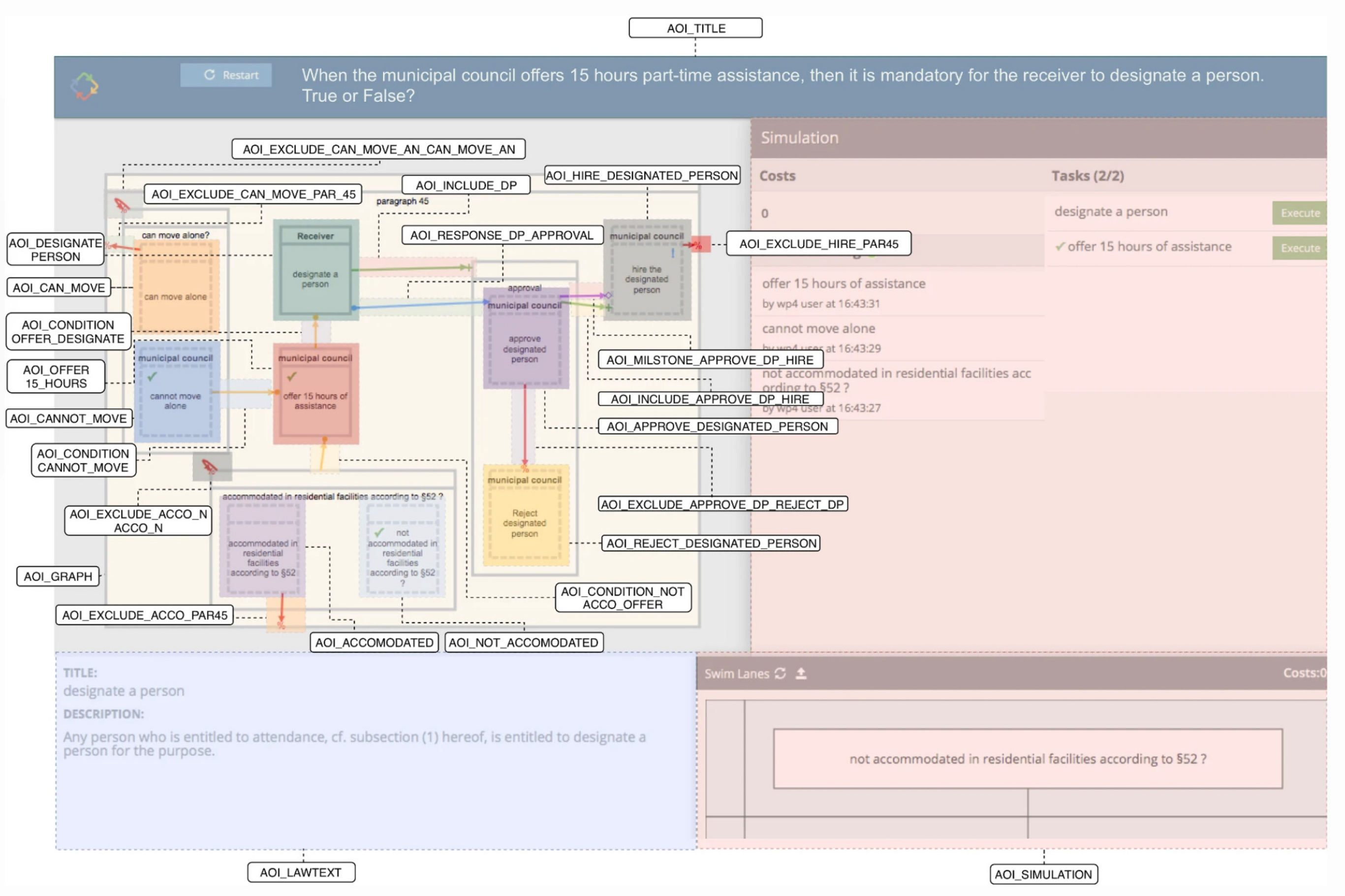
Participant 1
Tools: Celonis
VIEW PROFILE

Participant 2
Tools: Disco, Celonis
VIEW PROFILE

Participant 3
Tools: Disco
VIEW PROFILE

Example: Hybrid Process Models

Fine-grained Areas of Interst


When the municipal council offers 15 hours part-time assistance, then it is mandatory for the receiver to designate a person. True or False?

TITLE:
designate a person

DESCRIPTION:
Any person who is entitled to attendance, cf. subsection (1) hereof, is entitled to designate a person for the purpose.

offer 15 hours of assistance
by wp4 user at 16:43:31

cannot move alone
by wp4 user at 16:43:29

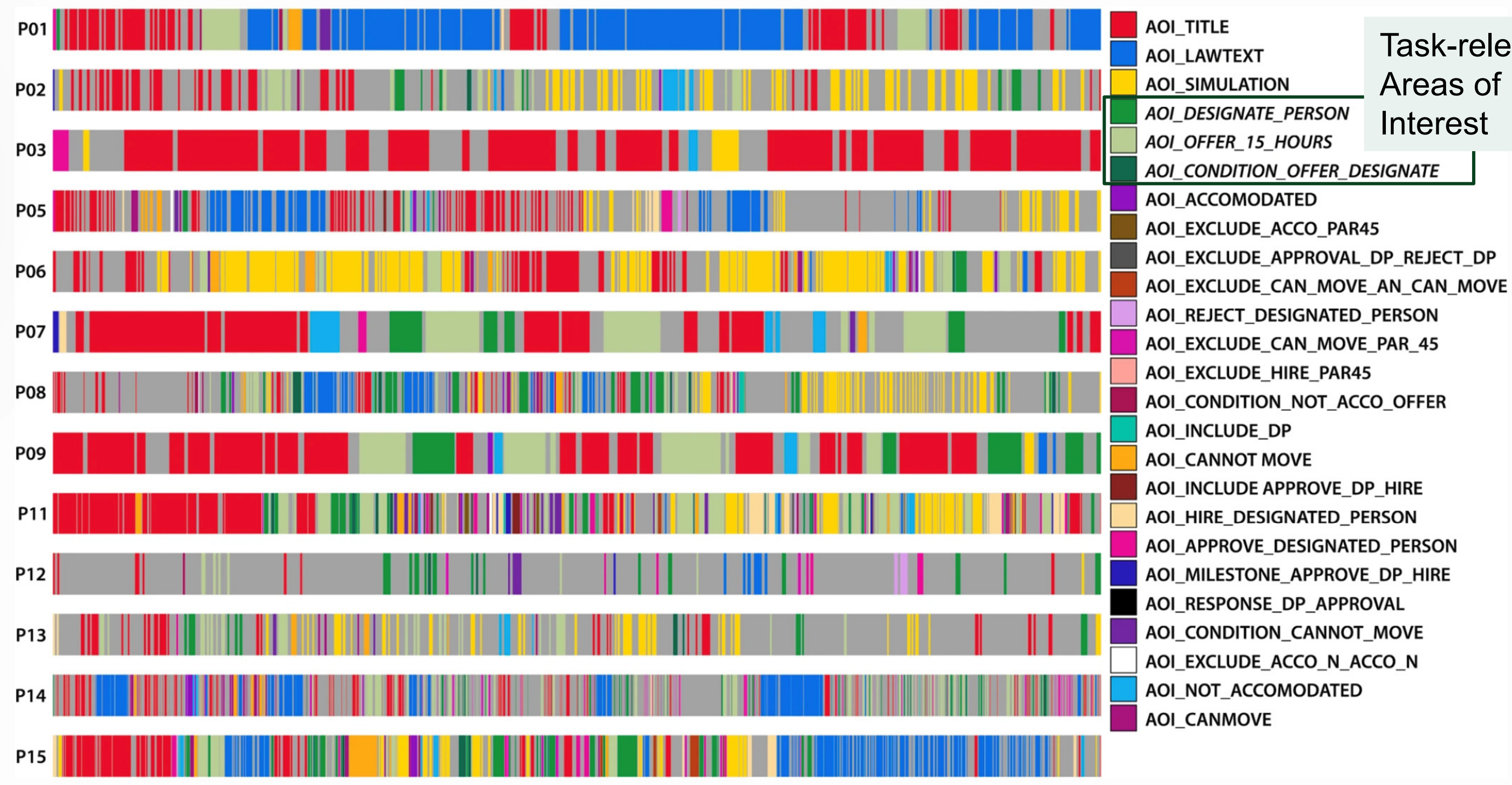
not accommodated in residential facilities according to §52 ?
by wp4 user at 16:43:27

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Universität St.Gallen

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.



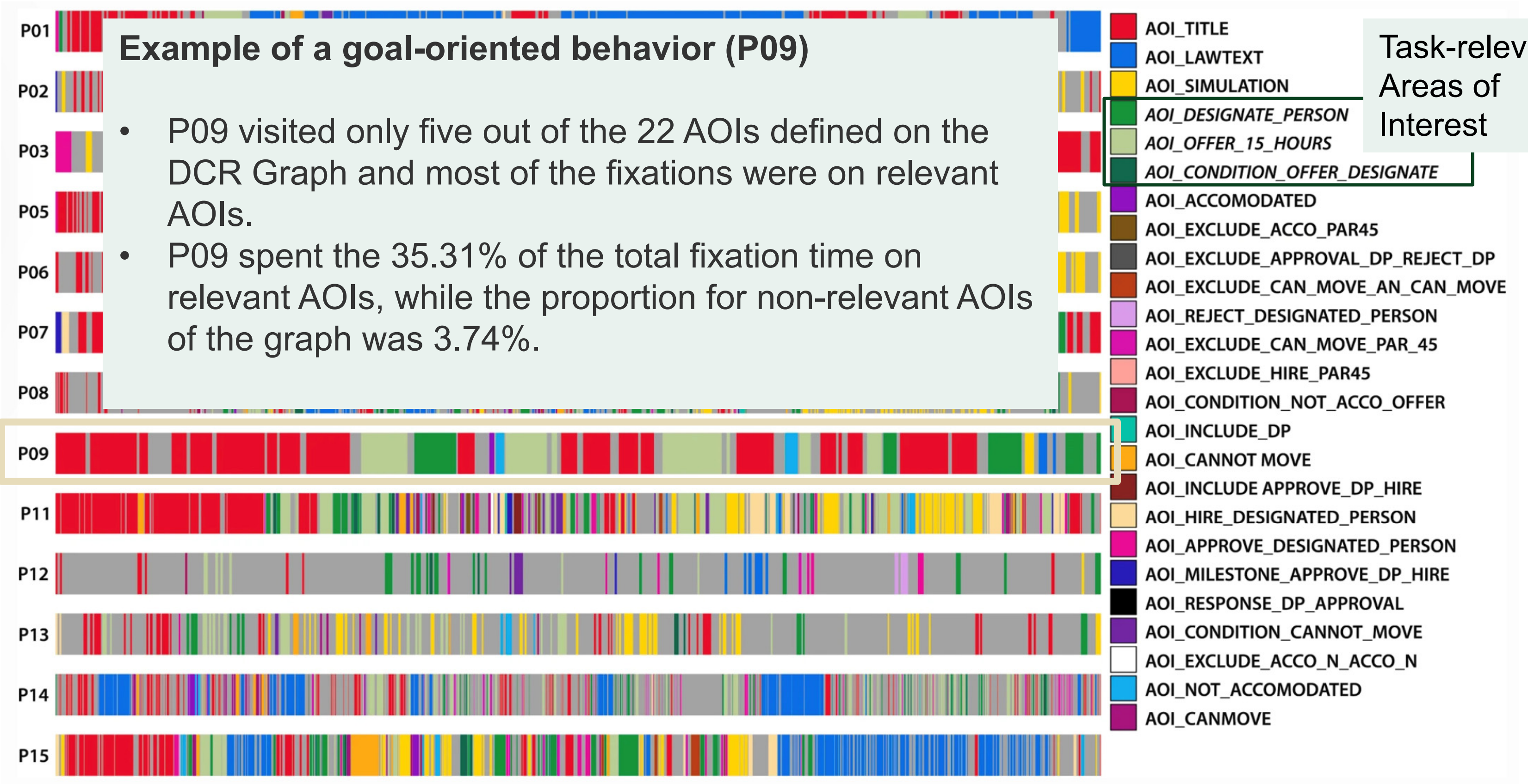
Task-relevant Areas of Interest

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Universität St.Gallen

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.



Example of a goal-oriented behavior (P09)

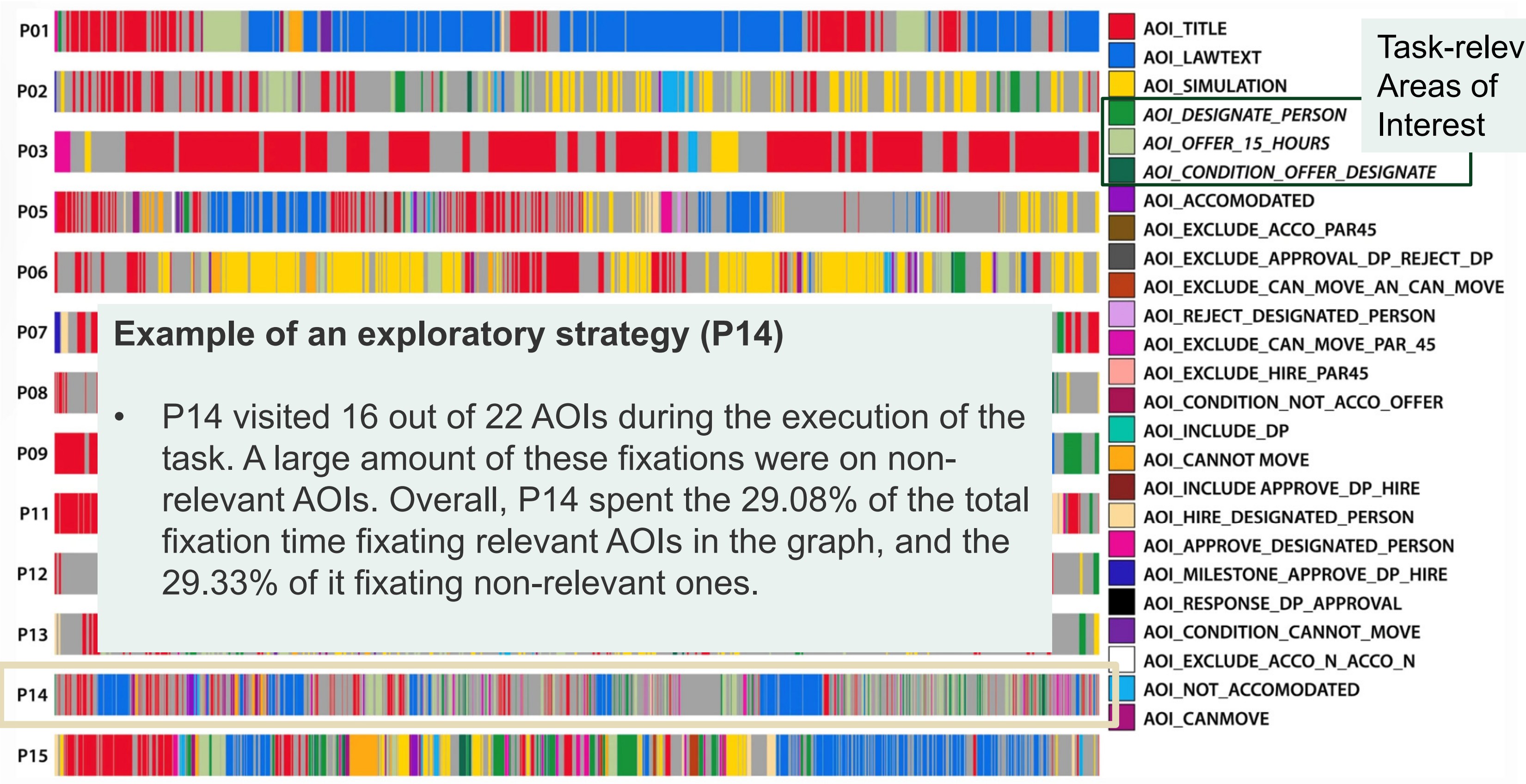
- P09 visited only five out of the 22 AOIs defined on the DCR Graph and most of the fixations were on relevant AOIs.
- P09 spent the 35.31% of the total fixation time on relevant AOIs, while the proportion for non-relevant AOIs of the graph was 3.74%.

Task-relevant Areas of Interest

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Example: Hybrid Process Models

Scarf-plot showing the sequences of fixations for participants solving a constraint task. Relevant AOIs of the DCR Graph for this task are labeled in italic.



Task-relevant Areas of Interest

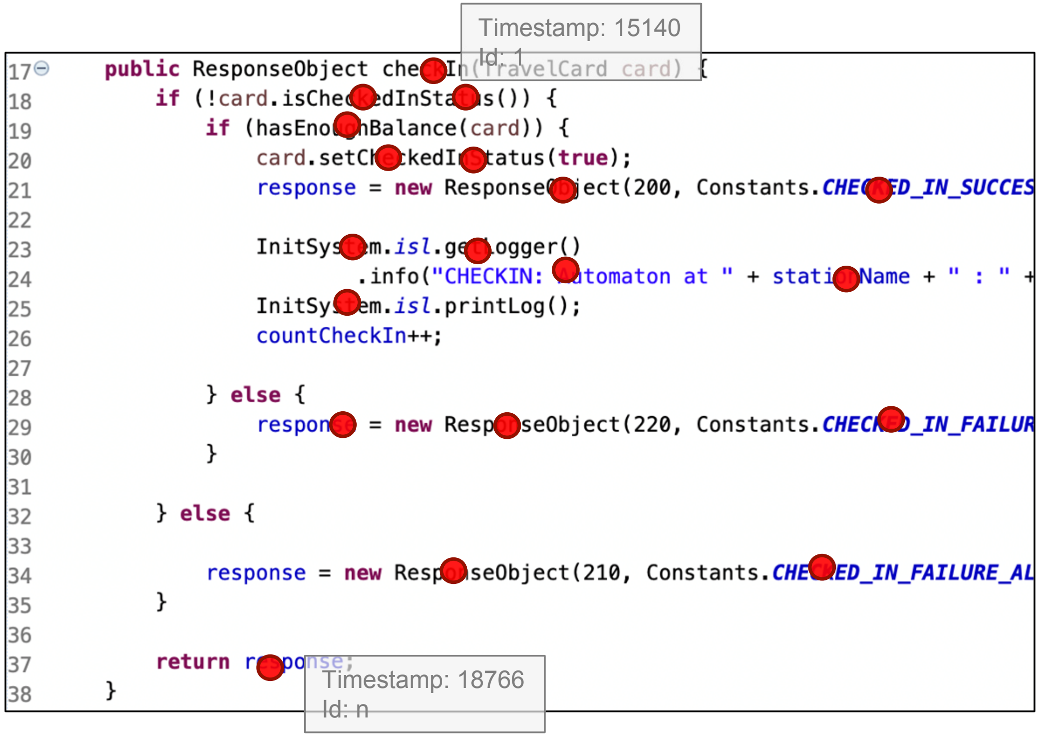
Example of an exploratory strategy (P14)

- P14 visited 16 out of 22 AOIs during the execution of the task. A large amount of these fixations were on non-relevant AOIs. Overall, P14 spent the 29.08% of the total fixation time fixating relevant AOIs in the graph, and the 29.33% of it fixating non-relevant ones.

Source: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud

Automated Mapping to Areas of Interest: The Case of Source-code

Eye-tracker



```
17    public ResponseObject checkIn(TravelCard card) {
18        if (!card.isCheckedInStatus()) {
19            if (hasEnoughBalance(card)) {
20                card.setCheckedInStatus(true);
21                response = new ResponseObject(200, Constants.CHECKED_IN_SUCCES
22
23                InitSystem.isl.getLogger()
24                      .info("CHECKIN: Automaton at " + stationName + " : " +
25                InitSystem.isl.printLog();
26                countCheckIn++;
27
28            } else {
29                response = new ResponseObject(220, Constants.CHECKED_IN_FAILUR
30            }
31
32        } else {
33
34            response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_AL
35        }
36
37        return response;
38    }
```

Sample of user’s gazes on the source-code

Classical approach of assigning AOIs manually

```
17    public ResponseObject checkIn(TravelCard card) {
18        if (!card.isCheckedInStatus()) {
19            if (hasEnoughBalance(card)) {
20                card.setCheckedInStatus(true);
21                response = new ResponseObject(200, Constants.CHECKED_IN_SUCCES
22
23                InitSystem.isl.getLogger()
24                      .info("CHECKIN: Automaton at " + stationName + " : " +
25                InitSystem.isl.printLog();
26                countCheckIn++;
27
28            } else {
29                response = new ResponseObject(220, Constants.CHECKED_IN_FAILUR
30            }
31
32        } else {
33
34            response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_AL
35        }
36
37        return response;
38    }
```

Each AOI refers to a distinct line of code

Automated Mapping to Areas of Interest: The Case of Source-code

Universität St.Gallen

What happens when the user scrolls down in the source-code editor?

```
17  public ResponseObject checkIn(TravelCard card) {
18      if (!card.isCheckedInStatus()) {
19          if (hasEnoughBalance(card)) {
20              card.setCheckedInStatus(true);
21              response = new ResponseObject(200, Constants.CHECKED_IN_SUCCES
22
23              InitSystem.isl.getLogger()
24                      .info("CHECKIN: Automaton at " + stationName + " : " +
25              InitSystem.isl.printLog();
26              countCheckIn++;
27
28          } else {
29              response = new ResponseObject(220, Constants.CHECKED_IN_FAILUR
30          }
31
32      } else {
33
34          response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_AL
35      }
36
37      return response;
38  }
```

Each AOI refers to a distinct line of code

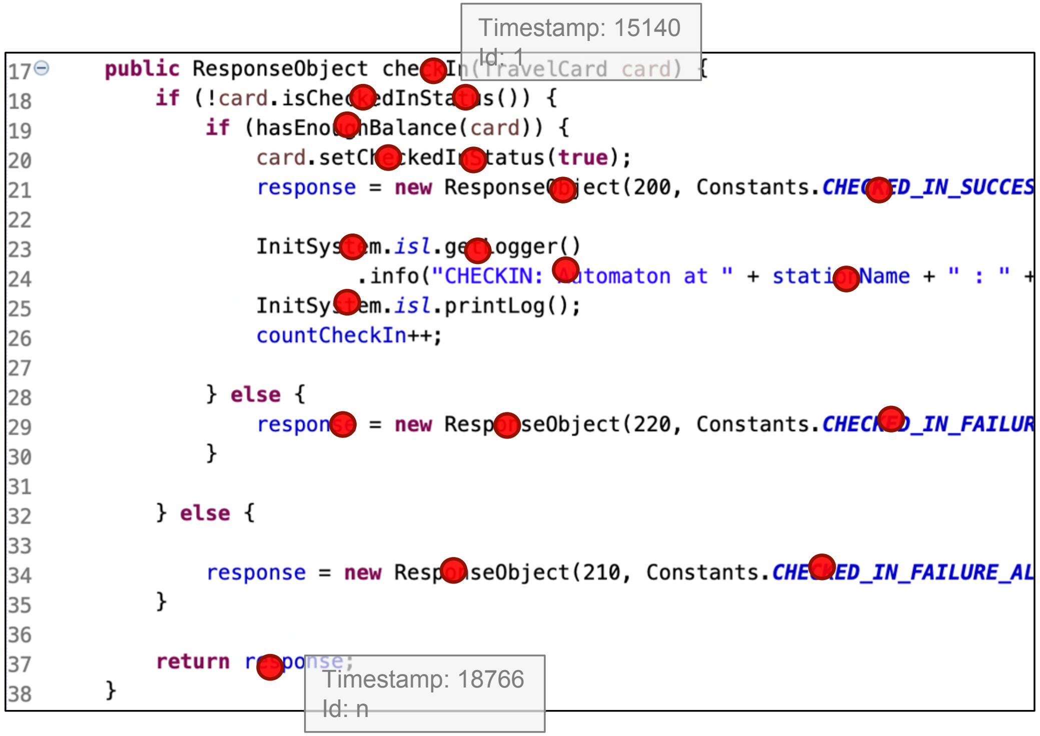
After scrolling

```
23              InitSystem.isl.getLogger()
24                      .info("CHECKIN: Automaton at " + stationName + " : " + Consta
25              InitSystem.isl.printLog();
26              countCheckIn++;
27
28          } else {
29              response = new ResponseObject(220, Constants.CHECKED_IN_FAILURE_LOW_B
30          }
31
32      } else {
33
34          response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_ALREADY)
35      }
36
37      return response;
38  }
39
40
41  public void checkInLog() {
42      InitSystem.isl.logContains(Constants.CHECKED_IN_SUCCESS);
43  }
44
45
46  public String getStationName() {
47      return stationName;
48  }
```

- Inconsistant mapping to AOIs
- Need to re-assign AOIS
- Time consuming (need to be done at every scroll event)

Automated Mapping to Areas of Interest: The Case of Source-code

Eye-tracker



```
17   public ResponseObject checkIn(TravelCard card) {
18       if (!card.isCheckedInStatus()) {
19           if (hasEnoughBalance(card)) {
20               card.setCheckedInStatus(true);
21               response = new ResponseObject(200, Constants.CHECKED_IN_SUCCES
22 
23               InitSystem.isl.getLogger()
24                       .info("CHECKIN: Automaton at " + stationName + " : " +
25               InitSystem.isl.printLog();
26               countCheckIn++;
27 
28           } else {
29               response = new ResponseObject(220, Constants.CHECKED_IN_FAILUR
30           }
31 
32       } else {
33 
34           response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_AL
35       }
36 
37       return response;
38   }
```

Timestamp: 15140
Id: 1

Timestamp: 18766
Id: n

Sample of user’s gazes on the source-code

iTrace

Automated mapping between gazes and areas of interest at the data collection

Gaze ID	Timestamp	Gaze X Pos.	Gaze Y Pos.	Line of code	Column of code
1	15140	1450	60	17	30
2	15148	1430	77	18	26
3	15156	1480	77	18	32
4	15164	1530	86	19	25
..	..	..	..	..	..

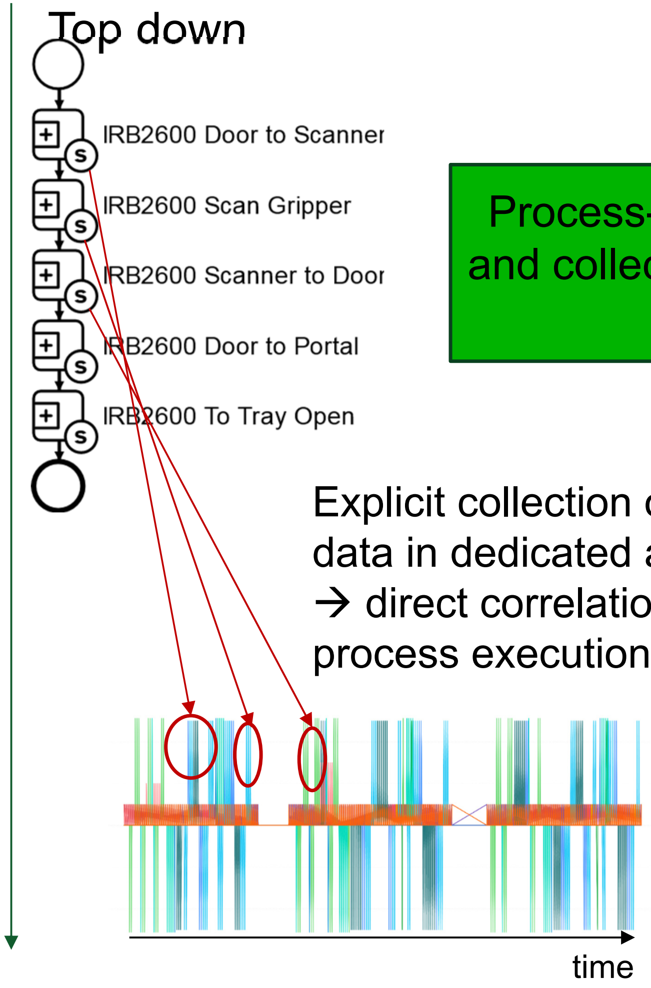
Sample of gaze file with automatically mapped AOI

Source: https://www.i-trace.org

Process Mining

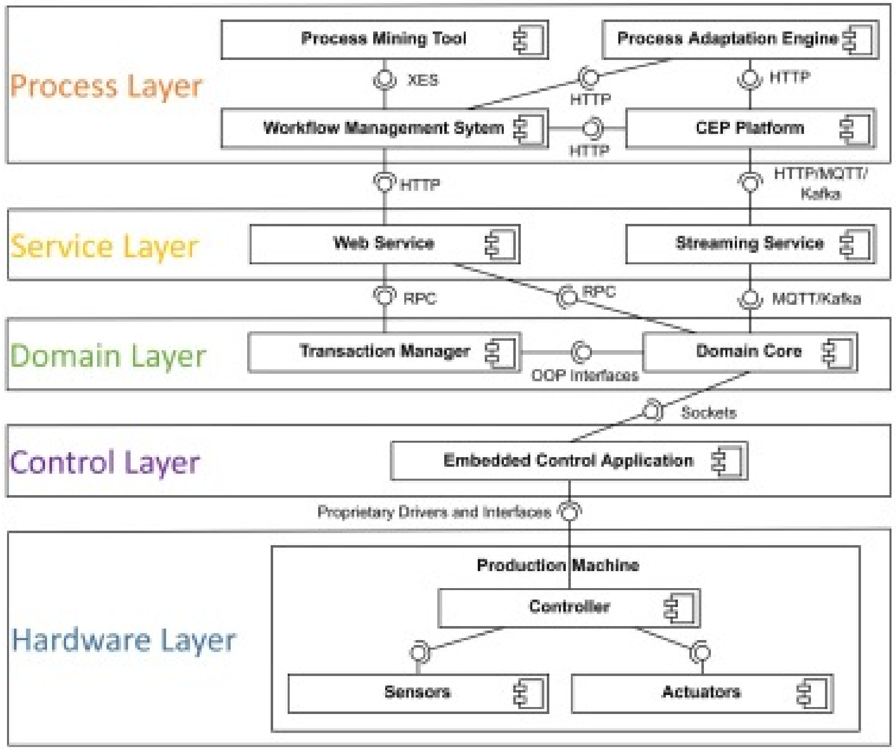
- Process Mining – a success story – from research to a multi billion market
- Gartner Quadrant
- Explanation of what it is

Process-driven Execution and Collection of IoT Data in Context

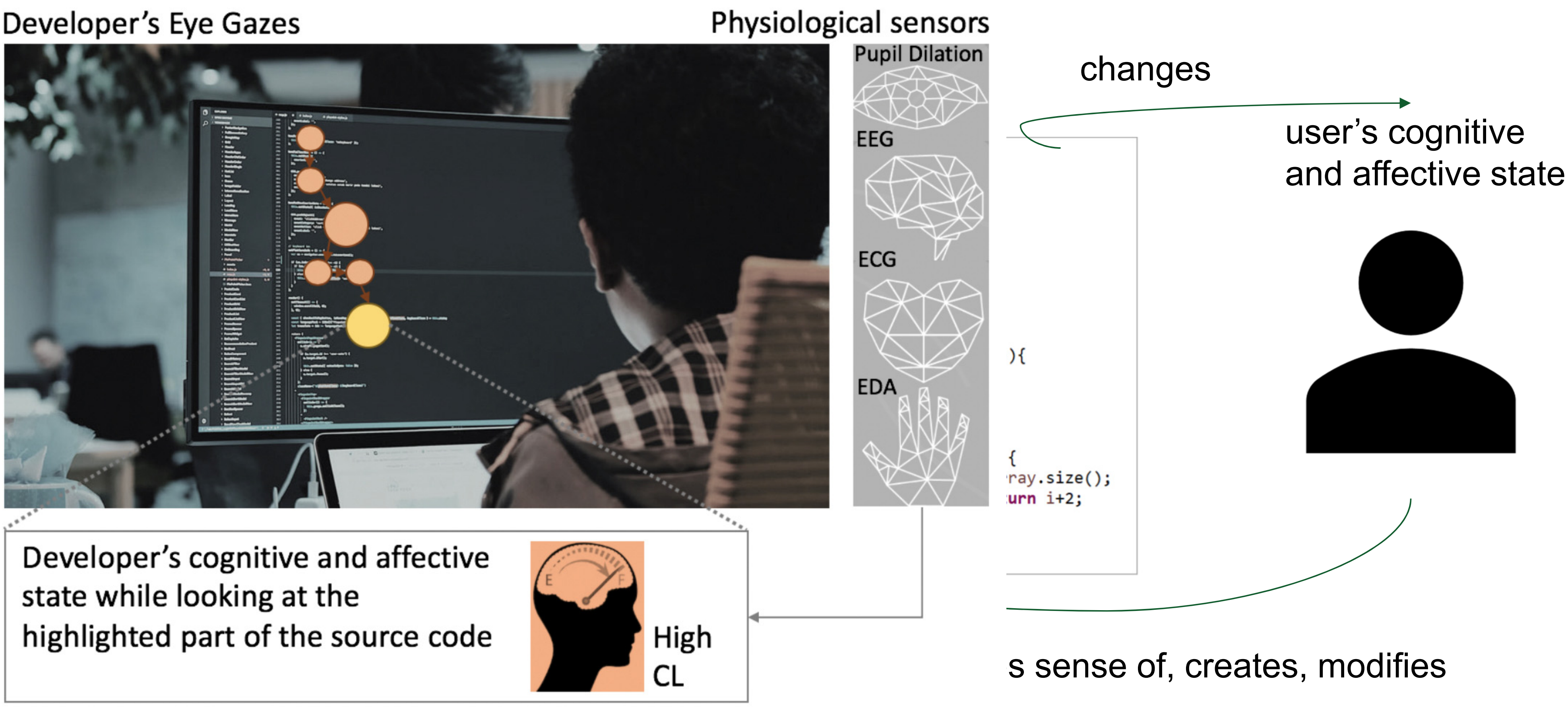


Process-driven execution and collection of IoT data in context

Explicit collection of sensor data in dedicated activities → direct correlation with process execution



The Process of Reading Source Code



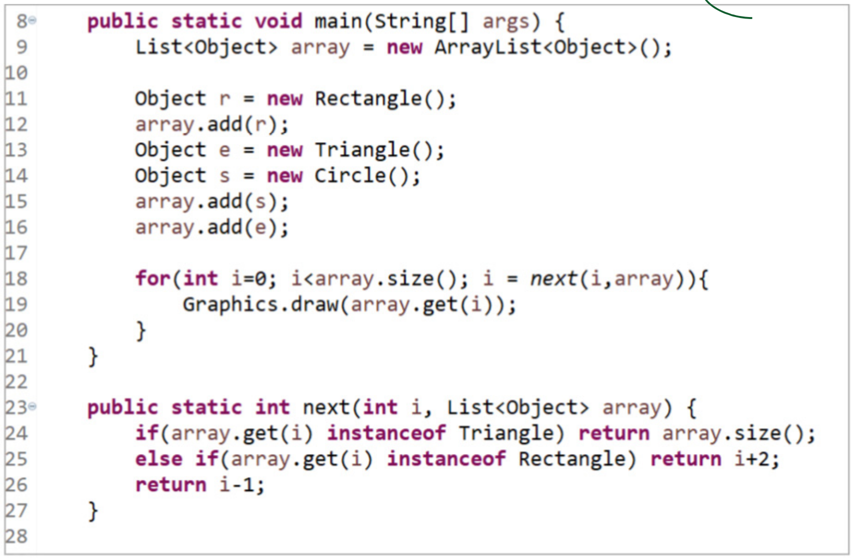
changes

user's cognitive and affective state

s sense of, creates, modifies

The Process of Reading Source Code

Software Design Artifacts



```
    public static void main(String[] args) {
        List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object> array) {
        if(array.get(i) instanceof Triangle) return array.size();
        else if(array.get(i) instanceof Rectangle) return i+2;
        return i-1;
    }

```

changes

user's cognitive
and affective state

measure

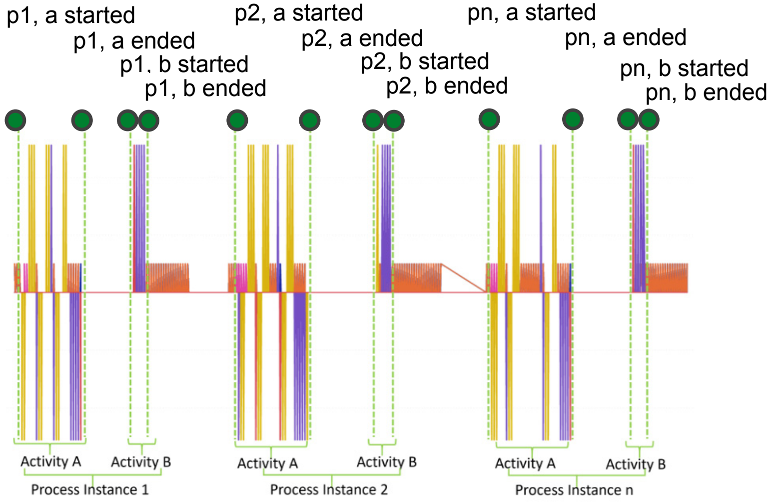
Biosensors

Reads, makes sense of, creates, modifies

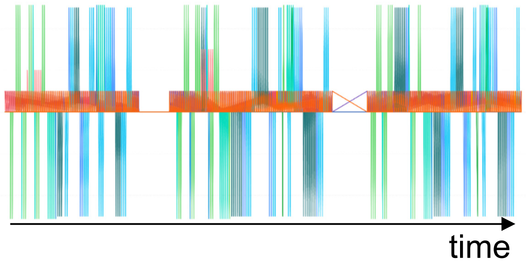
Process Activity Detection from Sensors

Process event sequences

Process Event Log ≡ **Process Event Stream**



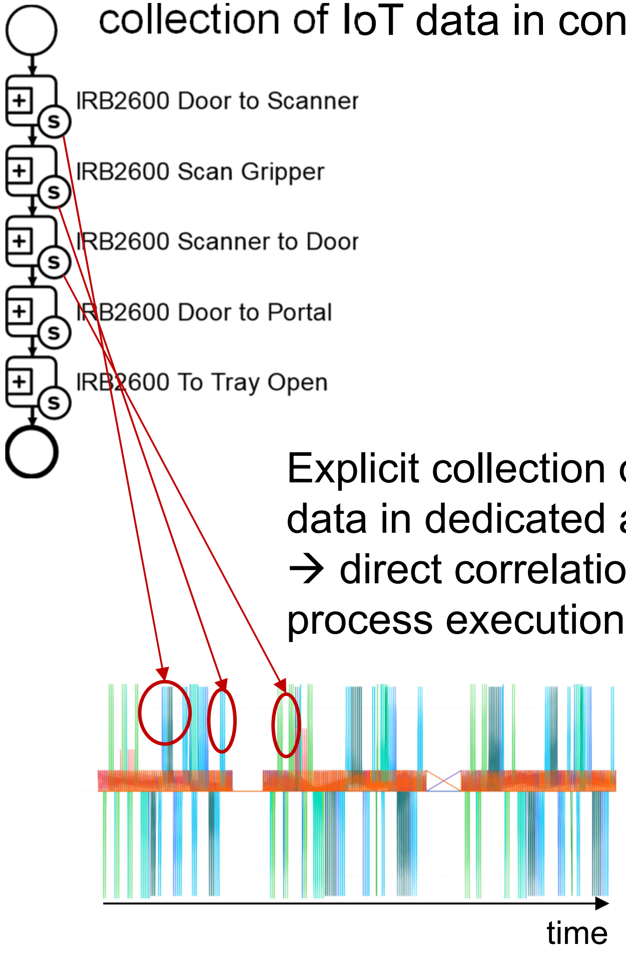
Event sequences of all sensors and actuators of smart factory



Bottom up

Top down

Process-driven execution and collection of IoT data in context



Explicit collection of sensor data in dedicated activities → direct correlation with process execution

Investigating and Supporting Human Work Proceeses

Our Goal

Using a process science lense

And digital trace data

Correlation

- Correlation
  - – Identifying my instance – what is my CaseID
    - A session is one process instance; a task is one instance
    - Each event belongs to one instance

Easy for one modality

For multiple modalities we need synchronization of event sequences

- Obejct-centric process mining; Linking events to smaller level elements
  - Relate my events to smaller level objects, e.g., a source code token

Abstraction

- What is my event?
- Granularity (fixed granularity problem)

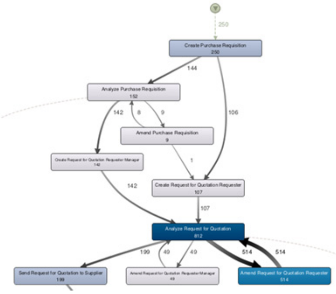
- When I change my granularity what is the impact (CAiSE paper)

Ambiguity

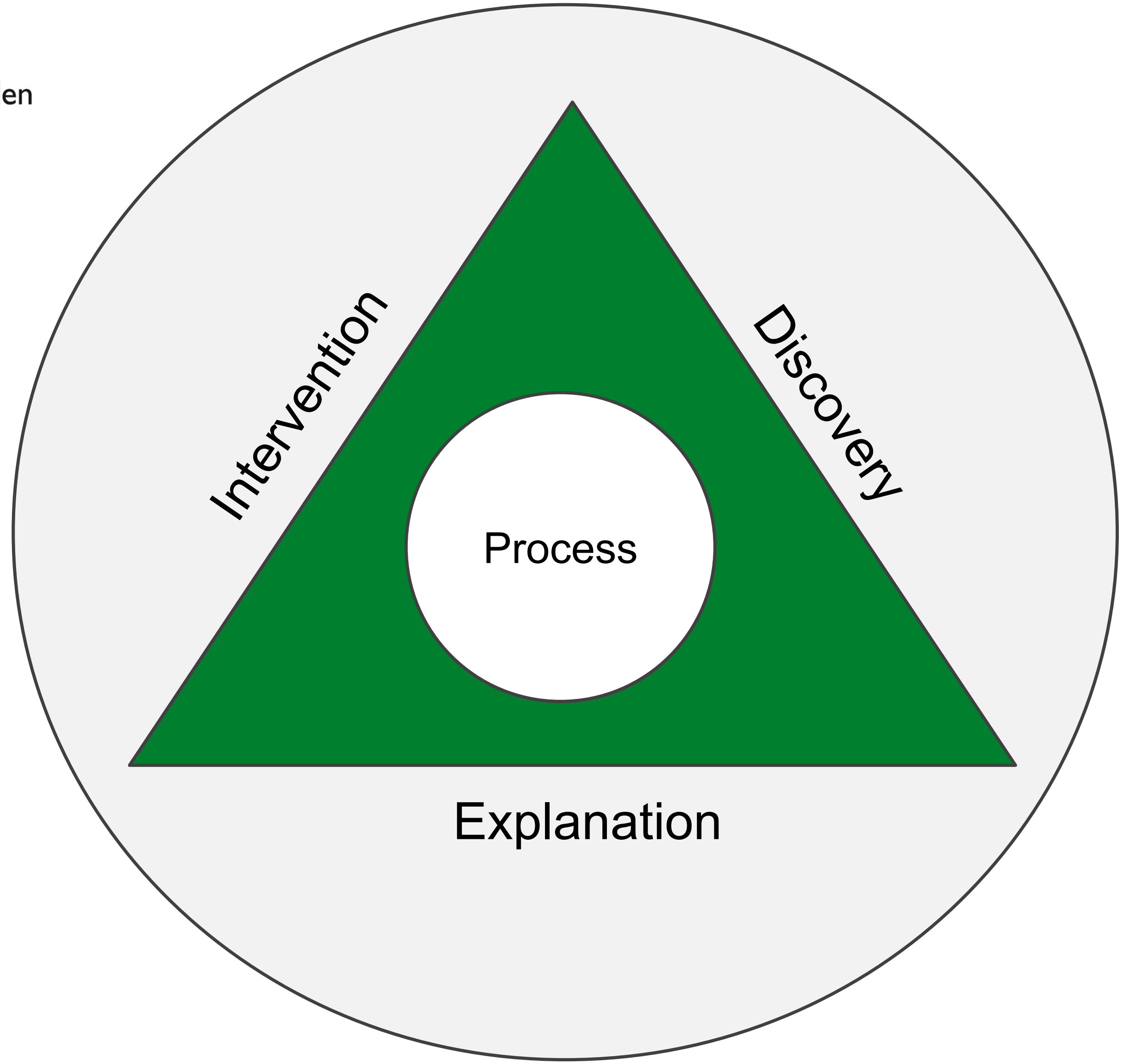
Unstructuredness

Selected Examples

Universität St.Gallen



Process of Process Mining

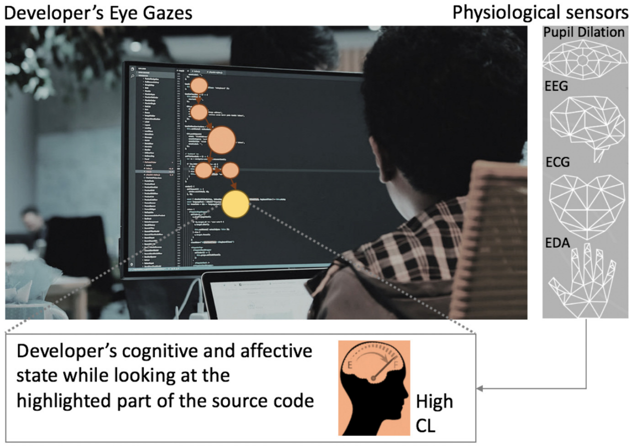


Cognitive Processes

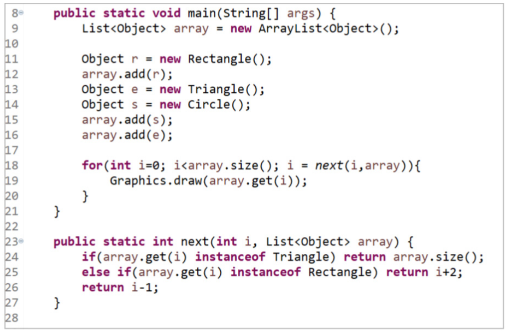
IoT-enabled processes

1. Unravelling Cognitive Processes

2. Tracing developers physiological and behavioural patterns to explain software deficiencies and intervene in fault-prone situations

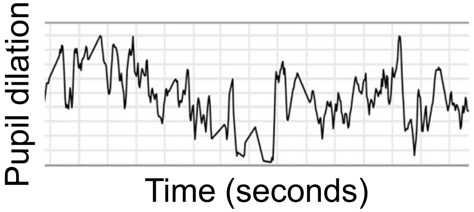


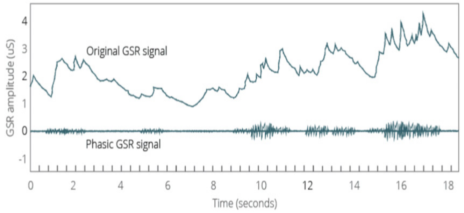
Software Artifact



```
    public static void main(String[] args) {
        List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object> array) {
        if(array.get(i) instanceof Triangle) return array.size();
        else if(array.get(i) instanceof Rectangle) return i+2;
        return i-1;
    }
```

Biosensors





Universität St.Gallen

Observability Through Sensors for Collecting Behavioral and Physiological Data

- Eye-tracking devices
  - – Collection of eye-related measures such as pupillary response data, eye blinks, gaze data
- Galvanic Skin Response sensor
  - – Collection of electro-dermal activity

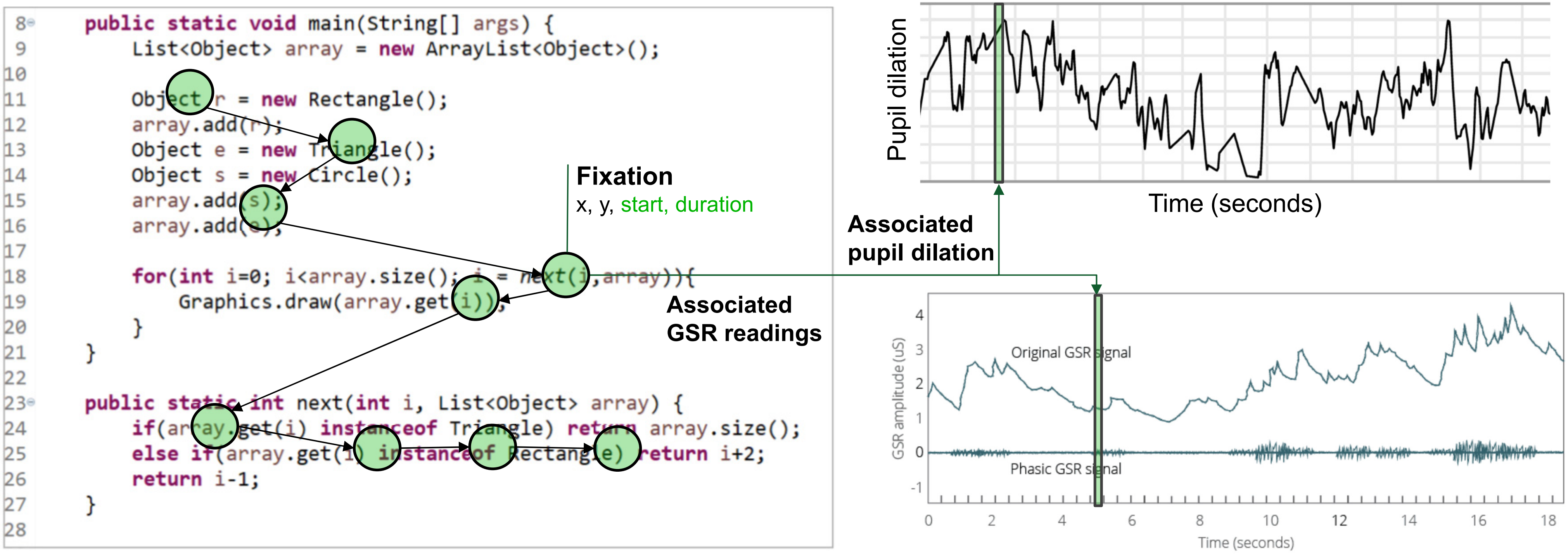
- Electrocardiogram
  - – Collects the electrical signal from the heart and allows measure heart-rate variability
- Heart rate monitor
  - – Collection of heart-related measures such as heart rate

Event Sequence Data

Potentially Multi-modal Data

Temporal Association of Events

Challenge: Synchronization of events



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Fixation
x, y, start, duration

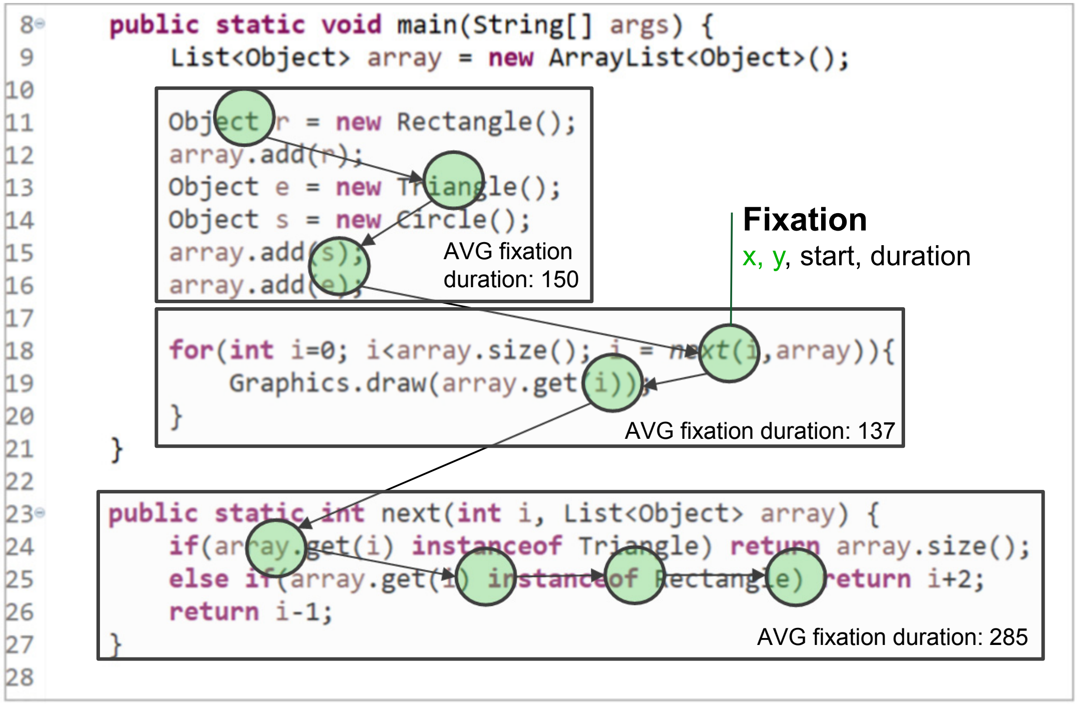
Associated GSR readings

Associated pupil dilation

Sequence of fixations over time

Challenge: Signals differ in terms of latency (time between stimulus and reaction)

Spatial Association of Software Design Artifacts and Events



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Software Design Artifact with Areas of Interest

Areas of Interest (AOIs)

Areas of interest allow for a **spatial grouping and aggregation** of fixations using their spatial properties

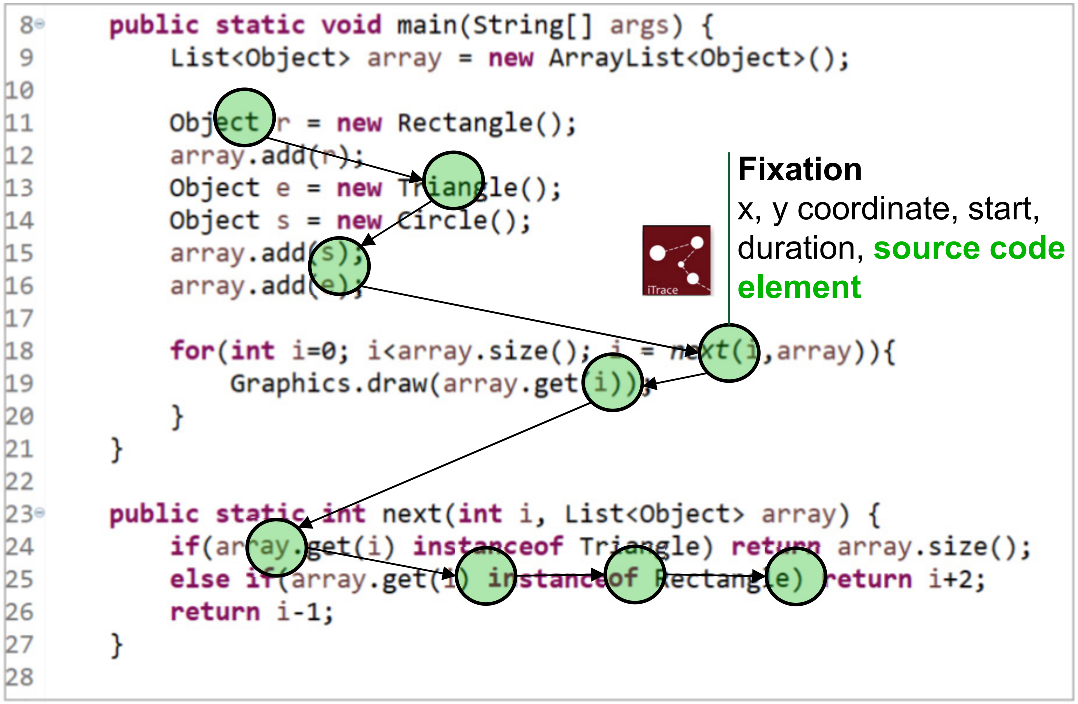
➜ **State-of-the art approach to analyze eye-tracking data**

Challenges:

- **Scrolling and zooming changes view of artifact**
- **Complex and dynamic artifacts**
- **Manual creation of AOIs practically not feasible**

Universität St.Gallen

Mapping of Events with Artifact



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Fixation
x, y coordinate, start, duration, **source code element**

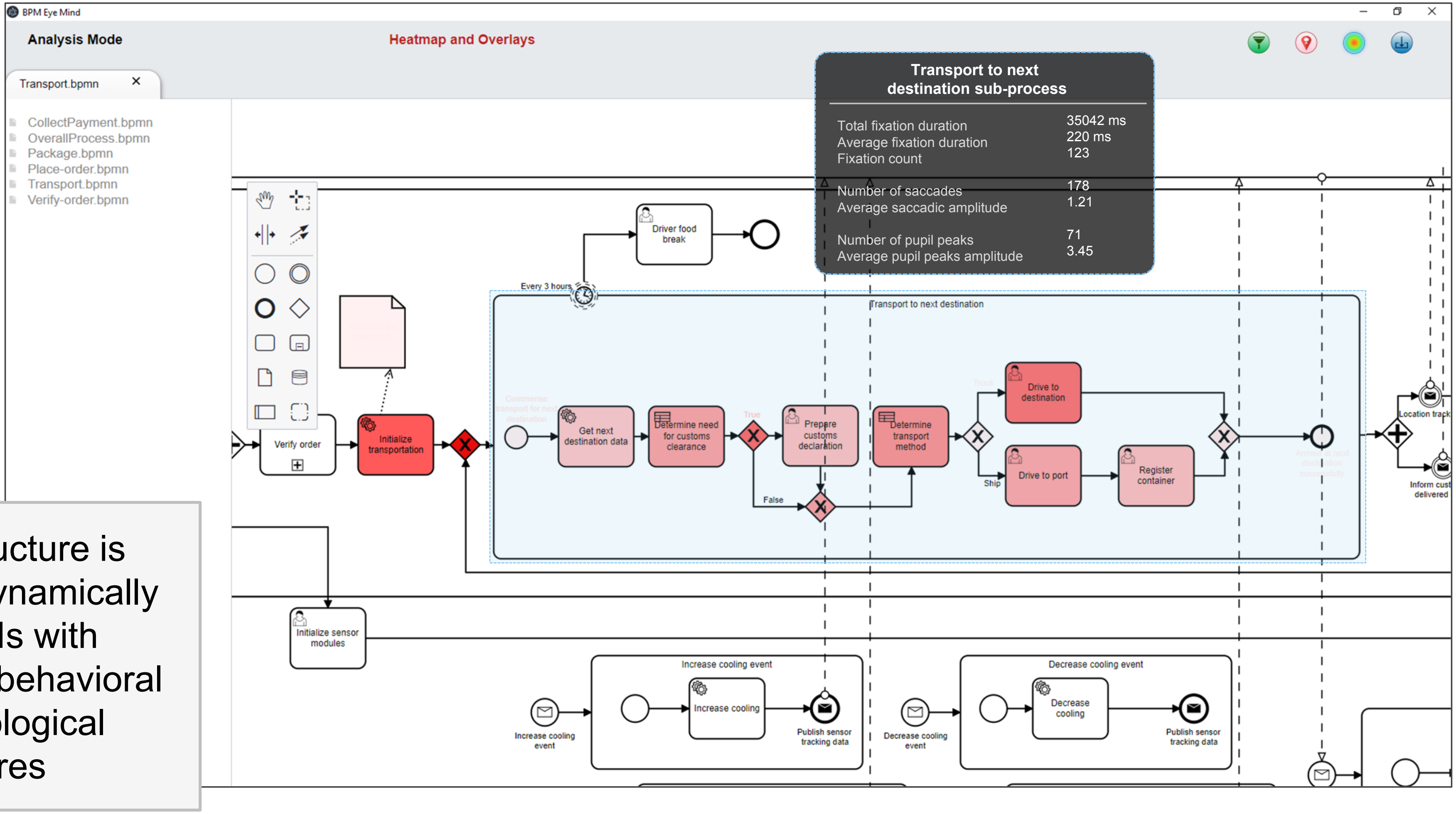
To which part of the artifcact does an event relate?

Enrichment of fixations with artifact information during data collection

Fixation event enrichment should not be an afterthought

Plugins for mapping gazes to artifacts: iTrace, EyeMind (tool developed in my team by Amine Abbad Andaloussi)

Eye Mind: Tool for Collecting and Analyzing Enriched Fixation Events



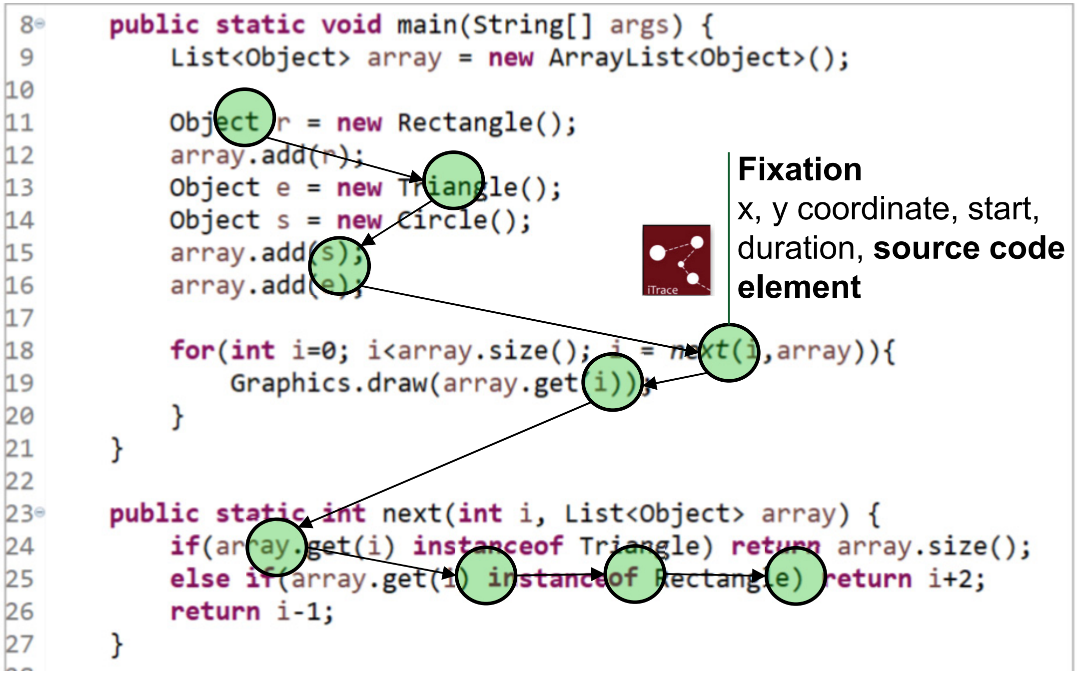
Transport to next destination sub-process

Total fixation duration	35042 ms
Average fixation duration	220 ms
Fixation count	123
Number of saccades	178
Average saccadic amplitude	1.21
Number of pupil peaks	71
Average pupil peaks amplitude	3.45

Process structure is exploited to dynamically create AOIs with projections of behavioral and physiological measures

Universität St.Gallen

Leveraging Artifact Information for Dynamic AOI Creation



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Fixation
x, y coordinate, start, duration, **source code element**

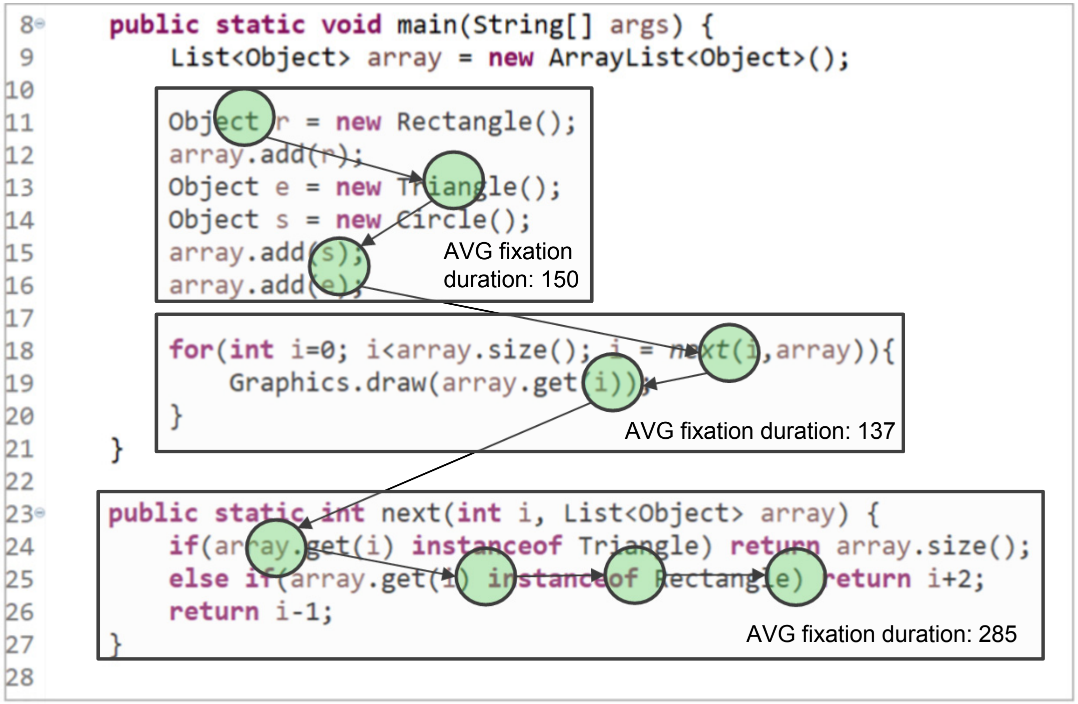
Mapping between fixation events and artifact allows to **create dynamic AOIs** leveraging artifact syntax and semantics**.**

Artifact syntax & semantics	*Abstract syntax tree* *Conditions, loops, concurrency, sequence flow, constraints ...*

Universität St.Gallen

1. Thierry Related Work

Areas of Interest
Grouping and Aggregation of Fixation Events



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

Sequence of AOIs over time

Areas of Interest (AOIs)

Fixations have a temporal and a spatial component. Areas of interest allow for a **spatial grouping and aggregation** of fixations

➜ **State-of-the art approach to analyze eye-tracking data**

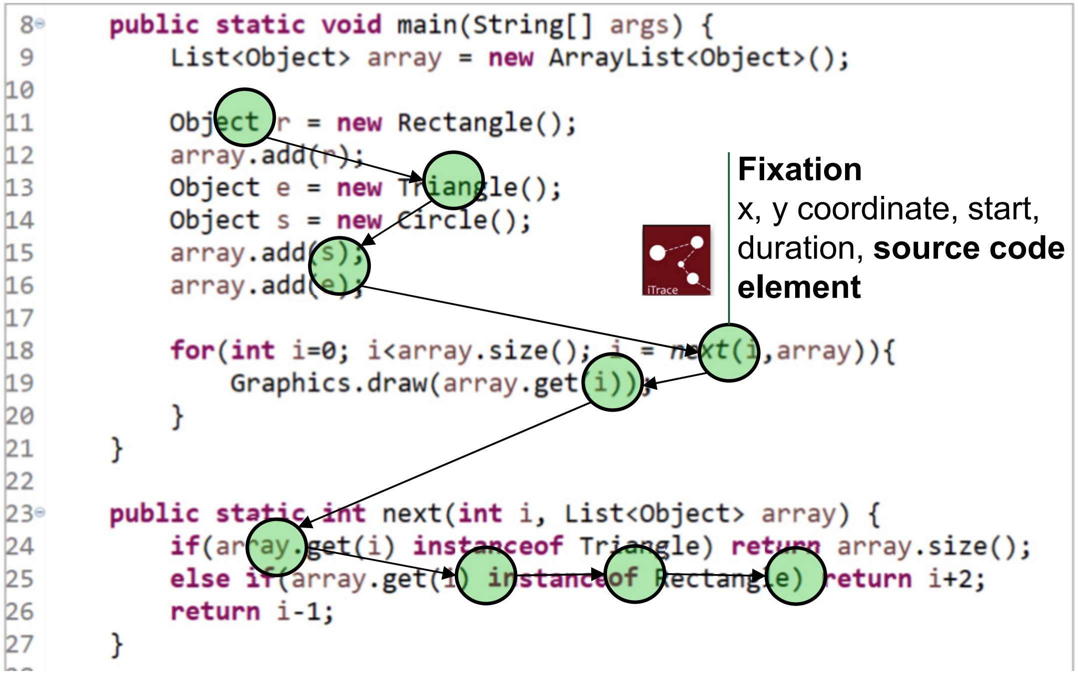
Challenges:

- **Scrolling and zooming changes view of artifact**
- **Complex and dynamic artifacts**
- **Manual creation of AOIs practically not feasible**

Mapping between fixation events and artifact allows to overcome these challenges

Contextualization of Fixation Events

Leveraging Artifact Metrics & Features



```
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();

    Object r = new Rectangle();
    array.add(r);
    Object e = new Triangle();
    Object s = new Circle();
    array.add(s);
    array.add(e);

    for(int i=0; i<array.size(); i = next(i,array)){
        Graphics.draw(array.get(i));
    }
}

public static int next(int i, List<Object> array) {
    if(array.get(i) instanceof Triangle) return array.size();
    else if(array.get(i) instanceof Rectangle) return i+2;
    return i-1;
}
```

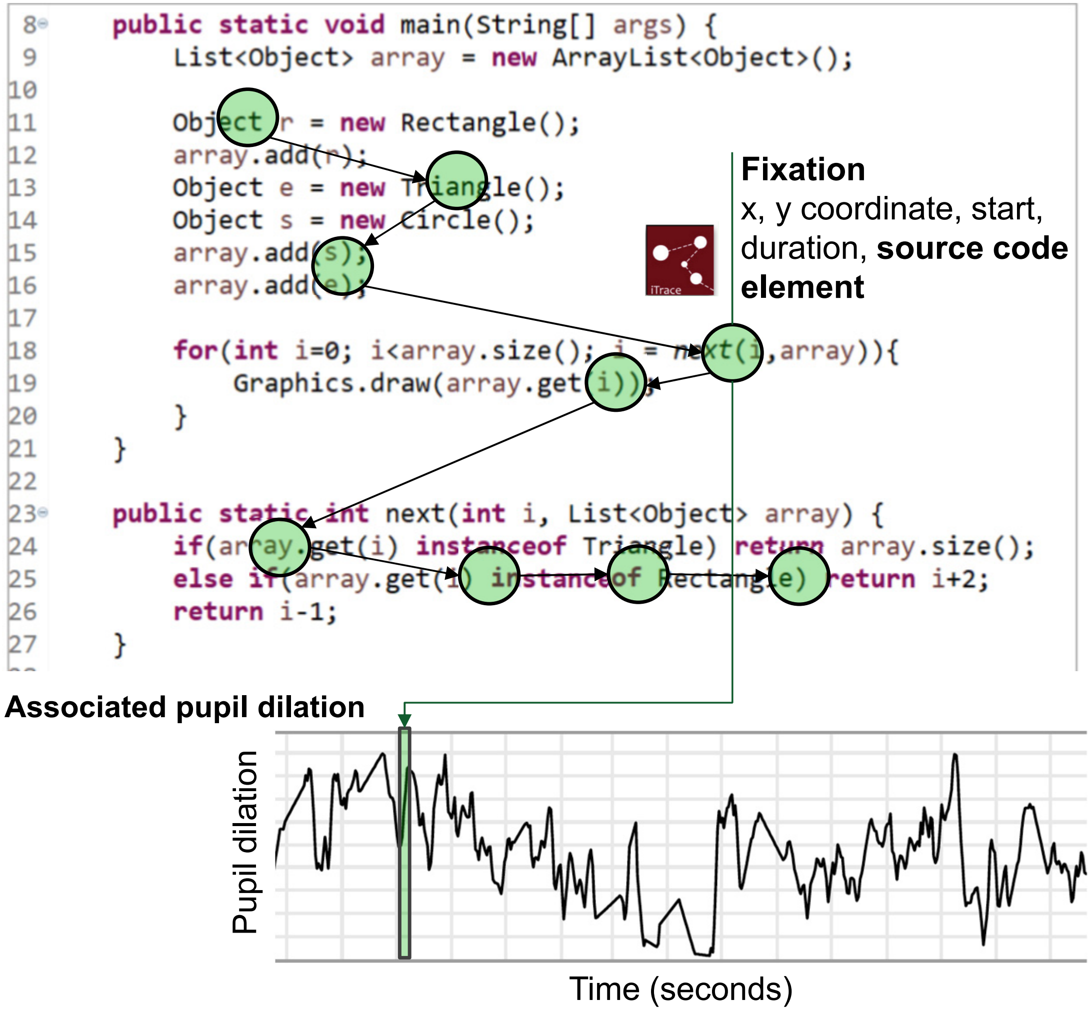
Fixation
x, y coordinate, start, duration, **source code element**

Mapping between fixation events and artifact additionally allows to **contextualize fixation events** with artifact metrics and properties

Artifact metrics & features	*Complexity metrics* *Lexical Anti-patterns* *Structural anti-patterns*

Contextualization of Fixation Events

Leveraging Artifact and Physiological & Behavioral Features



```
    public static void main(String[] args) {
        List<Object> array = new ArrayList<Object>();

        Object r = new Rectangle();
        array.add(r);
        Object e = new Triangle();
        Object s = new Circle();
        array.add(s);
        array.add(e);

        for(int i=0; i<array.size(); i = next(i,array)){
            Graphics.draw(array.get(i));
        }
    }

    public static int next(int i, List<Object> array) {
        if(array.get(i) instanceof Triangle) return array.size();
        else if(array.get(i) instanceof Rectangle) return i+2;
        return i-1;
    }
```

Fixation
x, y coordinate, start, duration, **source code element**

Associated pupil dilation

Mapping between fixation events and artifact additionally allows to **contextualize fixation events** with **artifact metrics & features**

Artifact metrics & features	*Complexity metrics* *Lexical Anti-patterns* *Structural anti-patterns*

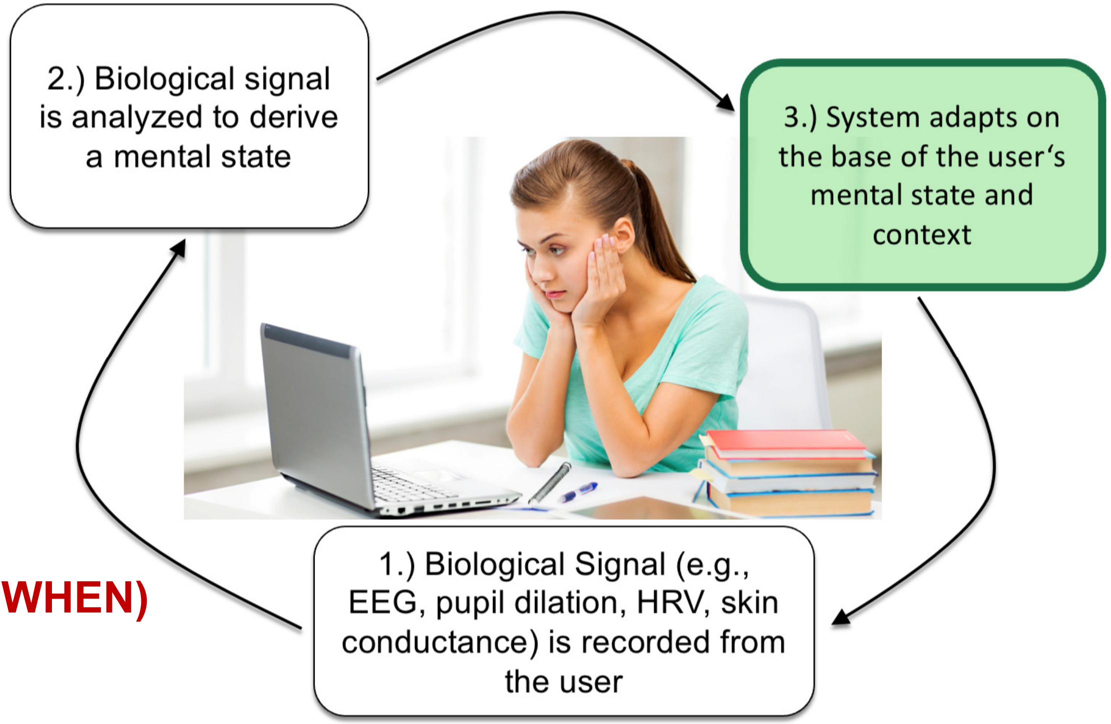
The temporal dimension can be used to **contextualize fixation events** with **behavioral & physiological features.**

Physiological & behavioral features	*Fixation-related measures* *Saccadic measures* *Pupil measures*

1. Contextualization of fixation events
2. Pinpoint difficult parts of code
3. Augmented Representations
   - E.g., for code reviews
4. Create Neuro-adaptive Software Systems

From Cognitive States to Neuro-adaptive Software Systems

Neuro-adaptive software systems are software systems that adapt to changes in user‘s mental state (i.e., cognitive or affective state).



2.) Biological signal is analyzed to derive a mental state

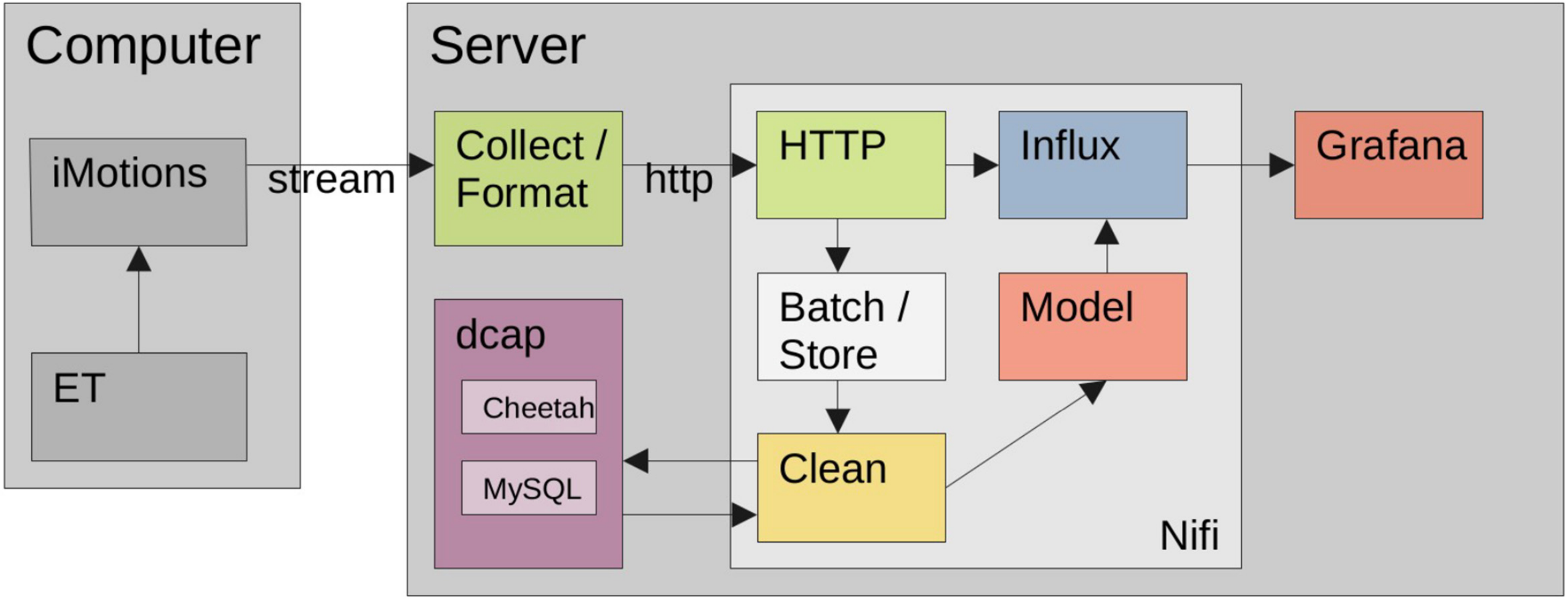
3.) System adapts on the base of the user‘s mental state and context

1.) Biological Signal (e.g., EEG, pupil dilation, HRV, skin conductance) is recorded from the user

Challenge:
Nature of Intervention (HOW) and Timing (WHEN)

Event-driven architectures: A Good Fit

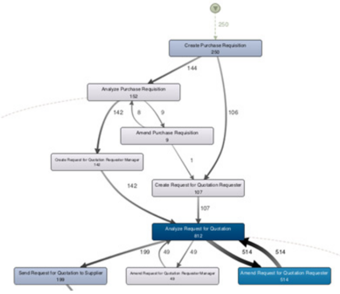
Prototype for Online Cognitive Load Prediction (and Offline Training)



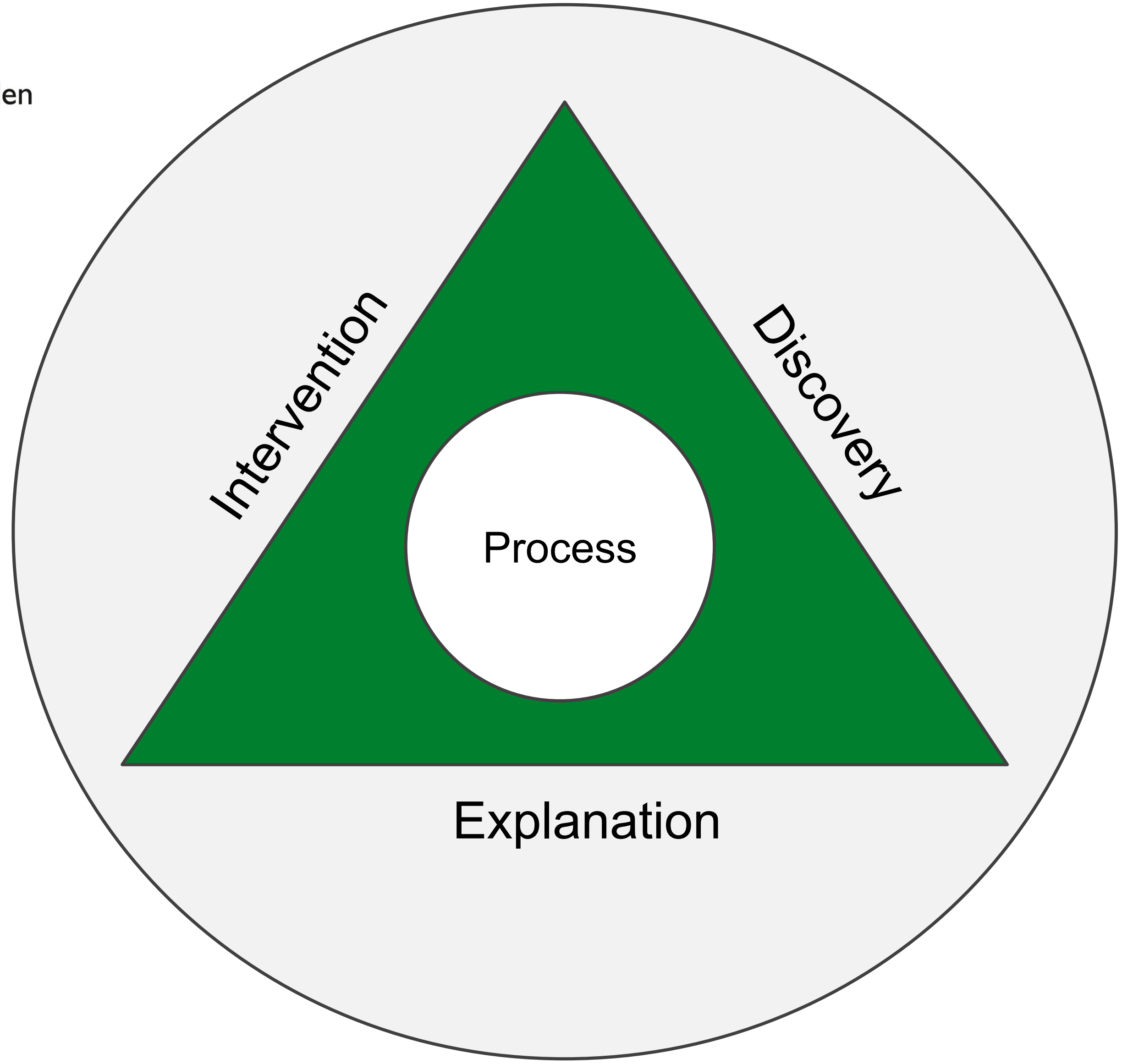
Prototype developed in our research group

supported by HASLERSTIFTUNG

Selected Examples



Process of Process Mining



Cognitive Processes

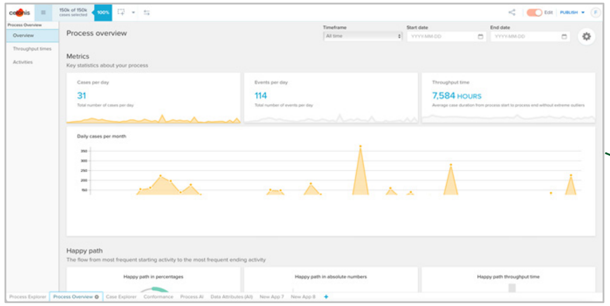
IoT-enabled business processes

Collection of Interaction Data and Verbal Data

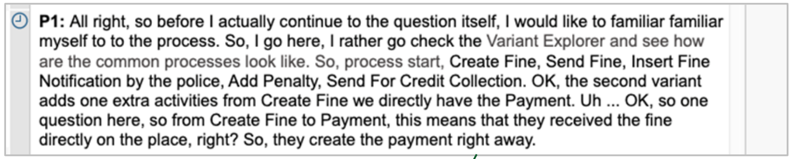
Raw Data

Interaction Data: Screen recordings

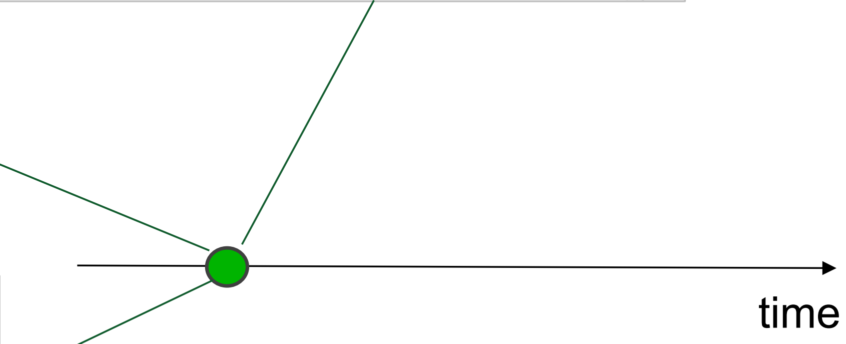
Screen recording



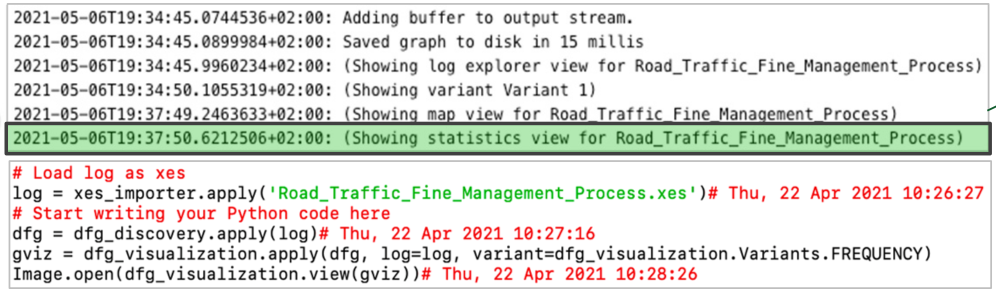
Verbal Data



P1: All right, so before I actually continue to the question itself, I would like to familiar familiar myself to to the process. So, I go here, I rather go check the Variant Explorer and see how are the common processes look like. So, process start, Create Fine, Send Fine, Insert Fine Notification by the police, Add Penalty, Send For Credit Collection. OK, the second variant adds one extra activities from Create Fine we directly have the Payment. Uh ... OK, so one question here, so from Create Fine to Payment, this means that they received the fine directly on the place, right? So, they create the payment right away.



Interaction Data: Application level logs



```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```

```
# Load log as xes
log = xes_importer.apply('Road_Traffic_Fine_Management_Process.xes')# Thu, 22 Apr 2021 10:26:27
# Start writing your Python code here
dfg = dfg_discovery.apply(log)# Thu, 22 Apr 2021 10:27:16
gviz = dfg_visualization.apply(dfg, log=log, variant=dfg_visualization.Variants.FREQUENCY)
Image.open(dfg_visualization.view(gviz))# Thu, 22 Apr 2021 10:28:26
```

Challenge: Synchronization of events

Challenge: Raw events have different abstraction levels

Creation of User Behavior Logs

Challenge: What are events indicating relevant state changes?

- **Relevance** depends on the **purpose** and is determined by the research/analysis question

  Example Questions:
  - **How was the evolution of tool usage over time?**
  - How did artifact usage evolve over time?
  - How are questions developed within process mining projects?
  - What are strategies for validating and verifying analysis artifacts/findings?

- Events indicating relevant state changes can refer to multiple **dimensions.** Each dimension can be described at different **abstraction levels.**

  Possible dimensions:
  - **Tool function**
  - Target artifact

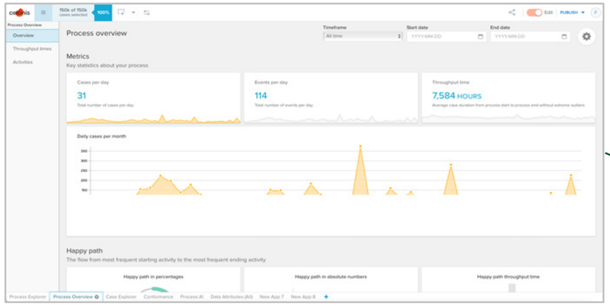
F. Zerbato, R. Seiger, G. Di Federico, A. Buarttin and B. Weber: Granularity in Process Mining: Can We Fix It? Problems@BPM 2021: 40-44.

Collection of Interaction Data and Verbal Data

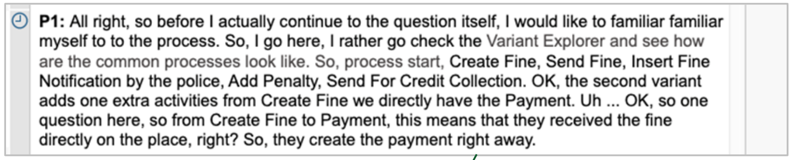
Raw Data

Interaction Data: Screen recordings

Screen recording

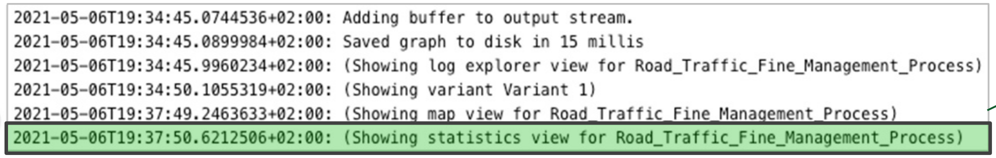


Verbal Data

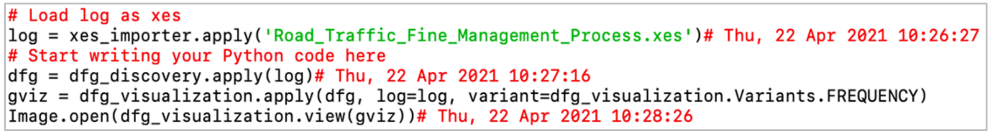


P1: All right, so before I actually continue to the question itself, I would like to familiar familiar myself to to the process. So, I go here, I rather go check the Variant Explorer and see how are the common processes look like. So, process start, Create Fine, Send Fine, Insert Fine Notification by the police, Add Penalty, Send For Credit Collection. OK, the second variant adds one extra activities from Create Fine we directly have the Payment. Uh ... OK, so one question here, so from Create Fine to Payment, this means that they received the fine directly on the place, right? So, they create the payment right away.

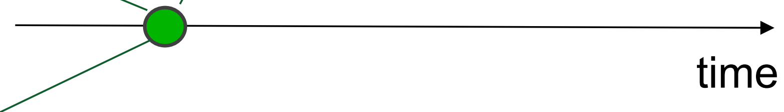
Interaction Data: Application level logs



```
2021-05-06T19:34:45.0744536+02:00: Adding buffer to output stream.
2021-05-06T19:34:45.0899984+02:00: Saved graph to disk in 15 millis
2021-05-06T19:34:45.9960234+02:00: (Showing log explorer view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:34:50.1055319+02:00: (Showing variant Variant 1)
2021-05-06T19:37:49.2463633+02:00: (Showing map view for Road_Traffic_Fine_Management_Process)
2021-05-06T19:37:50.6212506+02:00: (Showing statistics view for Road_Traffic_Fine_Management_Process)
```



```
# Load log as xes
log = xes_importer.apply('Road_Traffic_Fine_Management_Process.xes')# Thu, 22 Apr 2021 10:26:27
# Start writing your Python code here
dfg = dfg_discovery.apply(log)# Thu, 22 Apr 2021 10:27:16
gviz = dfg_visualization.apply(dfg, log=log, variant=dfg_visualization.Variants.FREQUENCY)
Image.open(dfg_visualization.view(gviz))# Thu, 22 Apr 2021 10:28:26
```



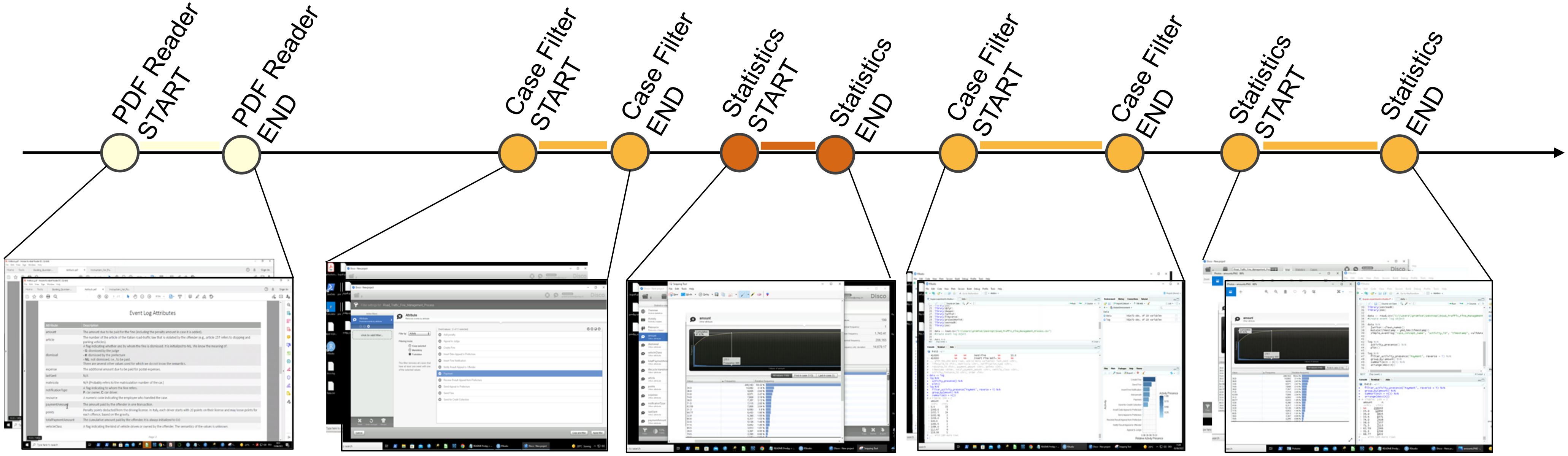
Challenge: Synchronization of events

Challenge: Raw events have different abstraction levels

User Behavior Log (Interaction Data)

Event sequence showing tool usage over time

Participant	Event	Timestamp
P1	Desktop view started	0:07:50,3
P1	Desktop view ended	0:08:13,4
…	…	…

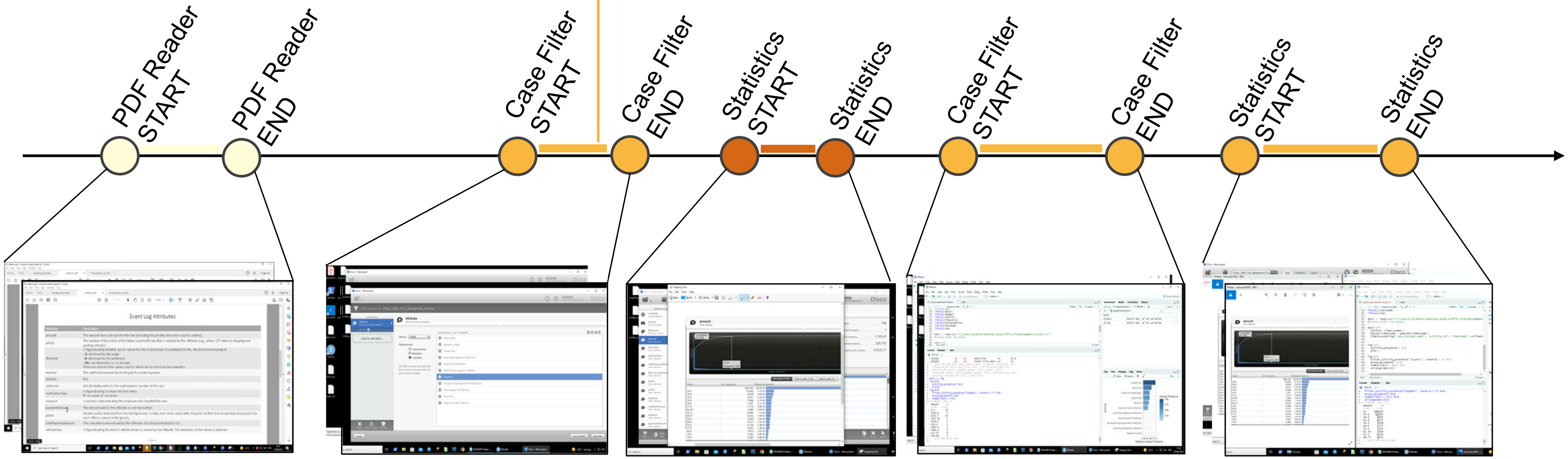


Universität St.Gallen

Interaction Data + Think Aloud

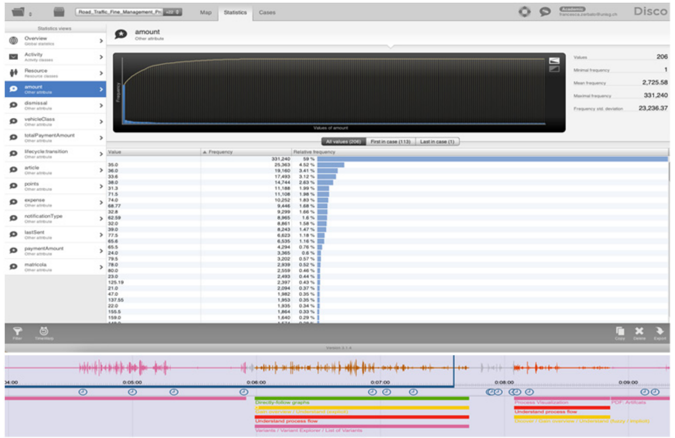
Think aloud data helps to interpret interaction data

[00:10:18 – 00:10:39]
"I'm trying to filter the Payment activity to see all the cases that don't have a payment. I've tried using the filter forbidden..."



Universität St.Gallen

Analyzing User Behavior Logs

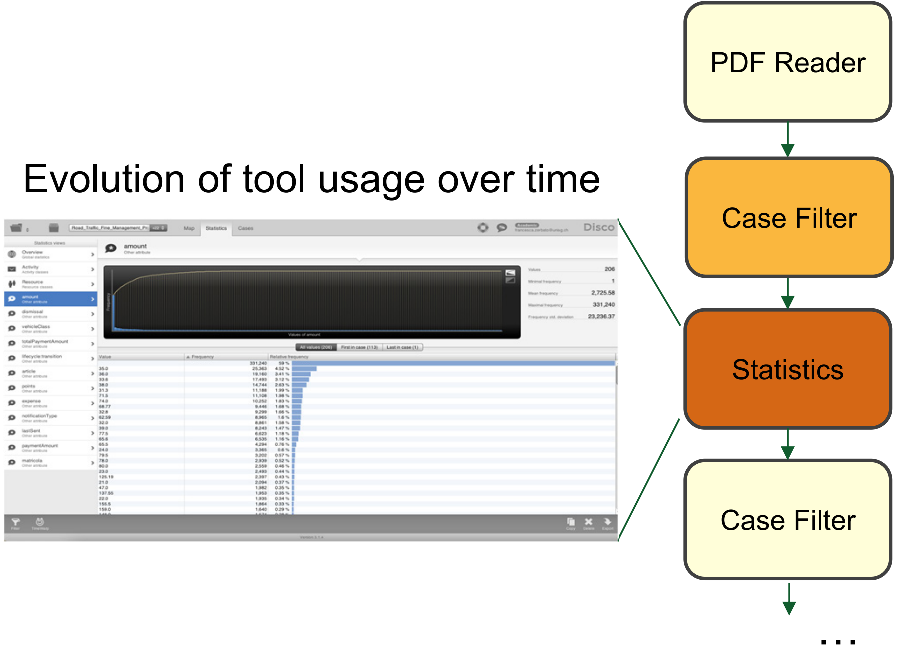


Event Sequence Data (Tool Perspective)

Participant	Event	Timestamp
P27	PDF Reader START	00:07:50,3
P27	PDF Reader END	00:08:13,4
P27	Case Filter START	00:08:14,4
P27	Case Filter END	00:08:48,7
P27	Statistics START	00:08:50,0
P27	Statistics END	00:11:18,1
P27	Case Filter START	00:11:30,6
P27	Case Filter END	00:11:49,8

Challenge: How can we analyze event sequences ?

Evolution of tool usage over time



PDF Reader

Case Filter

Statistics

Case Filter

…

Evolution of tool usage over time

13 tool functions, time intervals

6 tool functions, time intervals

6 tool functions, time points

What we put into the event sequences affects what analyses we can do later on and what patterns we can discover!

Pattern Discovery and Visualization

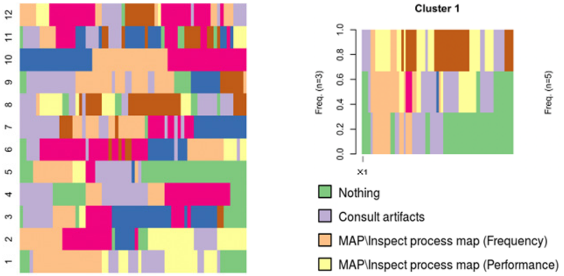
Challenge: How can we support pattern discovery through visualizations?

Discovering Patterns of Behavior in Event Sequences

Event Sequence Data

Participant	Event	Timestamp
P1	Desktop view started	0:07:50,3
P1	Desktop view ended	0:08:13,4
…	…	…

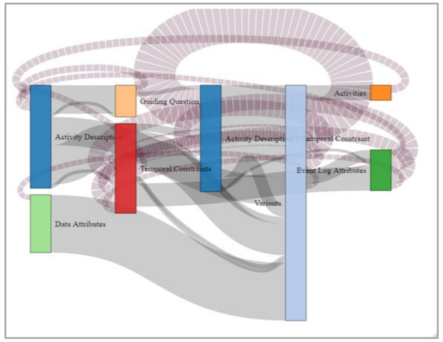
Hierarchical Cluster Analysis



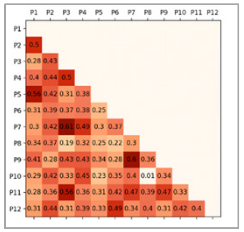
Sequential Pattern Mining Analysis

Patterns can emerge within one user or across multiple users.

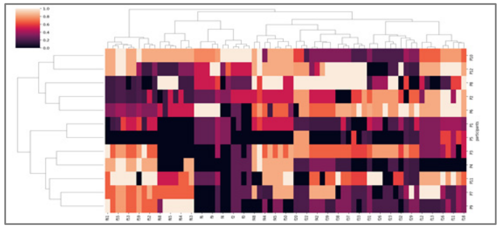
Scarf plot



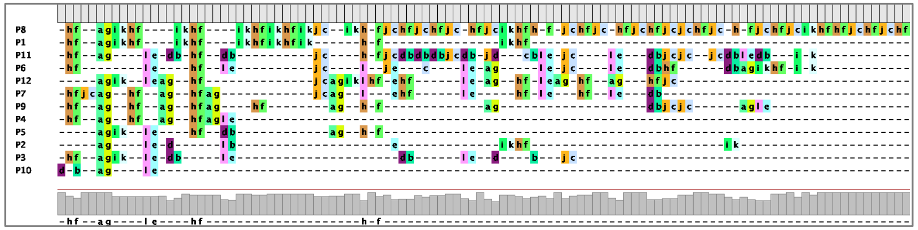
Sankey diagram



Correlation matrix



Dendogram

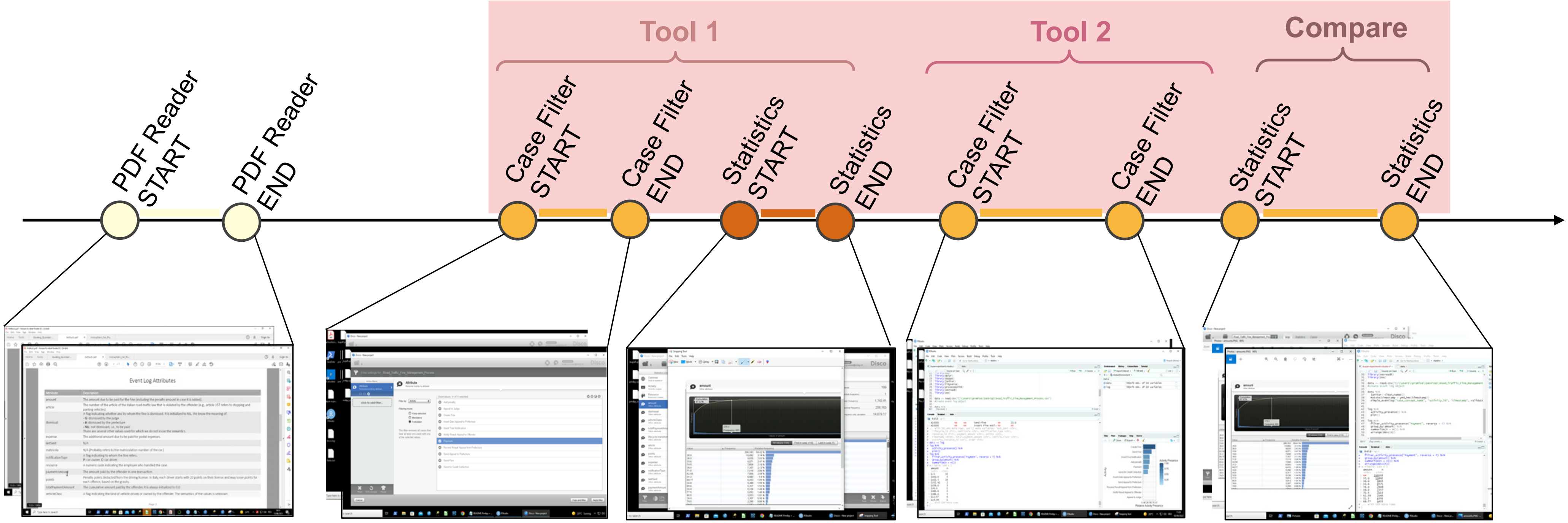


Multiple sequence alignment

Process Mining Strategies

Verify artifacts/findings

"And often I try to combine different tools to understand for sanity check if we have the same insights in different tools." (Interview Data)



Tool 1

Tool 2

Compare

PDF Reader START

PDF Reader END

Case Filter START

Case Filter END

Statistics START

Statistics END

Case Filter START

Case Filter END

Statistics START

Statistics END

Universität St.Gallen

Event Sequence Analysis and Visualization

S

Yi Guo, S

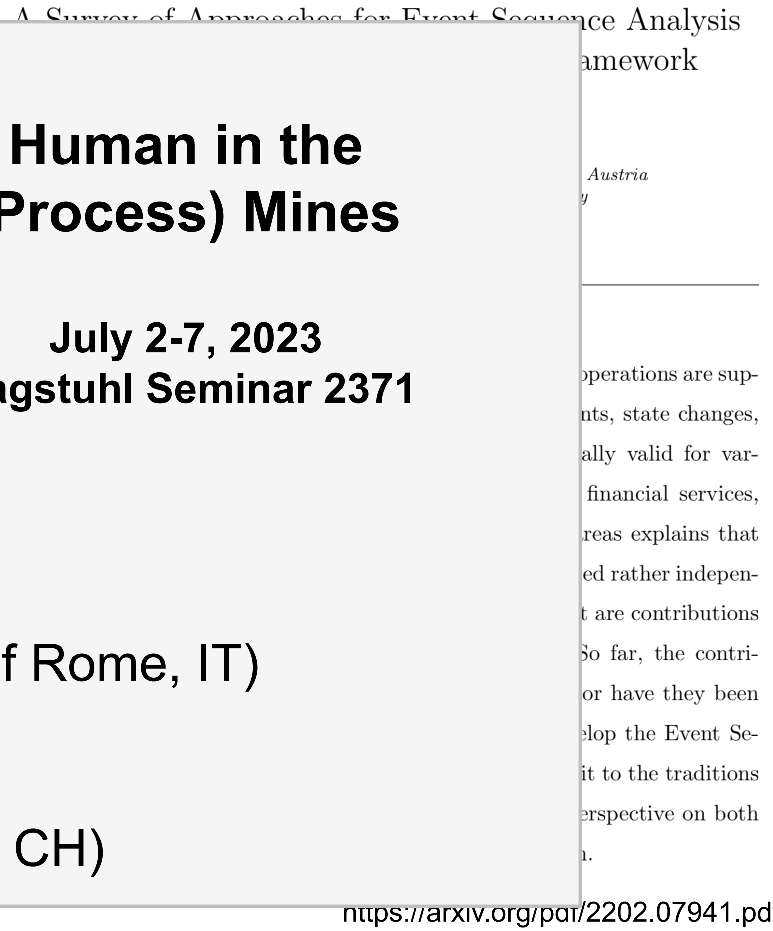
Abstract—Event sequence c
variety of applications rangin
and heterogeneous. This hig
resulting in ever-increasing n
insights from event sequence
our proposed design space, a
have also identified several re

Index Terms—Visual Analys

Coping w
Event Sec

Fan Du, Be

Abstract—The growing volum
challenges is needed for big d
In the case of temporal event
challenges users of visual ana
reduce the data volume and p
pattern simplification strategies, and (4) iterative strategies. For each strategy, we provide examples of the use and impact of this
strategy on volume and/or variety. Examples are selected from 20 case studies gathered from either our own work, the literature, or
based on email interviews with individuals who conducted the analyses and developers who observed analysts using the tools. Finally,
we discuss how these strategies might be combined and report on the feedback from 10 senior event sequence analysts.



A Survey of Approaches for Event Sequence Analysis
amework

Austria
y

operations are sup-
nts, state changes,
ally valid for var-
financial services,
reas explains that
ed rather indepen-
t are contributions
So far, the contri-
or have they been
elop the Event Se-
it to the traditions
erspective on both
n.

https://arxiv.org/pdf/2202.07941.pdf

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Human in the (Process) Mines

July 2-7, 2023
Dagstuhl Seminar 2371

Organizers:

Claudio Di Ciccio (Sapiezna University of Rome, IT)
Silvia Miksch (TU Wien, AT)
Pnina Soffer (Haifa University, IL)
Barbara Weber (University of St. Gallen, CH)

Summary

- **Events are everywhere** and come in many flavors
- It pays off to **carefully plan data collection** and to not just assume that things can be combined retrospectively (e.g., modeling and emitting domain events versus only reconstructing events from low-level interactions)
- Huge potential of **multi-modal data** and **contextualization of events**. Having an infrastructure to enable synchronized collection of multi-modal data and contextualization of events is instrumental
- Be aware of **confounding factors** when moving from discovery to explanation and intervention (they are often not latent!)
- Potential for **synergies** with **neighboring communities** working with event data and building event-driven systems and potential for **interdisciplinary research** taking a process science perspective

Thank You to My Team

New

Human and Cognitive Aspects

Amine Abbad Andaloussi
Postdoctoral Researcher

Thierry Sorg
PhD student

BPM & IoT

Ronny Seiger
Assistant Professor

Flemming Weyers
PhD student

Process Mining

Francesca Zerbato
Postdoctoral Researcher

Lisa Zimmermann
PhD student

.. and our collaborators

Events: Definition and Function

Multiple events can be related by time, causality, abstraction, or other relationships

Observation

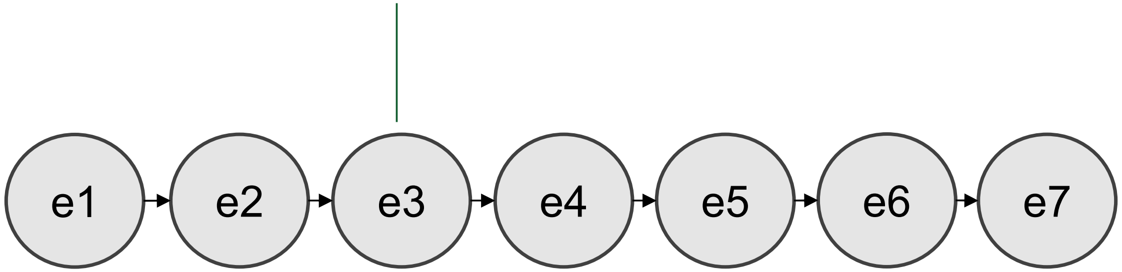
👁 👂 ✋ 👅 👃

Information Systems, Sensors

Occurence (Events)

Record about past occurence

Log, Stream, Event Sequence



Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)

Universität St.Gallen

The Many Flavors of Event Sequences
CAiSE 2016-2021

Overall 63 out of 216 papers related to events
Key role of event sequences in 52 papers

Process Event Logs (42)
Traces with events related to BP execution
Events related to BP execution (unkeyed)
Augmented Process Event Logs
Collection of Process Event Logs

Log of low-level Interactions (4)
Location data (keyed by device / entity)
Mouse clicks (including screen capture), keystrokes (unkeyed)
Low-level events (keyed by process ID)

Streams of (Raw) Events (4)
Stream of instantaneous and atomic event occurence
Events related to BP execution (keyed by instance ID)
Multiple streams with heart-related, movement-related, and vehicle state events

Cyber-physical System logs (1)
Events of different type related to CPS components (e.g., drones, metalurgical plant); keyed by scope

Customer Journeys (1)
Long-running traces of customer interactions

Universität St.Gallen

The Many Flavors of Event Sequences

CAiSE 2016-2021

Key role of event sequences in 52 papers

Observations

- Process event logs well represented (42 out of 52)
- Primary focus on logs versus streams (48 out of 52)
- Most works focus on a single log / stream (48 out of 52)
- Most works have logs/streams with keys (2 out of 52)
- Few works explicitly refer to distributed settings (3 out of 52)
- Few works refer to a data collection infrastructure (3 out of 52)

Process Event Logs (42)

Streams of (Raw) Events (4)

Log of low-level Interactions (4)

Cyber-physical system (CPS) logs (1)

Customer Journeys (1)

Events: Definition and Function

Events from a Complex Event Processing Lense

- Raw events
- Instantaneous events
- Simple events
- Complex events
- Derived events
- Composite events
- Virtual events

Observation

👁 👂 ✋ 👅 👃

Information Systems, Sensors

Occurence (Events)

Record about past occurence

Log, Event Sequence

e1 → e2 → e3 → e4 → e5 → e6 → e7

Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)

Universität St.Gallen

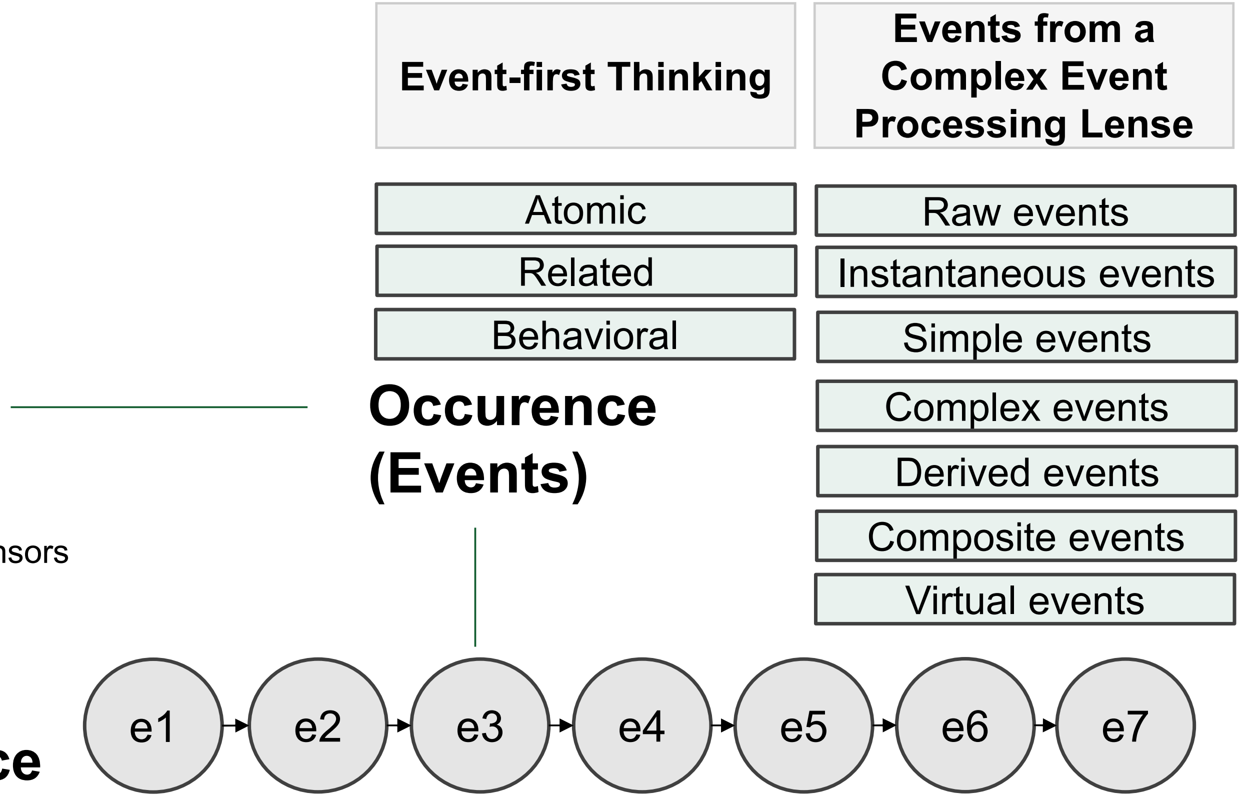
The Many Flavors of Events

Events seen with a Complex Event Processing lense

Raw event	An event object that records a real-world event.
Instantaneous event	An event that happens at a point in time.
Simple event	An event that is not viewed as summarizing, representing, or denoting a set of other events.
Complex event	An event that summarizes, represents, or denotes a set of other events.
Derived event (Synthetic, synthesized)	An event that is generated as a result of applying a method or process to one or more other events.
Composite event	A derived event that is created by combining a set of other simple or complex events (known as members) using a specific set of event constructors such as disjunction, conjunction, and sequence. It includes the member events from which it is derived.
Virtual Event	An event that does not happen in the real world, but is imagined, modeled, or simulated.

The Event Processing Glossary described events with a Complex Event Processing lense: https://complexevents.com/wp-content/uploads/2011/08/EPTS_Event_Processing_Glossary_v2.pdf

Events: Definition and Function



Event-first Thinking

- Atomic
- Related
- Behavioral

Events from a Complex Event Processing Lense

- Raw events
- Instantaneous events
- Simple events
- Complex events
- Derived events
- Composite events
- Virtual events

Observation

👁 👂 ✋ 👅 👃

Information Systems, Sensors

Occurence (Events)

Record about past occurence

Log, Stream, Event Sequence

e1 → e2 → e3 → e4 → e5 → e6 → e7

Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)

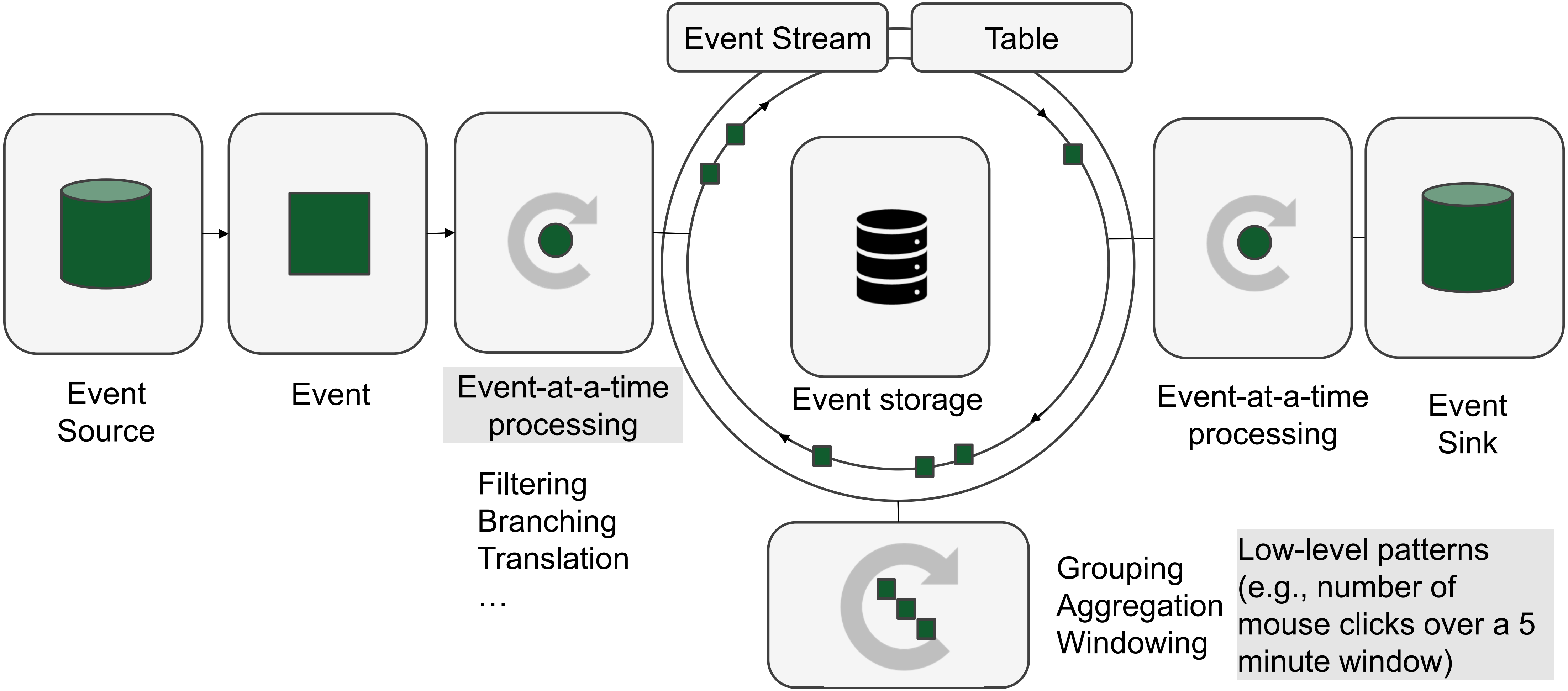
The Many Flavors of Events

Event-first Thinking

Atomic	Something happened (bid on an item, device temperature).
Related	A stream or sequence of events (tracking a price change, device metrics changes over time).
Behavioral	The accumulation of facts capture behavior. Collecting, remembering, and analyzing the facts allows us to recogonize and react to behavior.

Blog post on event-first thinking: https://www.confluent.io/blog/journey-to-event-driven-part-1-why-event-first-thinking-changes-everything/

Event Processing and Stream Processing



Event Stream

Table

Event Source

Event

Event-at-a-time processing

Filtering
Branching
Translation
…

Event storage

Stream processing

Grouping
Aggregation
Windowing

Low-level patterns (e.g., number of mouse clicks over a 5 minute window)

Event-at-a-time processing

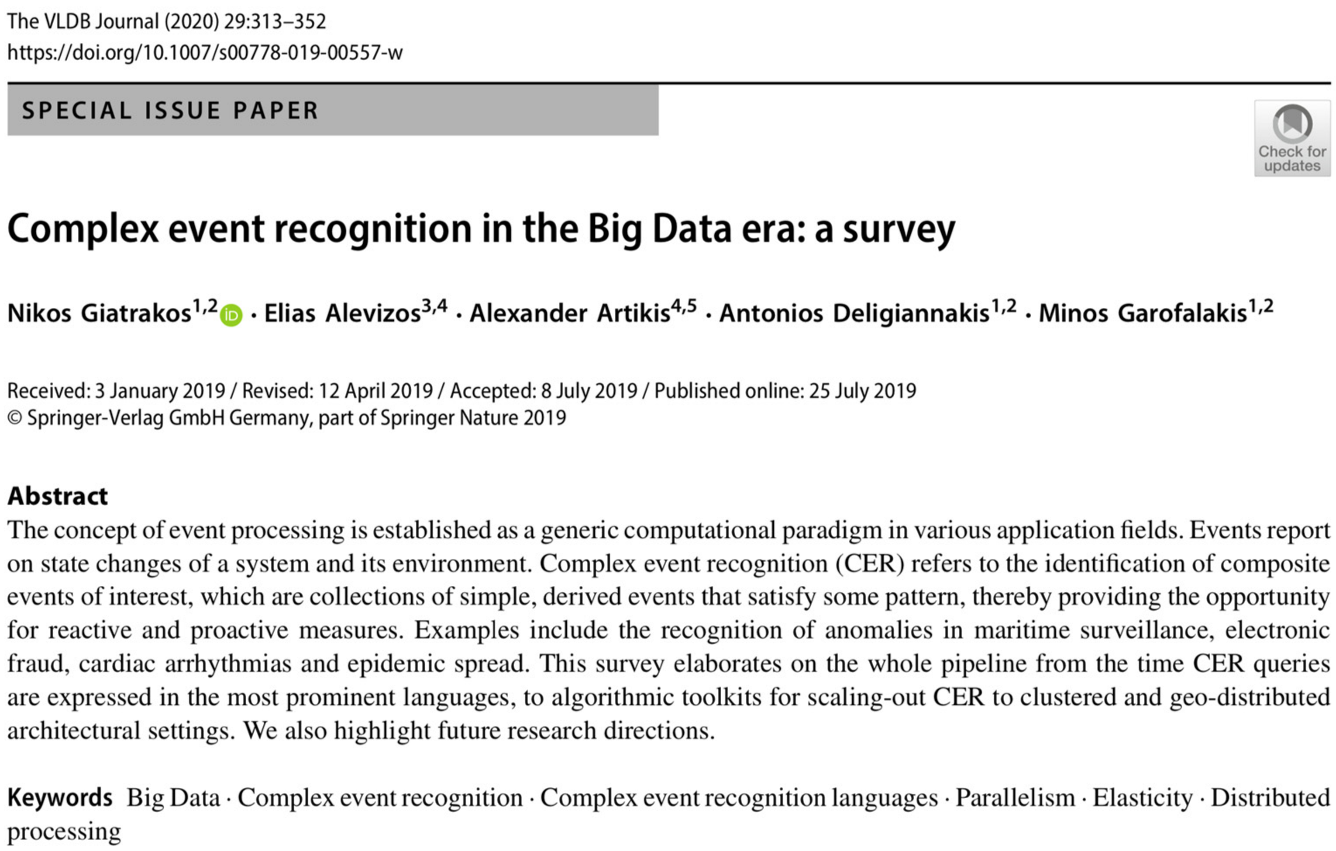
Event Sink

Complex Event Recognition

Complex patterns involving multiple events
- Using **conceptual hierarchies**
- Using **causal relationships**

Events report on **state changes of a system** and its environment

Identification of **composite events of interest** […] that satisfy some **patterns**, thereby providing the opportunity for **reactive and proactive measures**.



The VLDB Journal (2020) 29:313–352
https://doi.org/10.1007/s00778-019-00557-w

SPECIAL ISSUE PAPER

Check for updates

Complex event recognition in the Big Data era: a survey

Nikos Giatrakos[1,2] · Elias Alevizos[3,4] · Alexander Artikis[4,5] · Antonios Deligiannakis[1,2] · Minos Garofalakis[1,2]

Received: 3 January 2019 / Revised: 12 April 2019 / Accepted: 8 July 2019 / Published online: 25 July 2019

Abstract

The concept of event processing is established as a generic computational paradigm in various application fields. Events report on state changes of a system and its environment. Complex event recognition (CER) refers to the identification of composite events of interest, which are collections of simple, derived events that satisfy some pattern, thereby providing the opportunity for reactive and proactive measures. Examples include the recognition of anomalies in maritime surveillance, electronic fraud, cardiac arrhythmias and epidemic spread. This survey elaborates on the whole pipeline from the time CER queries are expressed in the most prominent languages, to algorithmic toolkits for scaling-out CER to clustered and geo-distributed architectural settings. We also highlight future research directions.

Keywords Big Data · Complex event recognition · Complex event recognition languages · Parallelism · Elasticity · Distributed processing

Giatrakos, N., Alevizos, E., Artikis, A. *et al.* Complex event recognition in the Big Data era: a survey. *The VLDB Journal* **29,** 313–352 (2020). https://doi.org/10.1007/s00778-019-00557-w

Distinction Event Stream Processing and Complex Event Processing: https://complexevents.com/2020/06/17/the-future-of-event-stream-analytics-and-cep/

Events: Definition and Function

Domain Driven Design and Microservices
- Domain Events
- Integration Events

Event-first Thinking
- Atomic
- Related
- Behavioral

Events from a Complex Event Processing Lense
- Raw events
- Instantaneous events
- Simple events
- Complex events
- Derived events
- Composite events
- Virtual events

Observation

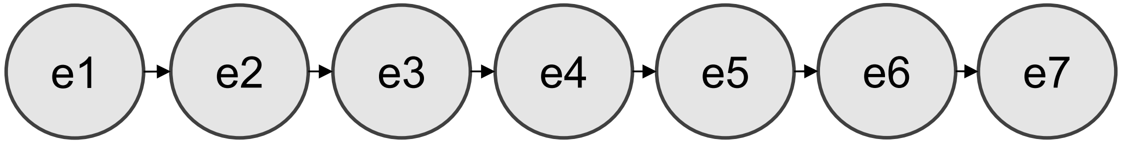
👁 👂 ✋ 👅 👃

Information Systems, Sensors

Occurence (Events)

Record about past occurence

Log, Stream, Event Sequence



Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)

Source: Evans, Eric. 2012. “Domain-Driven Design Reference.” http://domainlanguage.com/ddd/patterns/DDD_Reference_2011-01-31.pdf.

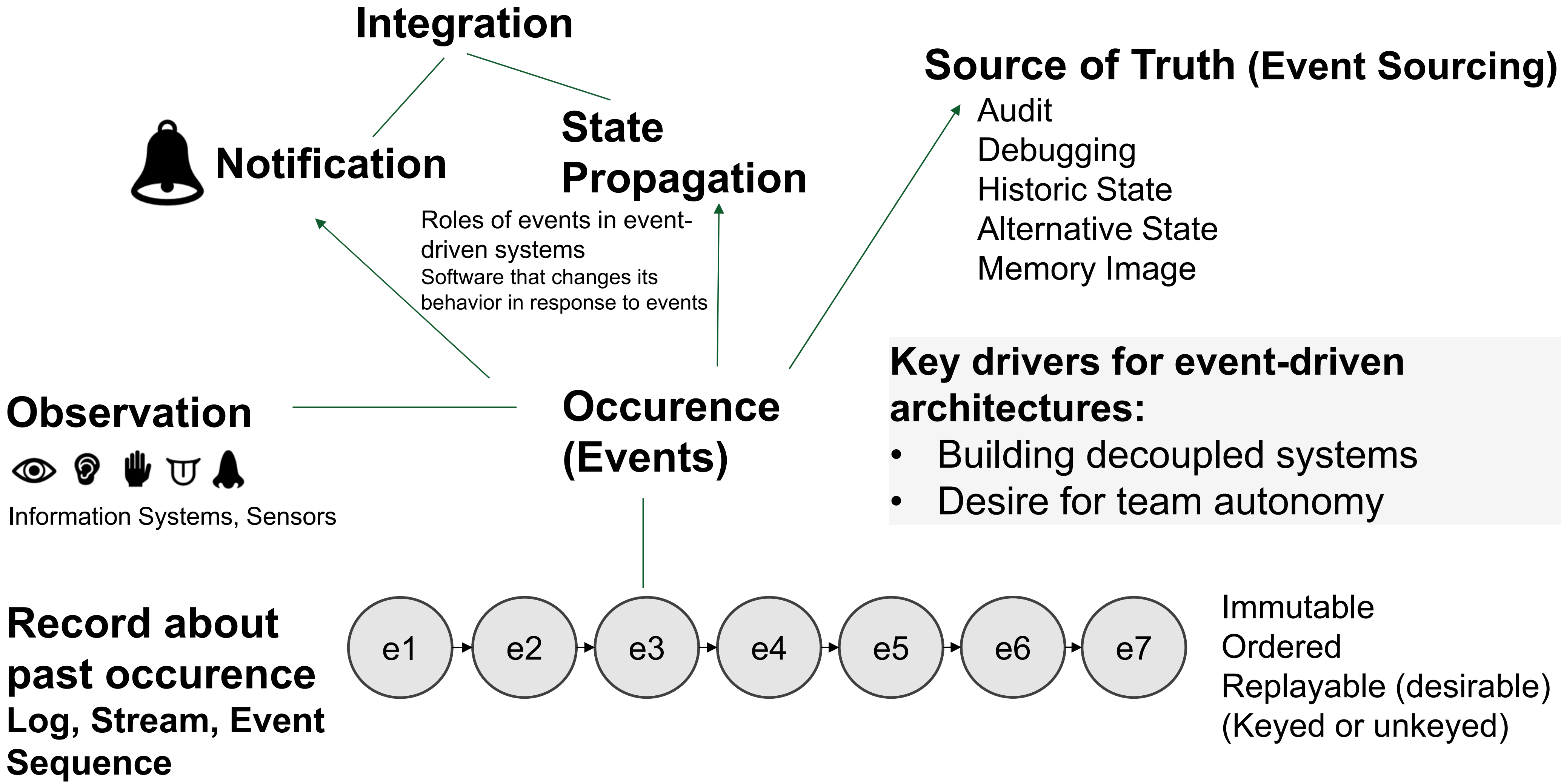
Universität St.Gallen

The Many Flavors of Events

Domain Driven Design and Microservices

Domain events	In domain-driven design **domain events** describe something that has occured in the business domain and is important to domain experts**.** Domain events are part of the domain **model** and expressed in an **ubiquitous language.**
Integration events	The notion of i**ntegration events** is used when publishing events **outside** the service **boundary**.

Events: Definition and Function



Integration

Notification

State Propagation

Roles of events in event-driven systems

Software that changes its behavior in response to events

Source of Truth (Event Sourcing)

Audit
Debugging
Historic State
Alternative State
Memory Image

Observation

Information Systems, Sensors

Occurence (Events)

Key drivers for event-driven architectures:

- Building decoupled systems
- Desire for team autonomy

Record about past occurence

Log, Stream, Event Sequence

Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)

Source: The Many Meanings of Event-driven Architecture by Martin Fowler

The Many Meaning of Event-driven Architecture
CAiSE 2016-2021

Overall 63 papers with focus on events
13 papers with system perspective (requirements, architecture, model-driven development)

Event detection (11)

- Monitoring
- Processing of events
  - Complex event processing
  - Event Processing
  - Stream processing
- Event prediction

Event production (4)

- Event modeling
- Event notification

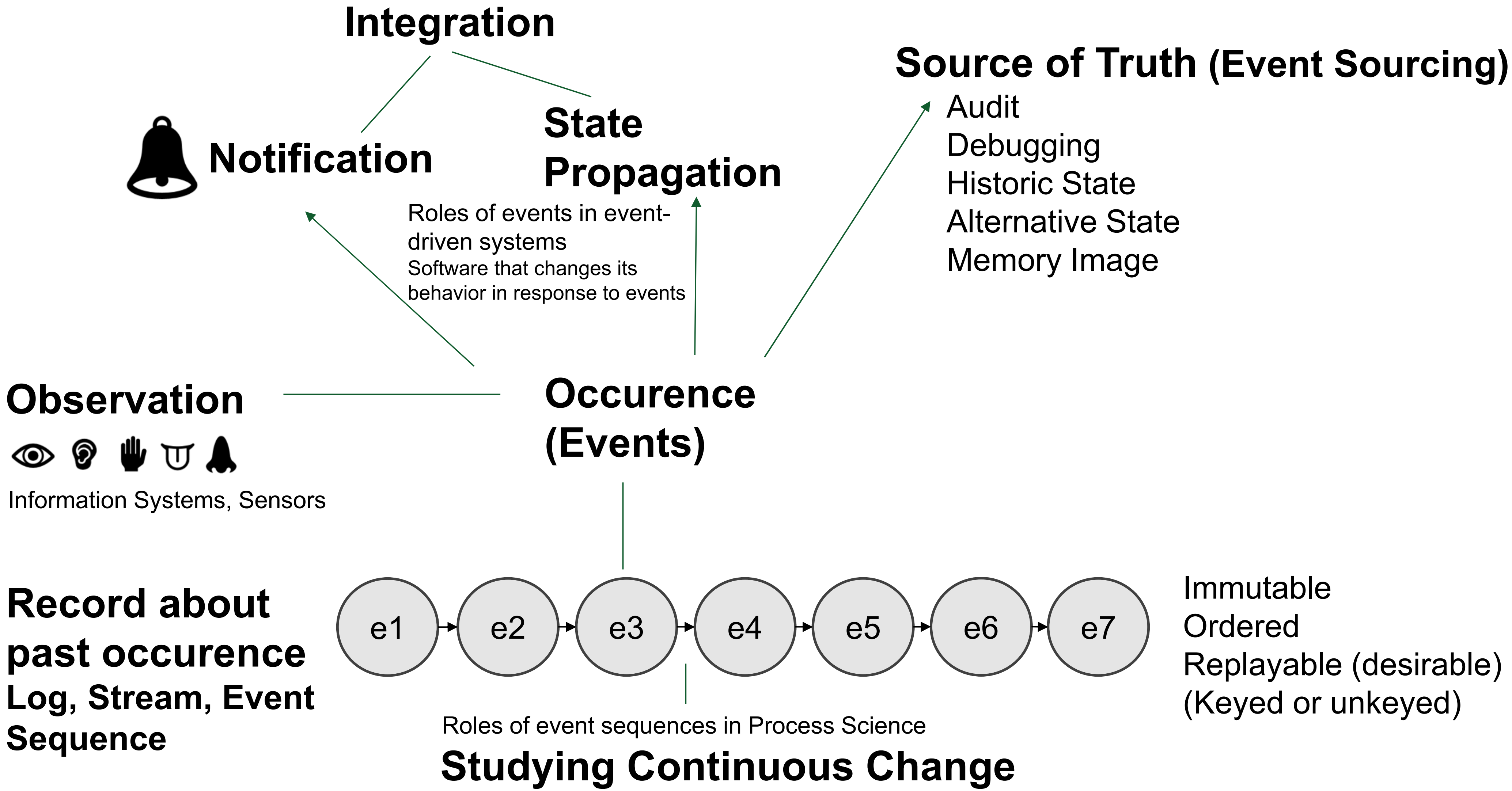
Event handling (8)

- Impact analysis on eff ect
- Reactive event handling
- Proactive event handling

Data Ingestion & Integration (2)

Universität St.Gallen

Events: Definition and Function



Integration

Notification

State Propagation

Roles of events in event-driven systems

Software that changes its behavior in response to events

Source of Truth (Event Sourcing)

- Audit
- Debugging
- Historic State
- Alternative State
- Memory Image

Observation

👁 👂 ✋ 👅 👃

Information Systems, Sensors

Occurence (Events)

Record about past occurence

Log, Stream, Event Sequence

- Immutable
- Ordered
- Replayable (desirable)
- (Keyed or unkeyed)

Roles of event sequences in Process Science

Studying Continuous Change