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

Presented by Barbara Weber
Digital technologies create an ever-increasing volume of digital traces.

Multiple events can be related by time, causality, abstraction, or other relationships.
Gartner published a market guide for process mining in 2018.

Inaugural publication of a Magic Quadrant for Process Mining tools in 2023.
The Potential of Process Mining

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

Creation of «current state» processes
Process Mining: The Big Picture

Source: van der Aalst & Carmona: Process Mining Handbook
**Event Data: The Starting Point for Process Mining**

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

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

Traces of Case 1, 17 and 19:

- **Case 1**
  - Inbound Call: 09.03.2010 08:05 - 09.03.2010 08:10
  - Handle Case: 11.03.2010 10:30 - 11.03.2010 10:32
  - Call Outbound: 11.03.2010 11:45 - 11.03.2010 11:52

- **Case 17**
  - Inbound Call: 04.03.2010 07:35 - 04.03.2010 07:46
  - Handle Case: 08.03.2010 11:16 - 08.03.2010 11:18
  - Inbound Call: 14.03.2010 17:53 - 14.03.2010 17:56

- **Case 19**
  - Inbound Email: 14.03.2010 14:08 - 18.03.2010 08:04
  - Handle Case: 18.03.2010 08:07 - 18.03.2010 08:09

Case ID, activity and at least one timestamp per event are the minimum requirements for an event log.
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**
Process Science Activities

Discovery

Goal. Capture and describe processes.

Example. Methods to create process representations from digital trace data and to identify patterns in processes.

Explanation

Goal. Understand why, how and when a process unfolds.

Example. Methods to study the context in which a pattern is situated.

Intervention

Goal. Intervene and shape the process into desired directions.

Example. Methods to develop and evaluate interventions.
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
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
## Process Science in Action

<table>
<thead>
<tr>
<th>Selection of Data Sources, Data Collection, and Event Log Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process Discovery</strong></td>
</tr>
<tr>
<td>Create “Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences</td>
</tr>
<tr>
<td><strong>Conformance Checking</strong></td>
</tr>
<tr>
<td>Process Monitoring</td>
</tr>
<tr>
<td><strong>Linking Data Sources and Contextualizing Events and Patterns</strong></td>
</tr>
<tr>
<td><strong>Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems</strong></td>
</tr>
<tr>
<td>Data-driven Tool Development</td>
</tr>
</tbody>
</table>

Leveraging Digital Trace Data to Investigate and Support Human-Centered Work Processes, ICEIS/ENASE Keynote on April 25th, 2023, Prague, Czech Republic
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}

Smart Factory equipped with sensors and actuators emitting events

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

Actuators:
- Motors
- Compressors
- Valves

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.

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.

Low

High
Process Observability Largely Differs

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

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

Tool interaction events during analysis (depending on the tool); large parts of the process occur in the reader’s mind.

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

Sensors and actuators emitting events.
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.

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

- Application logs (where available)
- Screen recordings to derive user interactions
- Think-aloud data
- Retrospective Interviews

Sensors and actuators emitting events:
- Eye-tracker
- Galvanic Skin Response
- Multiple Sensors
  - Proximity sensor
  - Presence sensor
  - Pressure sensor
  - Touch sensor
  - Flow sensor
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
Process-driven Execution and Collection of IoT Data in Context

- **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

- **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.
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**

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

WHO guidelines on drawing blood: best practices in phlebotomy

Low | High
**Process Activity / Event Detection from Sensors**

**Known process and activities**

- **Activity signatures to record IoT devices readings corresponding to events**

**Complex Event Processing queries**

**IoT to detect relevant state changes**

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.

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.
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.
Guiding Log Creation with Process Knowledge

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

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

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

TASK VIEW
Artifact

RESULT VIEW
Finding

DATA VIEW
Filtered Data

time
0 10 20 30

USER 1
VIEW PROFILE

USER 2
VIEW PROFILE

USER 3
VIEW PROFILE
Visualizing Event Sequences
Focus on Subsequences of Interest

USER 2
Tool: Celonis, Expertise: Good

TASK VIEW
Artifact

RESULT VIEW
Finding

DATA VIEW
Filtered Data

Other Tool
Activity Filter
Variants
Statistics
Activities
Process View
Original Log
Filtered Log
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

Process description

WHO guidelines on drawing blood:
best practices in phlebotomy

Next activity

Observed execution

Perform preliminary operations
Perform hand hygiene
Perform venipuncture
Monitor donor
Remove needle
Perform hand hygiene
Process Conformance Checking

Process description

WHO guidelines on drawing blood:
best practices in phlebotomy

Observed execution
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
[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..."
Providing Context to User Interaction Logs
Usage of Common Strategies

Process mining strategies derived from the analysis of interview data.

"Often I try to combine different tools to understand for sanity check if we have the same insights in different tools."
Associating a User’s Cognitive and Affective State With a Software Design Artifact

changes user’s cognitive and affective state

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

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

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

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

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

Detect nonconforming behavior during process execution.

WHO guidelines on drawing blood: best practices in phlebotomy.
Example of a recorded process mining analysis

<table>
<thead>
<tr>
<th>UId</th>
<th>Operation</th>
<th>I/O</th>
<th>Timestamp</th>
<th>User Annotations</th>
<th>Goals and Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_02</td>
<td>variantFilter(cases, keep, 75%)</td>
<td>L_2</td>
<td>07/10/22 10:01:18</td>
<td>filtered too much</td>
<td></td>
</tr>
<tr>
<td>R_01</td>
<td>nCases()</td>
<td>L_1</td>
<td>07/10/22 10:01:50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_02</td>
<td>variantFilter(cases, keep, 85%)</td>
<td>L_2</td>
<td>07/10/22 10:02:03</td>
<td>filtered too much</td>
<td></td>
</tr>
<tr>
<td>R_01</td>
<td>nCases()</td>
<td>L_1</td>
<td>07/10/22 10:02:32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_02</td>
<td>variantFilter(cases, keep, 90%)</td>
<td>L_2</td>
<td>07/10/22 10:03:11</td>
<td>good trade-off</td>
<td></td>
</tr>
</tbody>
</table>

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

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

University of Haifa

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

Supporting Software Maintenance With Psychophysiological Measures and Artifact Metrics

Universität St.Gallen

IPF Fellowship
Grant No.: 1031574
Thank You to My Team

Prof. Dr. Ronny Seiger
Assistant Professor

Thierry Sorg
PhD student

Dr. Francesca Zerbato
Senior Researcher

Dr. Amine Abbad Andaloussi
IPF Postdoctoral Fellow

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


- 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.

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 knowledge 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

<table>
<thead>
<tr>
<th>Phase</th>
<th>Goal</th>
<th>Exemplary Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discovery</strong></td>
<td>Capturing and describing processes</td>
<td>Techniques such as processes mining, to create <strong>representations of processes</strong> using <strong>digital event data</strong>; <strong>event-based architectures</strong> to organize <strong>data collection</strong> as well as computational methods to analyze the data and to <strong>identify patterns in processes</strong></td>
</tr>
<tr>
<td><strong>Explanation</strong></td>
<td>Understanding, why, how and when a process unfolds</td>
<td>Methods supporting sense-making around processes in a specific context, e.g., qualitative empirical research to study the <strong>context in which a pattern is situated</strong>. Leads to propositions or entire theories on <strong>cause effect relationships embedded in a situational context</strong></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Intervening and shaping the process into desired directions</td>
<td>Methods to <strong>develop and evaluate interventions</strong> 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.</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Selection of Data Sources, Data Collection, and Event Log Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Discovery</td>
</tr>
<tr>
<td>Create &quot;Current State“ Process Representations, Mine Behavior Pattern, Visualize Event Sequences, Create Augmented Representations</td>
</tr>
<tr>
<td>Conformance Checking</td>
</tr>
<tr>
<td>Process Monitoring</td>
</tr>
<tr>
<td>Linking Data Sources and Contextualizing Events and Patterns</td>
</tr>
<tr>
<td>(Supporting the identification of root-causes)</td>
</tr>
<tr>
<td>Interpretable (Bio-)Feedback, (Neuro-)Adaptive Software Systems</td>
</tr>
<tr>
<td>Data-driven Tool Development</td>
</tr>
</tbody>
</table>
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
(Tested 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

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

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

<table>
<thead>
<tr>
<th>ID</th>
<th>Tool Function</th>
<th>Tool</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>PDF Reader</td>
<td>Acrobat Reader</td>
<td>00:04:50,3</td>
<td>00:06:17,4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>Disco</td>
<td>00:09:38,3</td>
<td>00:11:09,9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>Disco</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>bupaR</td>
<td>00:14:00,7</td>
<td>00:15:09,9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>bupaR</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disco</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
</tbody>
</table>
**Biofeedback and Neuro-adaptive Software Systems**

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.

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

*Highlights of the mentally demanding parts of code to facilitate code review*

```java
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.getInstance().debug("Application started!");
    Logger.getInstance().debug("You have " + CPUIUtil.CPUAvailable() + " CPU(s) available");
}
```
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

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

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

Eye Mind: Process Model Augmented with Eye-tracking Metrics

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

Source: Prototype developed by Amine Abbad Andaloussi
Traditionally, not process-centric and focusses on specific tasks or decisions

Process Mining: At the Intersection of Data and Process Science

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

Source: van der Aalst & Carmona: Process Mining Handbook
Process Mining: The Big Picture

3 main types of process mining:
- discovery,
- conformance and
- enhancement

Source: van der Aalst et al.: Process Mining Manifesto
ERP and CRM Systems: Common Data Sources for Process Mining

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.

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}
Process Observability Largely Differs

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

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.

Equipped with sensors and actuators.

Low

High
Process Observability Largely Differs

Navigation, scrolling and zooming during model comprehension (depending on tool); large parts of the process occur in Tool interactions during analysis (depending on tool); large parts of the process occur in the analyst’s mind.

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

- Low
- High
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 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.

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}

Requires bridging the abstraction gap
Process Activity Detection from Sensors

### 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
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

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

---

*WHO guidelines on drawing blood:*
*best practices in phlebotomy*
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

**Process and activities** known, but not explicit; end-to-end visibility lacking; repetitive and well structured
Process-driven Execution and Collection of IoT Data in Context

- **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 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
Central goal is to gain a comprehensive understanding of the “process of process mining”

• 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

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

Two application domains
- Industry 4.0 and Industrial IoT
- Health care: clinical guidelines related to hygiene
**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

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

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

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

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

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, its TotalPaymentAmount, the cumulative amount paid by the offender, it's always initialized to zero. Ok. The amount paid by the offender in one transaction. This one is interesting.

<table>
<thead>
<tr>
<th>ID</th>
<th>Tool Function</th>
<th>Tool</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>PDF Reader</td>
<td>Acrobat Reader</td>
<td>00:04:50.3</td>
<td>00:06:17.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>Disco</td>
<td>00:09:38.3</td>
<td>00:11:09.9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>Disco</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>bupaR</td>
<td>00:14:00.7</td>
<td>00:15:09.9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>bupaR</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disco</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
</tbody>
</table>
Temporal Association of Events
Providing Context to Fixations

Challenge: Synchronization of events

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

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

Instantaneous 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.

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%.

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 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
Leveraging Digital Trace Data to Investigate and Support Human-Centered Work Processes, ICEIS/ENASE Keynote on April 25th, 2023, Prague, Czech Republic

Selected Examples

IoT-enabled business processes

Process

Intervention

Discovery

Explanation

Cognitive Processes

Selected Examples

Process of Process Mining

Universität St.Gallen
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 systems
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

Collected Data (After Some Preparation)

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Artifact</th>
<th>Timestamp (start)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBJ1</td>
<td>Task 1.1.0.md</td>
<td>2018-09-04T13:30:03.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBJ1</td>
<td>Kiosk.java</td>
<td>2018-09-04T13:36:05.015</td>
</tr>
<tr>
<td>SBJ1</td>
<td>Kiosk.java</td>
<td>2018-09-04T13:36:05.161</td>
</tr>
<tr>
<td>SBJ1</td>
<td>TravelCard.java</td>
<td>2018-09-04T13:36:17.192</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBJ1</td>
<td>Task 1.1.1.md</td>
<td>2018-09-04T13:38:02.015</td>
</tr>
<tr>
<td>SBJ1</td>
<td>Task 1.2.0.md</td>
<td>2018-09-04T13:38:58.043</td>
</tr>
</tbody>
</table>

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

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
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
Developers

...)

...
Artifact Types Used for Answering Domain Understanding Questions

Feature File

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

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

<table>
<thead>
<tr>
<th>Gaze ID</th>
<th>Timestamp</th>
<th>Gaze X Pos.</th>
<th>Gaze Y Pos.</th>
<th>Line of code</th>
<th>Column of code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15140</td>
<td>1450</td>
<td>60</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>15148</td>
<td>1430</td>
<td>77</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>15156</td>
<td>1480</td>
<td>77</td>
<td>18</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>15164</td>
<td>1530</td>
<td>86</td>
<td>19</td>
<td>25</td>
</tr>
</tbody>
</table>

Sample of gaze file with automatically mapped AOIs

Source: https://www.i-trace.org
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
Automated Mapping to Areas of Interest: The Case of Process Models

Automated mapping of gazes to process model elements

Instantaneous 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

Process structure is exploited to dynamically create AOIs with projections of behavioral and physiological measures.
Example: Hybrid Process Models
Stimulus with 3 Areas of Interest

AOI: Graph
AOI: simulation
AOI: law text

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

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

<table>
<thead>
<tr>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>04.03.2010 07:35</td>
<td>04.03.2010 07:46</td>
</tr>
<tr>
<td>04.03.2010 07:53</td>
<td>04.03.2010 07:55</td>
</tr>
<tr>
<td>08.03.2010 11:16</td>
<td>08.03.2010 11:18</td>
</tr>
<tr>
<td>09.03.2010 08:05</td>
<td>09.03.2010 08:10</td>
</tr>
<tr>
<td>11.03.2010 10:30</td>
<td>11.03.2010 10:32</td>
</tr>
<tr>
<td>11.03.2010 11:15</td>
<td>11.03.2010 11:19</td>
</tr>
<tr>
<td>11.03.2010 11:45</td>
<td>11.03.2010 11:52</td>
</tr>
<tr>
<td>14.03.2010 14:08</td>
<td>18.03.2010 08:04</td>
</tr>
<tr>
<td>14.03.2010 17:53</td>
<td>14.03.2010 17:56</td>
</tr>
<tr>
<td>18.03.2010 08:06</td>
<td>18.03.2010 08:07</td>
</tr>
<tr>
<td>18.03.2010 08:07</td>
<td>18.03.2010 08:08</td>
</tr>
<tr>
<td>18.03.2010 08:09</td>
<td>18.03.2010 08:09</td>
</tr>
</tbody>
</table>
What are common strategies used in the analysis stage?

Process mining strategies derived from the analysis of interview data.

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

<table>
<thead>
<tr>
<th>ID</th>
<th>Tool Function</th>
<th>Tool</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>PDF Reader</td>
<td>Acrobat Reader</td>
<td>00:04:50.3</td>
<td>00:06:17.4</td>
</tr>
<tr>
<td></td>
<td>Case Filter</td>
<td>Disco</td>
<td>00:09:38.3</td>
<td>00:11:09.9</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>Disco</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>Disco</td>
<td>00:14:00.7</td>
<td>00:15:09.9</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>bupaR</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>bupaR</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
<tr>
<td></td>
<td>Case Filter</td>
<td>bupaR</td>
<td>00:11:46.1</td>
<td>00:12:34.3</td>
</tr>
</tbody>
</table>

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..."
Creating User Interaction Logs

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, its 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.

<table>
<thead>
<tr>
<th>ID</th>
<th>Tool Function</th>
<th>Tool</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>Disco</td>
<td>00:09:38,3</td>
<td>00:11:09,9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>Disco</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter</td>
<td>bupaR</td>
<td>00:14:00,7</td>
<td>00:15:09,9</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>bupaR</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics</td>
<td>Disco</td>
<td>00:11:46,1</td>
<td>00:12:34,3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Exploration and Hypotheses Testing are Core Components of Process Mining

Work by Pnina Soffer and Her Team
Creation of Participant Profiles

**Participant 1**
Tools: Celonis

**Participant 2**
Tools: Disco, Celonis

**Participant 3**
Tools: Disco
Example: Hybrid Process Models
Fine-grained 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.

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 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%.

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.

- 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

Classical approach of assigning AOIs manually

Sample of user's gazes on the source-code

Each AOI refers to a distinct line of code
Automated Mapping to Areas of Interest: The Case of Source-code

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

Each AOI refers to a distinct line of code

- Inconsistent 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

Sample of user’s gazes on the source-code

```
public ResponseObject checkIn(TravelCard card) {
  if (!card.isCheckInStatus()) {
    if (hasEnoughBalance(card)) {
      card.setCheckInStatus(true);
      response = new ResponseObject(200, Constants.CHECKED_IN_SUCCESS);
    }
    InitSystem.isl.getLogger().
    .info("CHECKIN: Automaton at " + staticName + ": " +
    InitSystem.isl.printLog();
    countCheckIn++;
  } else {
    response = new ResponseObject(220, Constants.CHECKED_IN_FAILURE);
  }
  } else {
    response = new ResponseObject(210, Constants.CHECKED_IN_FAILURE_ALL);
  }
  return response;
```

Sample of gaze file with automatically mapped AOIs

<table>
<thead>
<tr>
<th>Gaze ID</th>
<th>Timestamp</th>
<th>Gaze X Pos.</th>
<th>Gaze Y Pos.</th>
<th>Line of code</th>
<th>Column of code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15140</td>
<td>1450</td>
<td>60</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>15148</td>
<td>1430</td>
<td>77</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>15156</td>
<td>1480</td>
<td>77</td>
<td>18</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>15164</td>
<td>1530</td>
<td>86</td>
<td>19</td>
<td>25</td>
</tr>
</tbody>
</table>

Source: https://www.i-trace.org
• 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

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

Top down
- IRB2600 Door to Scanner
- IRB2600 Scan Gripper
- IRB2600 Scanner to Door
- IRB2600 Door to Portal
- IRB2600 To Tray Open
The Process of Reading Source Code

Developer’s Eye Gazes

Physiological sensors

changes

user’s cognitive and affective state

Developer’s cognitive and affective state while looking at the highlighted part of the source code

High CL

s sense of, creates, modifies

Software Design Artifacts

user’s cognitive and affective state

Changes
The Process of Reading Source Code

Software Design Artifacts

changes

measure

Biosensors

user's cognitive and affective state

Reads, makes sense of, creates, modifies

```java
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();
    Object r = new Rectangle();
    array.add(r);
    Object t = new Triangle();
    Object c = new Circle();
    array.add(t);
    array.add(c);
    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;
}
```
Process Activity Detection from Sensors

Event sequences of all sensors and actuators of smart factory

Process event sequences

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 Processes

Our Goal

Using a process science lens

And digital trace data
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

- Object-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
Process of Process Mining

IoT-enabled processes

Process

Explanation

Discovery

Intervention

Selected Examples

Cognitive Processes
1. Unravelling Cognitive Processes

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

```java
public static void main(String[] args) {
    List<Object> array = new ArrayList();
    array.add(3);
    Object x = new Triangle();
    Object y = new Circle();
    array.add(x);
    array.add(y);
    for (int i = 0; i < array.size(); i++) {
        if (array.get(i) instanceof Triangle) return array.size();
        else if (array.get(i) instanceof Circle) return i + 2;
        return i + 4;
    }
}

public static int next(int i, List<Object> array) {
    return i + 1;
}
```

Biosensors

Developer’s cognitive and affective state while looking at the highlighted part of the source code.

High CL.

Developer’s Eye Gazes

Physiological sensors

Pupil dilation

Time (seconds)
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

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

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
Mapping of Events with Artifact

To which part of the artifact 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

Process structure is exploited to dynamically create AOIs with projections of behavioral and physiological measures.
Leveraging Artifact Information for Dynamic AOI Creation

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

![Diagram of code snippet with fixation events and artifact syntax & semantics]

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

```java
public static void main(String[] args) {
    List<Object> array = new ArrayList<Object>();
    Object r = new Rectangle();
    array.add(r);
    Object e = new Ellipse();
    Object s = new Circle();
    array.add(e);
    array.add(s);
    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;
}
```
1. Thierry Related Work
Areas of Interest
Grouping and Aggregation of Fixation Events

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

Sequence of AOIs over time
Contextualization of Fixation Events
Leveraging Artifact Metrics & Features

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

---

Fixation x, y coordinate, start, duration, **source code element**
Contextualization of Fixation Events
Leveraging Artifact and Physiological & Behavioral Features

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

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
Collection of Interaction Data and Verbal Data

Raw Data
Interaction Data: Screen recordings

Verbal Data

Challenge: Synchronization of events

Interaction Data: Application level logs

Challenge: Raw events have different abstraction levels
Create the 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

Collection of Interaction Data and Verbal Data

Raw Data

Interaction Data: Screen recordings

Verbal Data

Challenge: Synchronization of events

Challenge: Raw events have different abstraction levels
User Behavior Log (Interaction Data)

Event sequence showing tool usage over time

<table>
<thead>
<tr>
<th>Participant</th>
<th>Event Description</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Desktop view started</td>
<td>0:07:50,3</td>
</tr>
<tr>
<td>P1</td>
<td>Desktop view ended</td>
<td>0:08:13,4</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Diagram showing sequence of tool usage with timestamps:
Think aloud data helps to interpret interaction data

"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..."
Analyzing User Behavior Logs

Challenge: How can we analyze event sequences?

Event Sequence Data (Tool Perspective)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Event</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>PDF Reader START</td>
<td>00:07:50,3</td>
</tr>
<tr>
<td>P27</td>
<td>PDF Reader END</td>
<td>00:08:13,4</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter START</td>
<td>00:08:14,4</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter END</td>
<td>00:08:48,7</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics START</td>
<td>00:08:50,0</td>
</tr>
<tr>
<td>P27</td>
<td>Statistics END</td>
<td>00:11:18,1</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter START</td>
<td>00:11:30,6</td>
</tr>
<tr>
<td>P27</td>
<td>Case Filter END</td>
<td>00:11:49,8</td>
</tr>
</tbody>
</table>

Evolution of tool usage over time

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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Event</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Desktop view started</td>
<td>0:07:50,3</td>
</tr>
<tr>
<td>P1</td>
<td>Desktop view ended</td>
<td>0:08:13,4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Hierarchical Cluster Analysis

Patterns can emerge within one user or across multiple users.

Sequential Pattern Mining Analysis

Patterns (Support = 40%)

Length 2    Length 3

Correlation matrix

Multiple sequence alignment

Sankey diagram

Dendogram

Scarf plot
"And often I try to combine different tools to understand for sanity check if we have the same insights in different tools." (Interview Data)
Event Sequence Analysis and Visualization

Coping with Event Sequence

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

Human in the
(Process) Mines

July 2-7, 2023
Dagstuhl Seminar 2371

Abstract—The growing volume of event sequences is needed for big data applications. In the case of temporal event data, challenges of users of visual analytics tools include a reduction of the data volume and the complexity of the data. This contribution provides an overview of 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.

• **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

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

Record about past occurrence
- Log, Stream, Event Sequence

\[
\begin{align*}
e_1 &\rightarrow e_2 &\rightarrow e_3 &\rightarrow e_4 &\rightarrow e_5 &\rightarrow e_6 &\rightarrow e_7 \\
\end{align*}
\]
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 occurrence
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
The Many Flavors of Event Sequences
CAiSE 2016-2021

Key role of event sequences in 52 papers

- 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)

Observations

- 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

Observation

Information Systems, Sensors

Occurrence (Events)

Record about past occurrence

Log, Event Sequence

Events from a Complex Event Processing Lense

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

Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)
### The Many Flavors of Events

**Events seen with a Complex Event Processing lense**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw event</td>
<td>An event object that records a real-world event.</td>
</tr>
<tr>
<td>Instantaneous event</td>
<td>An event that happens at a point in time.</td>
</tr>
<tr>
<td>Simple event</td>
<td>An event that is not viewed as summarizing, representing, or denoting a set of other events.</td>
</tr>
<tr>
<td>Complex event</td>
<td>An event that summarizes, represents, or denotes a set of other events.</td>
</tr>
<tr>
<td>Derived event</td>
<td>An event that is generated as a result of applying a method or process to one or more other events.</td>
</tr>
<tr>
<td>Composite event</td>
<td>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.</td>
</tr>
<tr>
<td>Virtual Event</td>
<td>An event that does not happen in the real world, but is imagined, modeled, or simulated.</td>
</tr>
</tbody>
</table>

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

Occurrence (Events)
e1 → e2 → e3 → e4 → e5 → e6 → e7

Record about past occurrence
Log, Stream, Event Sequence

Immutable Ordered Replayable (desirable)
(Keyed or unkeyed)
The Many Flavors of Events
Event-first Thinking

<table>
<thead>
<tr>
<th>Atomic</th>
<th>Something happened (bid on an item, device temperature).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related</td>
<td>A stream or sequence of events (tracking a price change, device metrics changes over time).</td>
</tr>
<tr>
<td>Behavioral</td>
<td>The accumulation of facts capture behavior. Collecting, remembering, and analyzing the facts allows us to recognize and react to behavior.</td>
</tr>
</tbody>
</table>

Event Processing and Stream Processing

Event Source → Event → Event-at-a-time processing → Event Stream → Table → Event-at-a-time processing → Event Sink

- Event-at-a-time processing
  - Filtering
  - Branching
  - Translation
  - Windowing
  - Aggregation
  - Grouping

Stream processing

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

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.

Complex event recognition in the Big Data era: a survey

Nikos Giatrakos\textsuperscript{1,2}, Elias Alevizos\textsuperscript{3,4}, Alexander Artikis\textsuperscript{4,5}, Antonios Deligiannakis\textsuperscript{1,2}, Minos Garofalakis\textsuperscript{1,2}

Received: 3 January 2019 / Revised: 12 April 2019 / Accepted: 8 July 2019 / Published online: 25 July 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 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


**Events: Definition and Function**

<table>
<thead>
<tr>
<th>Domain Driven Design and Microservices</th>
<th>Event-first Thinking</th>
<th>Events from a Complex Event Processing Lense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Events</td>
<td>Atomic</td>
<td>Raw events</td>
</tr>
<tr>
<td>Integration Events</td>
<td>Related</td>
<td>Instantaneous events</td>
</tr>
<tr>
<td></td>
<td>Behavioral</td>
<td>Simple events</td>
</tr>
</tbody>
</table>

**Observation**
Information Systems, Sensors

**Occurrence (Events)**

- e1 → e2 → e3 → e4 → e5 → e6 → e7

**Record about past occurrence**
Log, Stream, Event Sequence

### The Many Flavors of Events

**Domain Driven Design and Microservices**

| Domain events | In domain-driven design **domain events** describe something that has occurred 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 **integration 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

Occurrence (Events)

Key drivers for event-driven architectures:
• Building decoupled systems
• Desire for team autonomy

Record about past occurrence
Log, Stream, Event Sequence

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

Data Ingestion & Integration (2)

Event handling (8)
- Impact analysis on effect
- Reactive event handling
- Proactive event handling
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

Record about past occurrence
Log, Stream, Event Sequence

Occurrence (Events)

Roles of event sequences in Process Science

Studying Continuous Change

Immutable
Ordered
Replayable (desirable)
(Keyed or unkeyed)