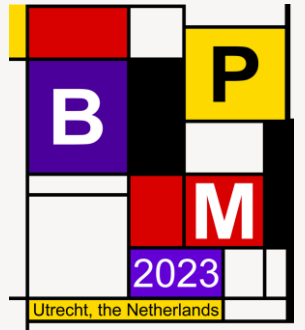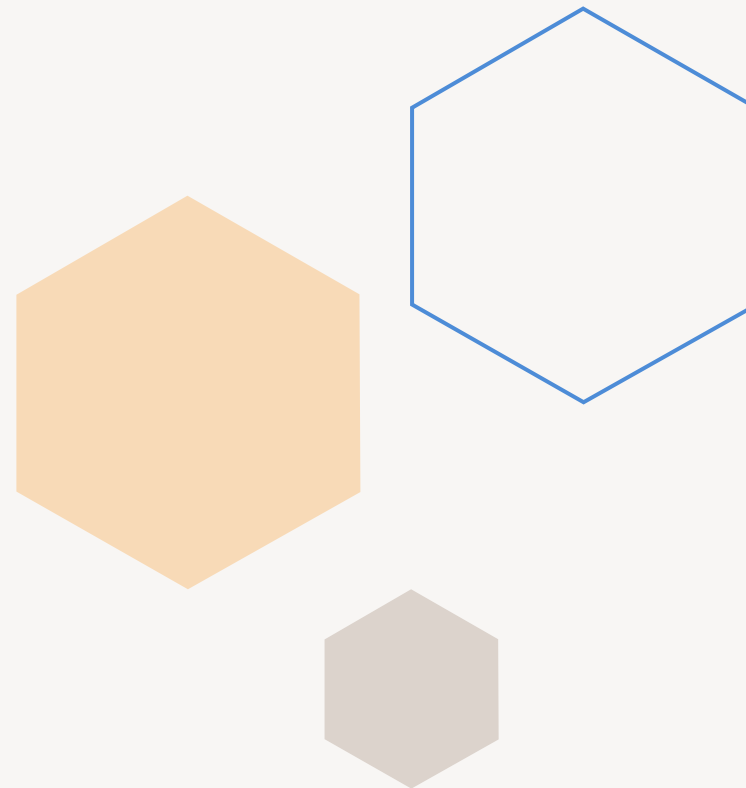# Task Support for Process Mining

## From formulating questions to evaluating results

21st BPM Conference
Utrecht, 13 September 2023

# Meet the Organisers



**Iris Beerepoot**

Utrecht University



**Francesca Zerbato**

University of St. Gallen



**Barbara Weber**

University of St. Gallen

.



**Pnina Soffer**

University of Haifa

## Advisory Group

- Xixi Lu, Utrecht University
- Niels Martin, Hasselt University

- Vinicius Stein Dani, Utrecht University
- Lisa Zimmermann, University of St. Gallen

Outline

Project Lifecycle

Define Questions

Data Collection & Preparation

Mining & Analysis

Results

# Aims and Scope

Existing methodologies provide *high-level* guidance

▼

Task Support for Process Mining

▲

Case studies focus on results obtained in specific settings

- Overview of existing support for Process Mining tasks

- Real scenario walkthrough

- Pointers to relevant literature and open discussion

# Our Focus

*Organizational and Ecosystem* level

**Individual and Group levels**
look at the work practices and behavior of individuals and interactions between them

*Technical* level

vom Brocke, J., Jans, M., Mendling, J., & Reijers, H. A. (2021). A five-level framework for research on process mining. *Business & Information Systems Engineering*, 1-8.

# The Process Mining Lifecycle

| Define Questions | Data Collection & Preparation | Mining & Analysis | Results |

Emamjome, F., Andrews, R., & ter Hofstede, A. H. (2019). A case study lens on process mining in practice. In *On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21–25, 2019, Proceedings* (pp. 127-145). Springer International Publishing.

# PM²: A Process Mining Project Methodology
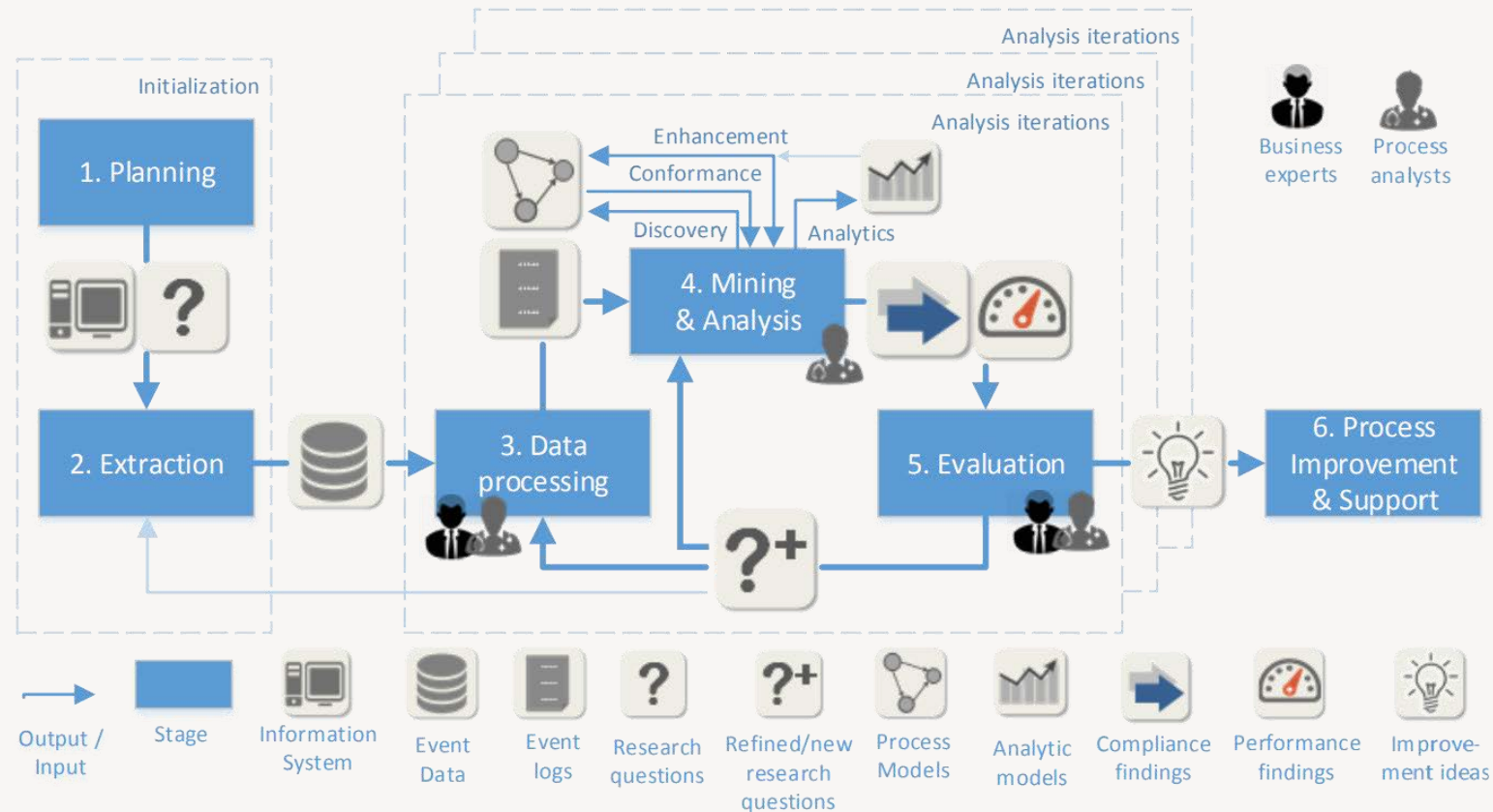


Van Eck, M. L., Lu, X., Leemans, S. J., & Van Der Aalst, W. M. (2015, May). PM: a process mining project methodology. In *International conference on advanced information systems engineering* (pp. 297-313). Cham: Springer International Publishing.

# Our Scenario

K&B Landscape Architecture and Urbanism

45 employees

Project-focused

No experience with process mining

What can process mining do for us?

Define Questions

# Define Questions

**Goal**
Specify the scope and the objectives of the project and formulate questions.

Questions "should be":
- Related to a specific process
- Answerable using event data
- Concrete

Scope
Appendectomy process, from hospital admission to post-operative care.

Objective
Transparency; Process path understanding and variability.

Questions
- What are common paths that patients take during the process?
- Are there any variations in the paths taken for different patient groups?

The question and its type help to choose the analysis approach and the techniques and tools to use.

Gurgen Erdogan, T., & Tarhan, A. (2018). A goal-driven evaluation method based on process mining for healthcare processes. *Applied Sciences*, *8*(6), 894.

# What if we don't have Questions?

"Sometimes coming up with <u>good</u> questions <u>at the start</u> of a project is <u>difficult.</u>"

"It is often <u>very hard</u> to identify the <u>correct</u> question."

"A data-driven project is powered by the <u>availability of event data</u>. There is no concrete question or goal…"

Our scenario

K&B design is interested in exploring the use of process mining to understanding the way people work.

Scope
Working patterns of people.

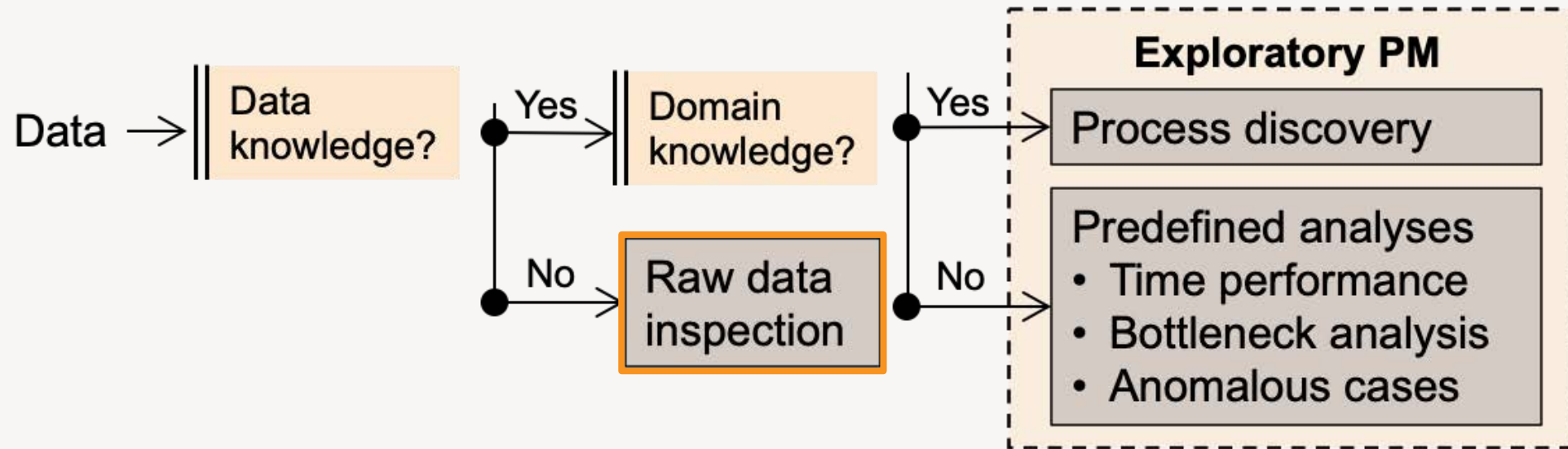Goal
Process understanding; Transparency; Improvement.

Questions
…?

# Question Formulation using Process Mining

Process mining analysis can help analysts to formulate questions.

**Steps.**



**Goals.**
- Assess what analyses can be done on the provided data;
- Generate data-driven insights and hypotheses that can inspire questions.

Zerbato, F., Koorn, J. J., Beerepoot, I., Weber, B., & Reijers, H. A. (2022, September). On the Origin of Questions in Process Mining Projects. In *International Conference on Enterprise Design, Operations, and Computing* (pp. 165-181). Cham: Springer International Publishing.

# Question Formulation using Process Mining

**Raw data inspection**

**Strategy.** Understand the data to learn about its structure, how it is formatted, what process steps are recorded and what data attributes are available.

*Rationale*

- Planning: Look into what can be analyzed and how
- Profiling: Check the data quality

(?) What kinds of questions can be asked based on the data available in the organization?

Data → Data knowledge? → Yes → ...
→ No → Raw data inspection
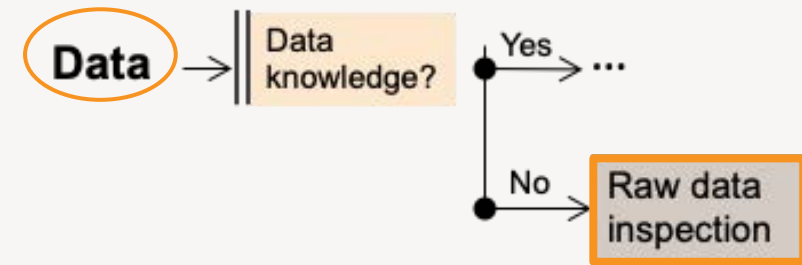
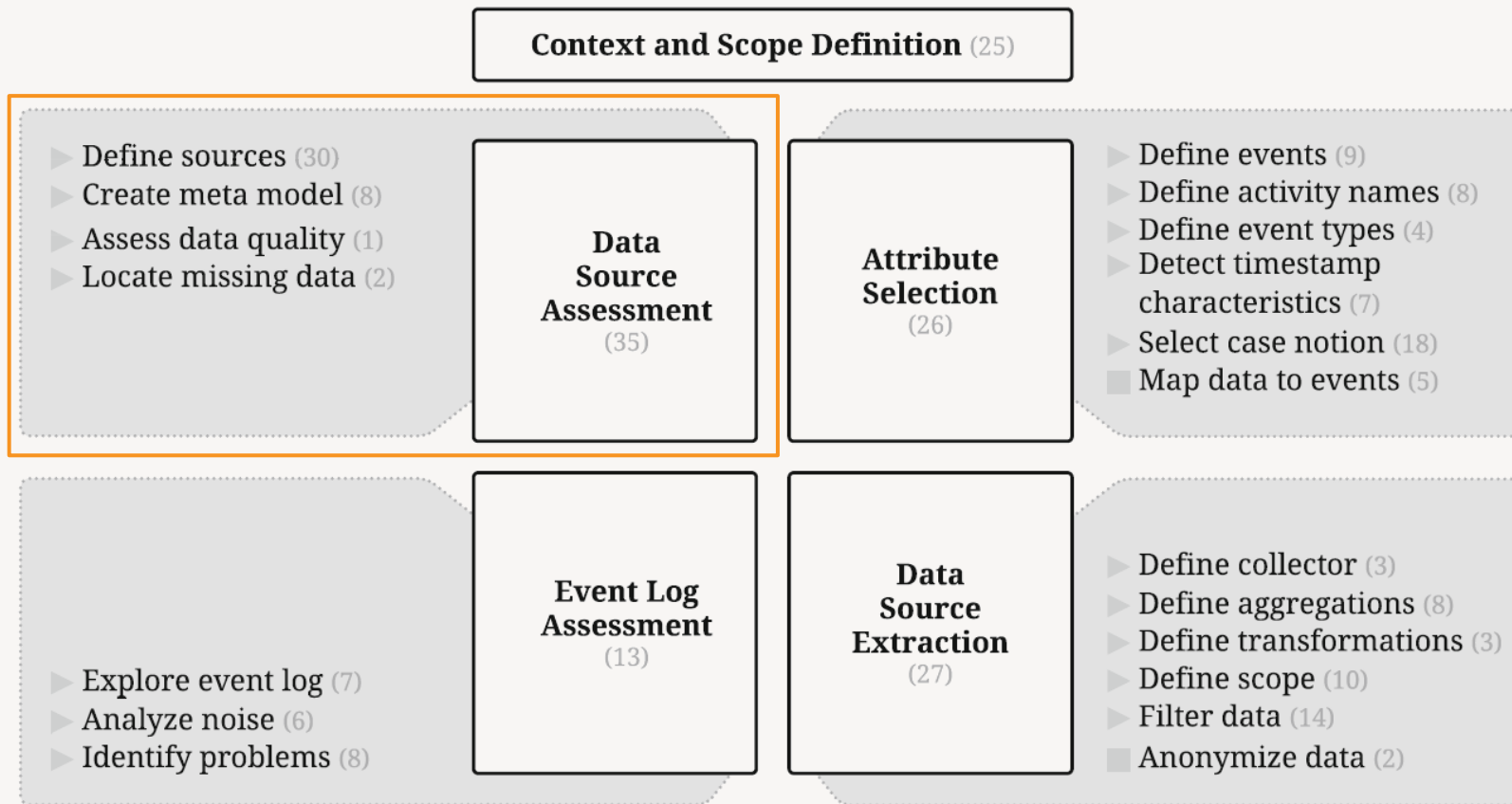We need to first extract the data!

→ Data Collection & Preparation

# Mapping Processes to Data
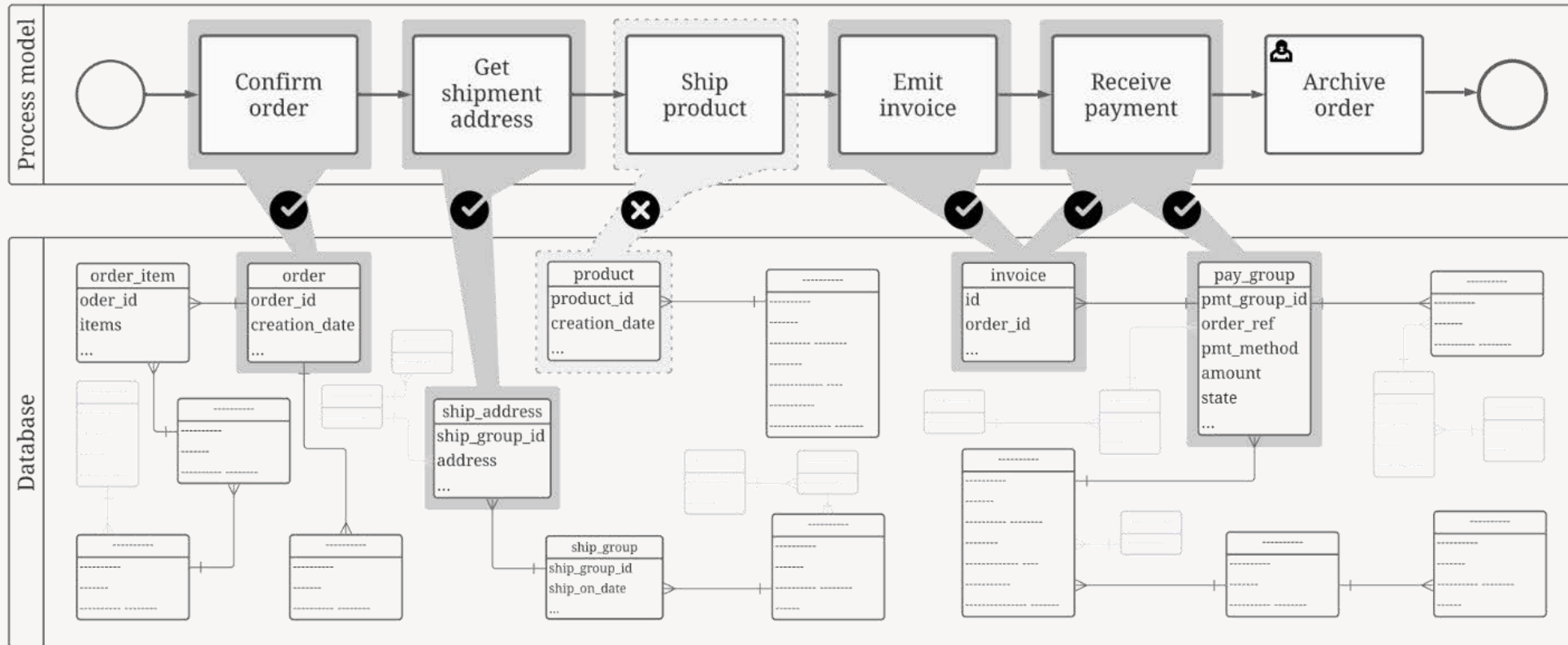


Stein Dani, V., Leopold, H., van der Werf, J. M. E., & Reijers, H. A. (2022, September). Supporting Event Log Extraction Based on Matching. In *International Conference on Business Process Management* (pp. 322-333). Cham: Springer International Publishing.

# Manual Tasks in Event Log Extraction

**Context and Scope Definition** (25)

- ▶ Define sources (30)
- ▶ Create meta model (8)
- ▶ Assess data quality (1)
- ▶ Locate missing data (2)

**Data Source Assessment** (35)

**Attribute Selection** (26)

- ▶ Define events (9)
- ▶ Define activity names (8)
- ▶ Define event types (4)
- ▶ Detect timestamp characteristics (7)
- ▶ Select case notion (18)
- ▪ Map data to events (5)

**Event Log Assessment** (13)

- ▶ Explore event log (7)
- ▶ Analyze noise (6)
- ▶ Identify problems (8)

**Data Source Extraction** (27)

- ▶ Define collector (3)
- ▶ Define aggregations (8)
- ▶ Define transformations (3)
- ▶ Define scope (10)
- ▶ Filter data (14)
- ▪ Anonymize data (2)

Stein Dani, V., Leopold, H., van der Werf, J. M. E., Lu, X., Beerepoot, I., Koorn, J. J., & Reijers, H. A. (2021, September). Towards understanding the role of the human in event log extraction. In *International Conference on Business Process Management* (pp. 86-98). Cham: Springer International Publishing.
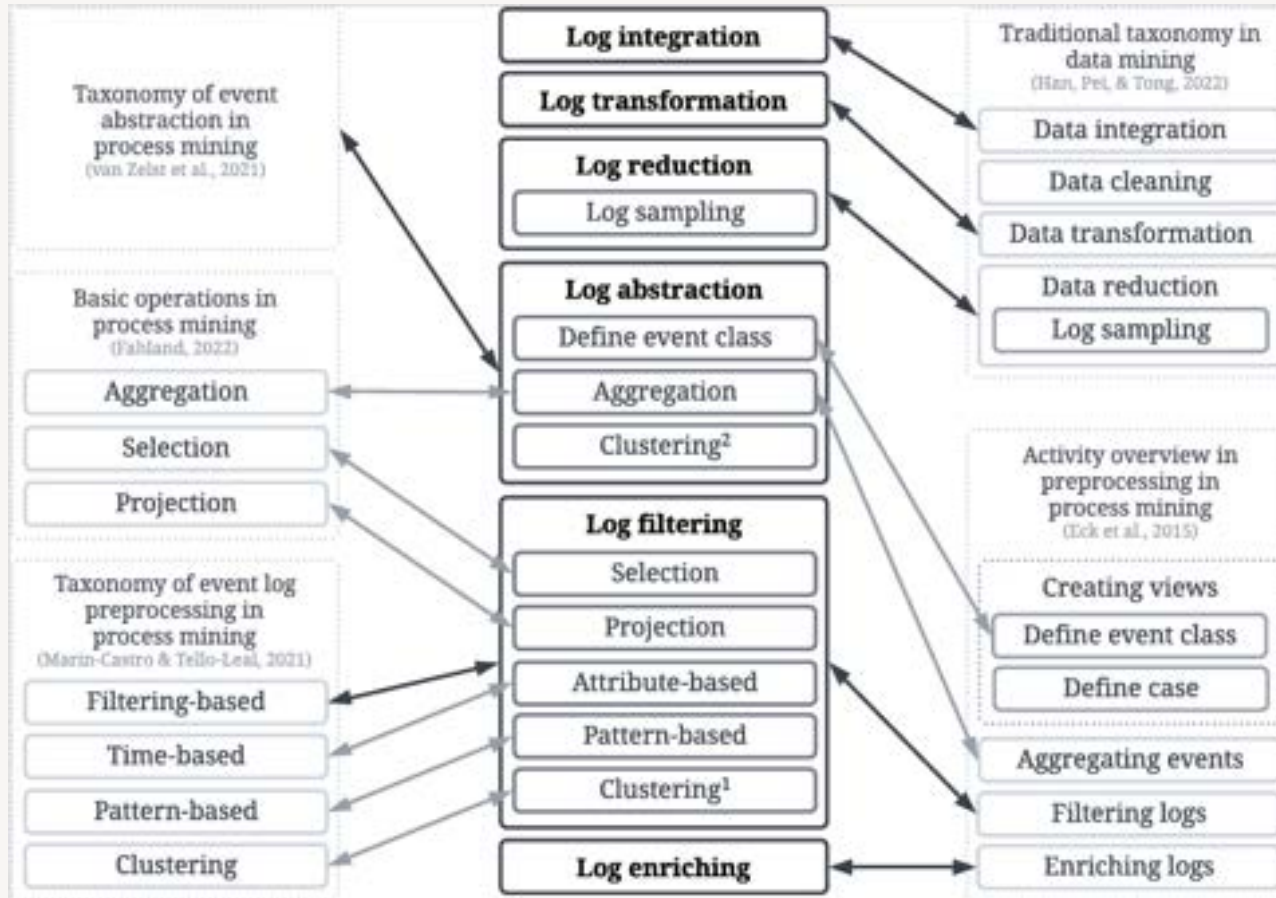
# Data Source Assessment

**AUTODESK**
AutoCAD

**AUTODESK** Revit

Until 2019, a system called PSO was used for registering hours!

**AFAS** software

# Event Pre-processing



Liu, Y., Stein Dani, V., Beerepoot, I., Lu, X. (forthcoming).

# Log Integration and Transformation

**PSO**

| Date | Employee name | Project | Item | | Number of hours |
|------|---------------|---------|------|--|----------------|
| 1-3-2016 | Beerepoot, I.M. | 100.00 | 220 Overleg/presentatie onderzoek | | 5 |

**AFAS**

| Date | Employee name | Project | Project phase | Number of hours |
|------|---------------|---------|---------------|-----------------|
| 1-2-2021 | Iris Beerepoot | 1428.00 | Overleg | 3 |

!

| Date | Employee name | Project | Activity | Number of hours |
|------|---------------|---------|----------|-----------------|
| 1-3-2016 | Beerepoot | 100.00 | 220 Overleg/presentatie onderzoek | 5 |
| 1-2-2021 | Beerepoot | 1428.00 | Overleg | 3 |

**Integrated log**

Define questions

Define
Questions
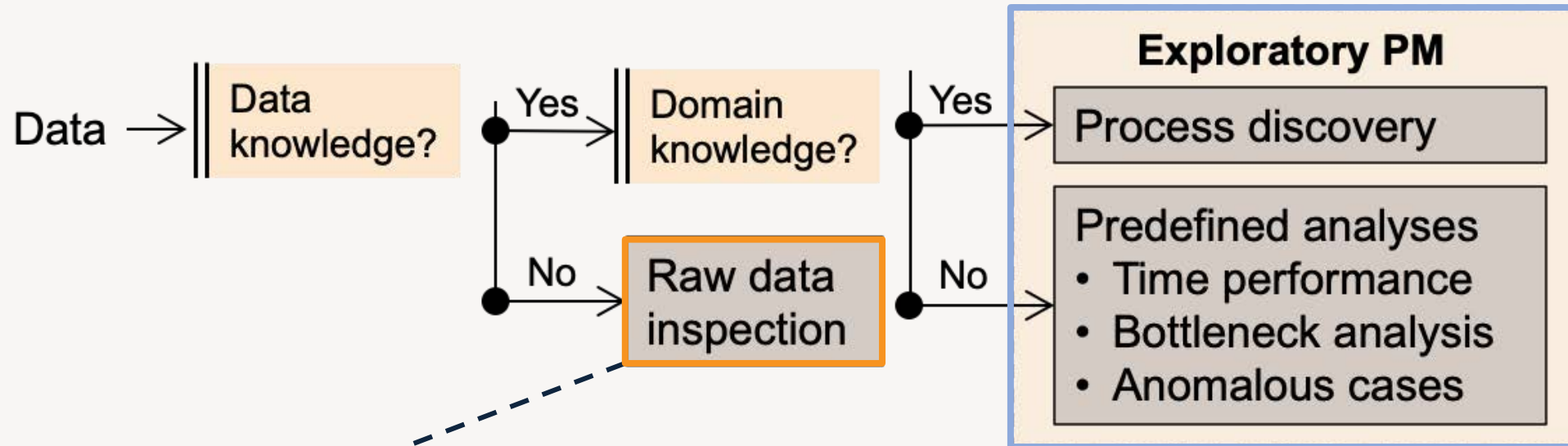
Data
Collection &
Preparation

Mining &
Analysis

Results

# Define Questions

# Question Formulation with Exploratory PM

Data → Data knowledge? —Yes→ Domain knowledge? —Yes→ **Exploratory PM**: Process discovery

Data knowledge? —No→ Raw data inspection

Domain knowledge? —No→ Predefined analyses
- Time performance
- Bottleneck analysis
- Anomalous cases

Focus on projects
Q1. What is the general order of activities across projects?

# Let Exploratory Analysis Generate Questions

*Second Iteration*

**Strategy:** Define questions that capture stakeholders' needs; use standard hypotheses and templates to keep it focused and interactive

**Strategy:** Map the questions to the data to identify concrete entry points for the analysis

*Rationale*
- Identify concrete entry points for the analysis
- Involve stakeholders in question formulation
  - Bring visualizations and descriptives to the table
- Generate hypotheses collaboratively

Mining & Analysis

Define
Questions

Data
Collection &
Preparation

Mining &
Analysis

Results
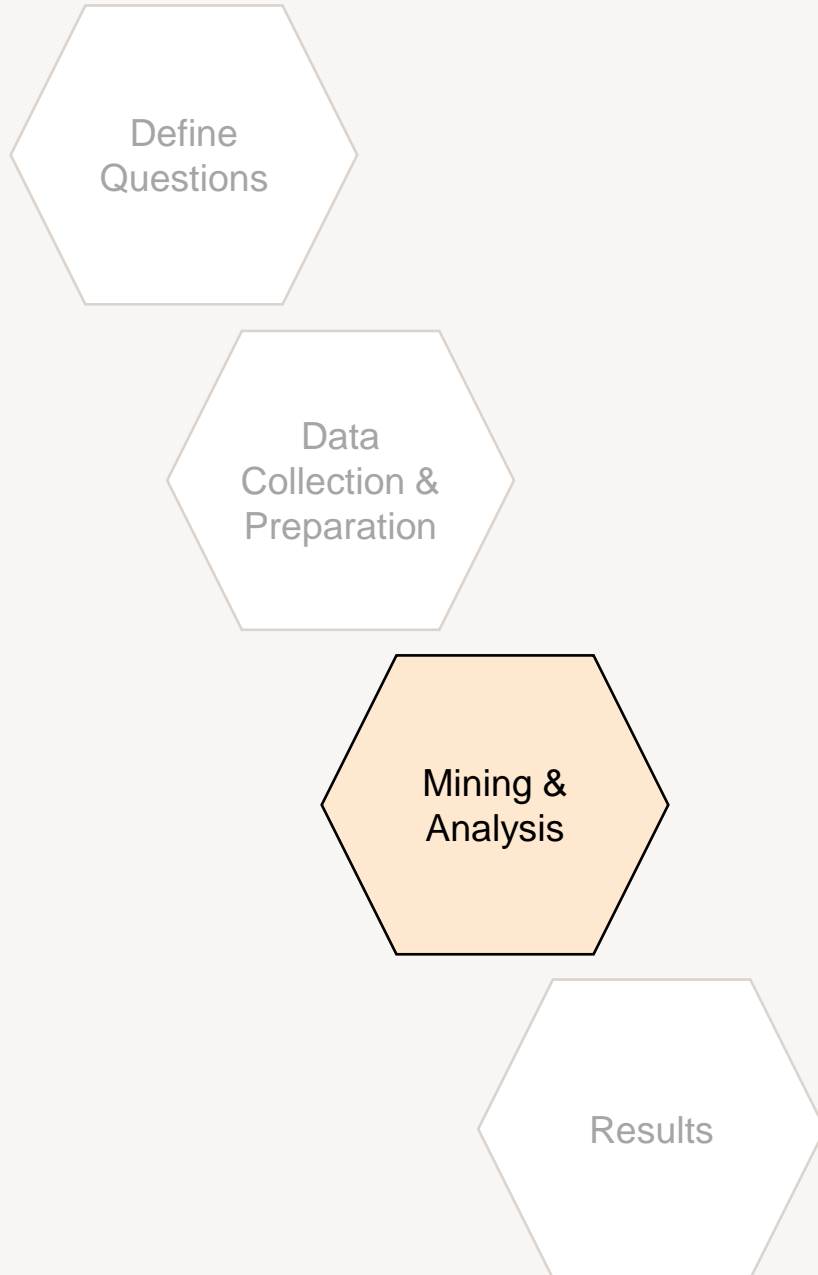
# Mining & Analysis

# Mining & Analysis

**Goal**
Apply process mining techniques to the data prepared to answer the questions and obtain process-related insights.

The analysis:

- Depends on the question
- Requires analysts to produce artifacts and (intermediate) results that are consumed in later analysis stages or evaluated

**Produce Artifacts and Results**
- Apply process mining techniques
- Create artifact and visualizations
- …

**Consume Artifacts and Results**
- Interpret the results based on the questions and the context
- …

**Plan and Organize**
- Choose the analysis approach
- Choose the tool
- Prioritize analysis directions
- Track and document the analysis
- …

Planning activities help to keep the analysis focus.

# Exploratory Analysis

**Strategy:** Understand the process to get familiar with the control flow in light of the question.
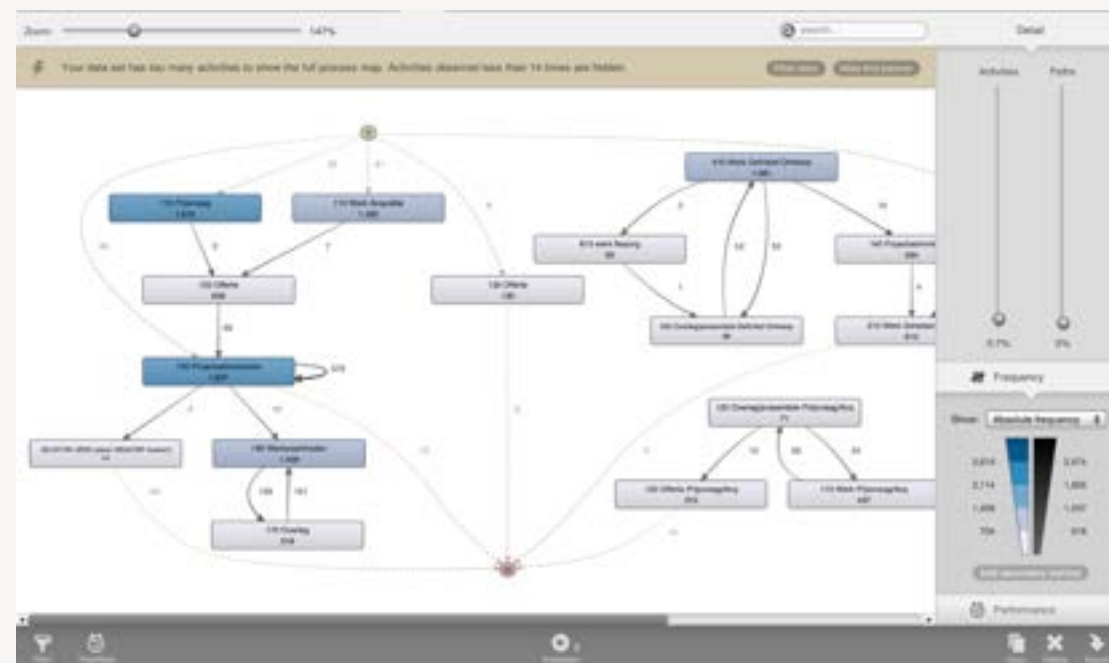
*Rationale*

- Characterize: Assess whether the data is suitable to answer the question

- Confirm: Check domain understanding

💡 Process understanding is often based on visuals and descriptives; it can be done collaboratively.

Q1: What is the general order of activities across projects?
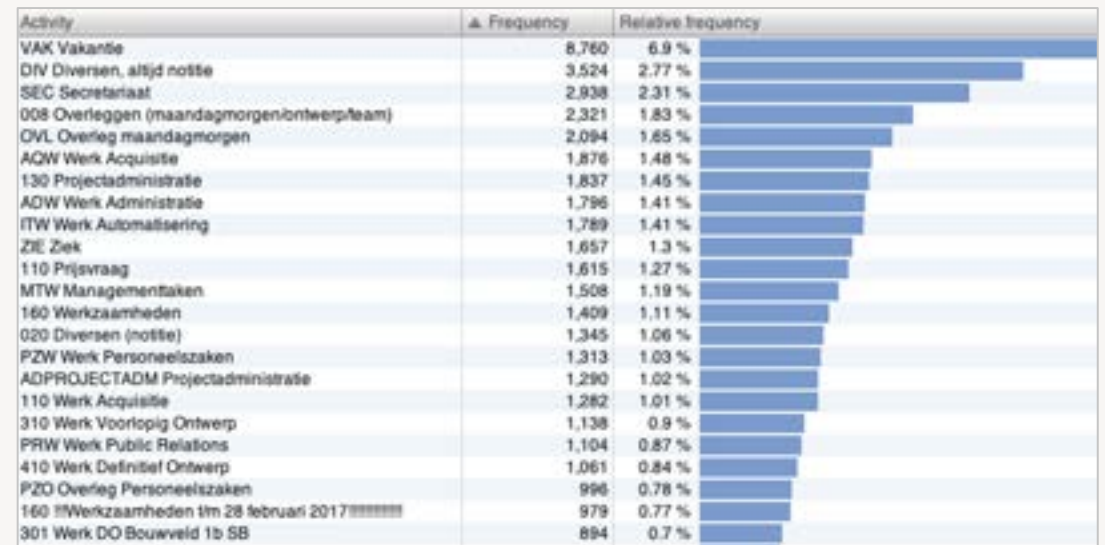
# Exploratory Analysis: Outcomes

## Questions

Q1: What is the general order of activities across projects?

Q2. What is the average duration of projects? What are projects with long lead times?

Q3. Are resources involved in multiple projects? What are handoffs patterns between resources across projects?

## Observations

| Activity | ▲ Frequency | Relative frequency |
|---|---|---|
| VAK Vakantie | 8,760 | 6.9 % |
| DIV Diversen, altijd notitie | 3,524 | 2.77 % |
| SEC Secretariaat | 2,938 | 2.31 % |
| 008 Overleggen (maandagmorgen/ontwerp/team) | 2,321 | 1.83 % |
| OVL Overleg maandagmorgen | 2,094 | 1.65 % |
| AQW Werk Acquisitie | 1,876 | 1.48 % |
| 130 Projectadministratie | 1,837 | 1.45 % |
| ADW Werk Administratie | 1,796 | 1.41 % |
| ITW Werk Automatisering | 1,789 | 1.41 % |
| ZIE Ziek | 1,657 | 1.3 % |
| 110 Prijsvraag | 1,615 | 1.27 % |
| MTW Managementtaken | 1,508 | 1.19 % |
| 160 Werkzaamheden | 1,409 | 1.11 % |
| 020 Diversen (notitie) | 1,345 | 1.06 % |
| PZW Werk Personeelszaken | 1,313 | 1.03 % |
| ADPROJECTADM Projectadministratie | 1,290 | 1.02 % |
| 110 Werk Acquisitie | 1,282 | 1.01 % |
| 310 Werk Voorlopig Ontwerp | 1,138 | 0.9 % |
| PRW Werk Public Relations | 1,104 | 0.87 % |
| 410 Werk Definitief Ontwerp | 1,061 | 0.84 % |
| PZO Overleg Personeelszaken | 996 | 0.78 % |
| 160 !!Werkzaamheden t/m 28 februari 2017!!!!!!!!!! | 979 | 0.77 % |
| 301 Werk DO Bouwveld 1b SB | 894 | 0.7 % |

- The data is highly variable
- A lot of time is allocated to activities that do not belong to "real" projects

**Data Collection & Preparation**
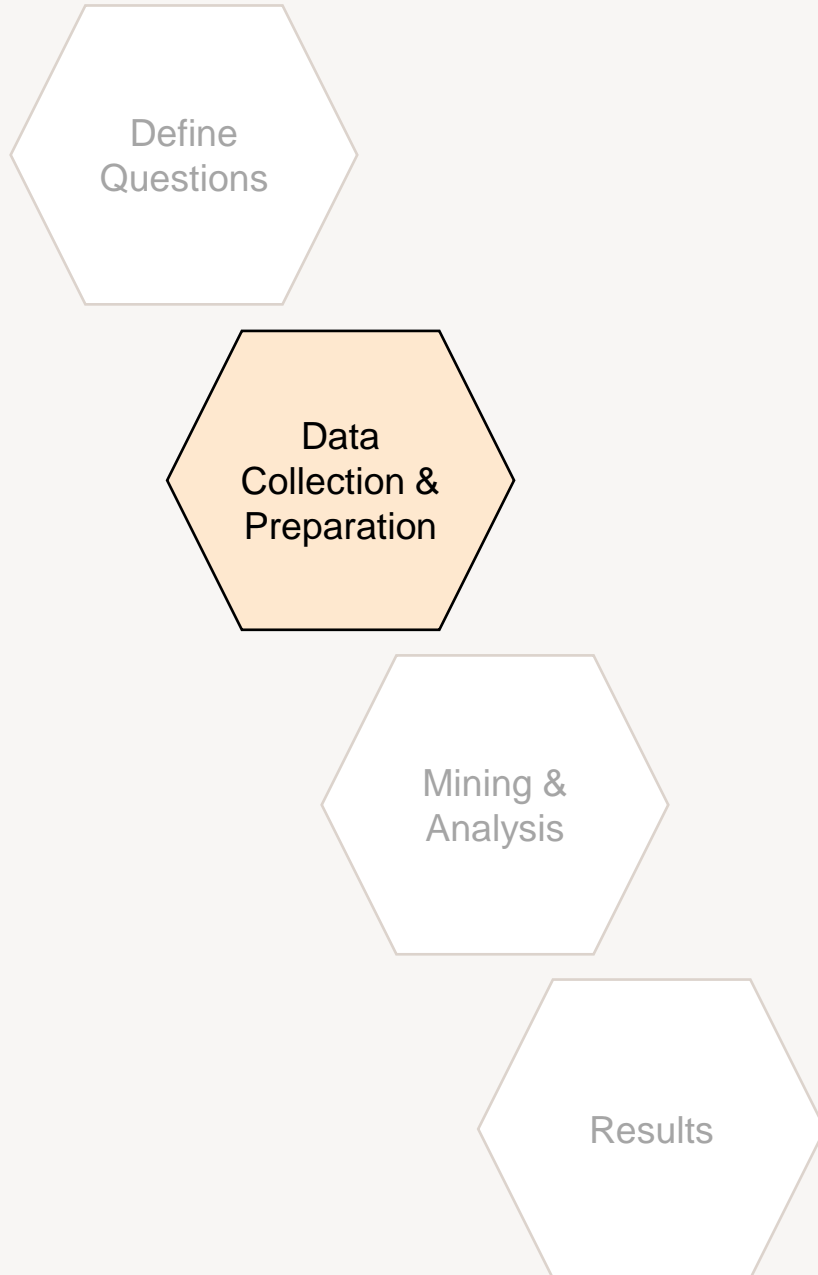
Data Collection &
Preparation
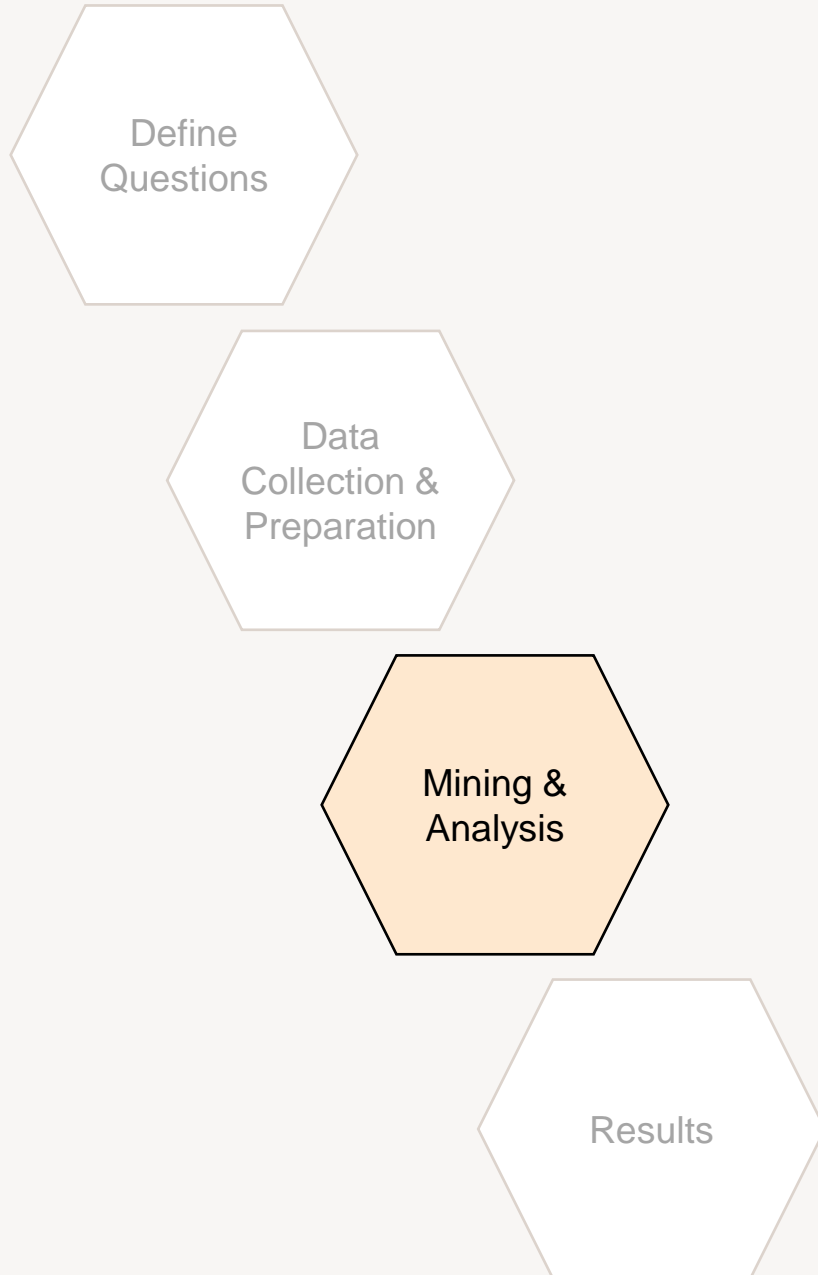
# Log Transformation and Abstraction

**Integrated log**

| Date | Employee name | Project | Activity | Number of hours |
|------|---------------|---------|----------|-----------------|
| 1-3-2016 | Beerepoot | 100.00 | 220 Overleg/presentatie onderzoek | 5 |
| 1-2-2021 | Beerepoot | 1428.00 | Overleg | 3 |

✓ Remove capitalization
✓ Merge projects (domain knowledge)
✓ Abstract activities into higher-level ones (domain knowledge)

**New log**

| Date | Employee name | Project | Activity | Number of hours |
|------|---------------|---------|----------|-----------------|
| 1-3-2016 | Beerepoot | 100 | overleg | 5 |
| 1-2-2021 | Beerepoot | 1428 | overleg | 3 |

Mining & Analysis

Define
Questions

Data
Collection &
Preparation

Mining &
Analysis

Results

# Mining & Analysis

# A General Overview

Q1. What is the general order of activities across projects?

Sketch

Draft design

Final design

# Pattern Discovery

**Strategy:** Discover patterns in the data or relationships among observed phenomena or scenarios.

*Rationale*

- Find interesting things: keep an eye open for aspects that stand out

- Generate hypotheses

Q2. What is the average duration of projects? What are projects with long lead times?

Zerbato, F., Soffer, P., & Weber, B. (2022, September). Process mining practices: evidence from interviews. In *International Conference on Business Process Management* (pp. 268-285). Cham: Springer International Publishing.
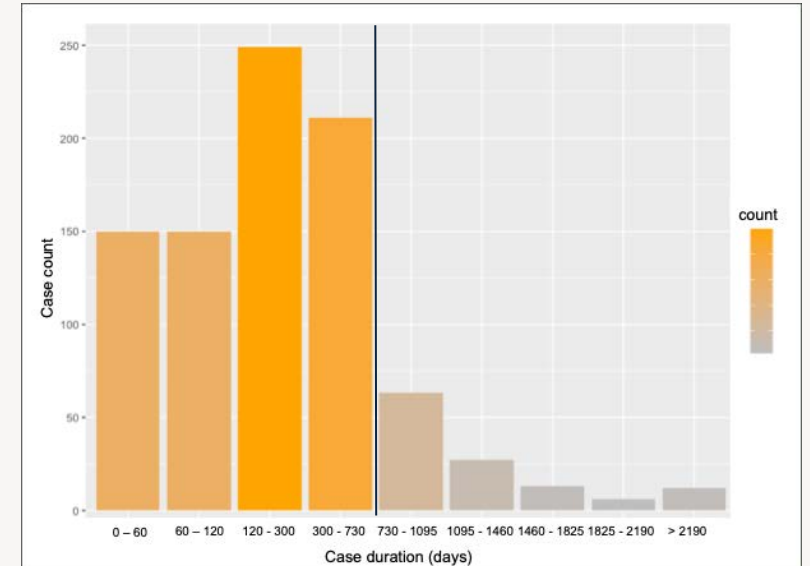
- Case duration spans 0-2190 days, which correspond to roughly 7 years.
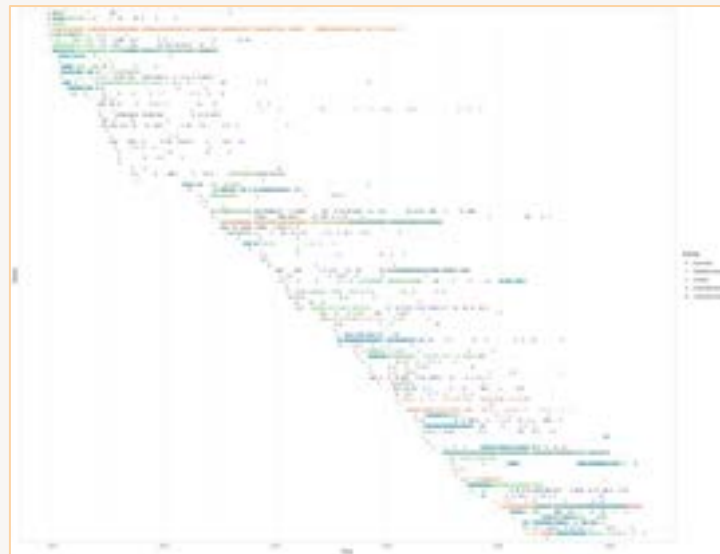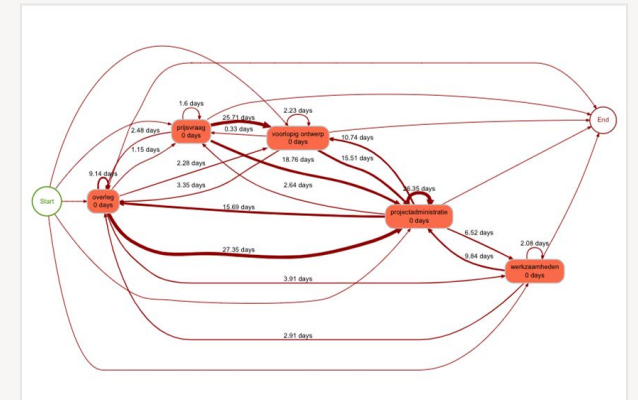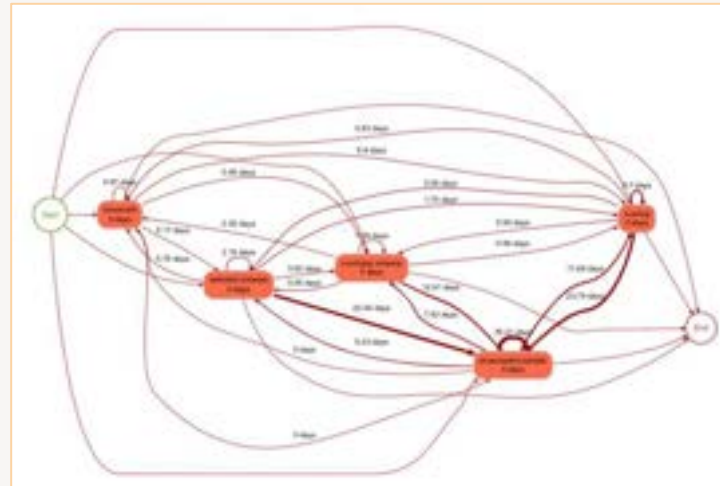- The majority of cases takes up to 2 years.

# Comparative Analysis

**Strategy:** Classify and compare cases to describe groups and find differences.

*Rationale*

- Identify subsets of the data relevant to answer the question

- Serve as a basis for finding correlations

Q2. What is the average duration of projects? What are projects with long lead times?

# Resource Analysis

**Strategy:** Look for correlations to find relations among different process characteristics.

*Rationale*

- Analyze the impact of other process perspectives on the control flow

- Generate new hypotheses

- Hints for root causes

Q3. Are resources involved in multiple projects? What are handoffs patterns between resources across projects?
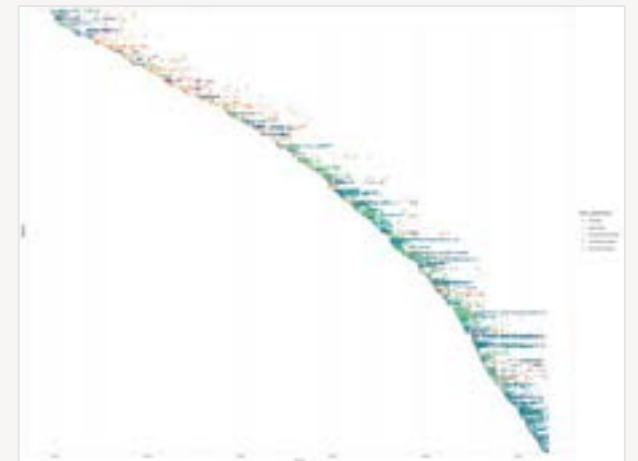
Define
Questions

Data
Collection &
Preparation
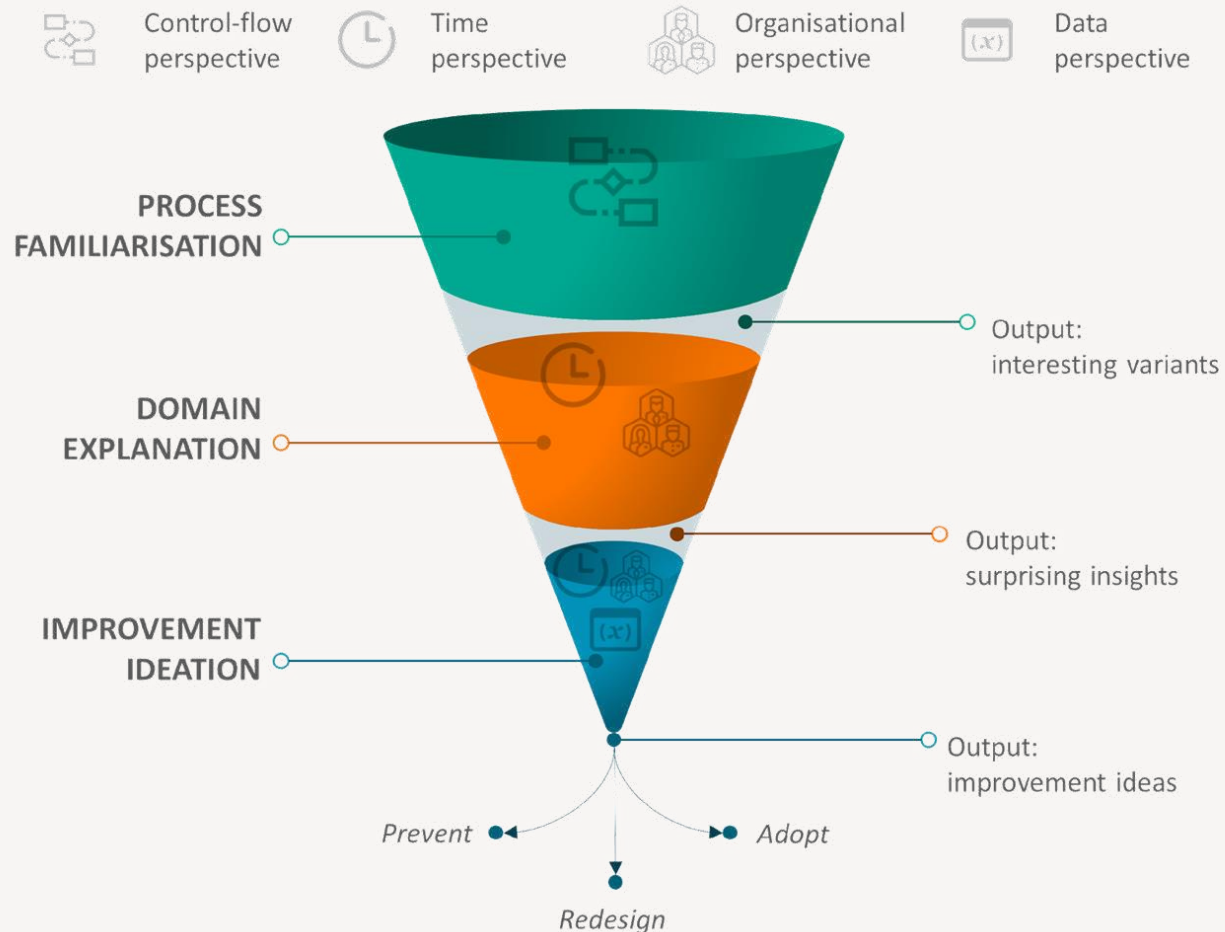
Mining &
Analysis

Results

# Results

# Evaluating Results
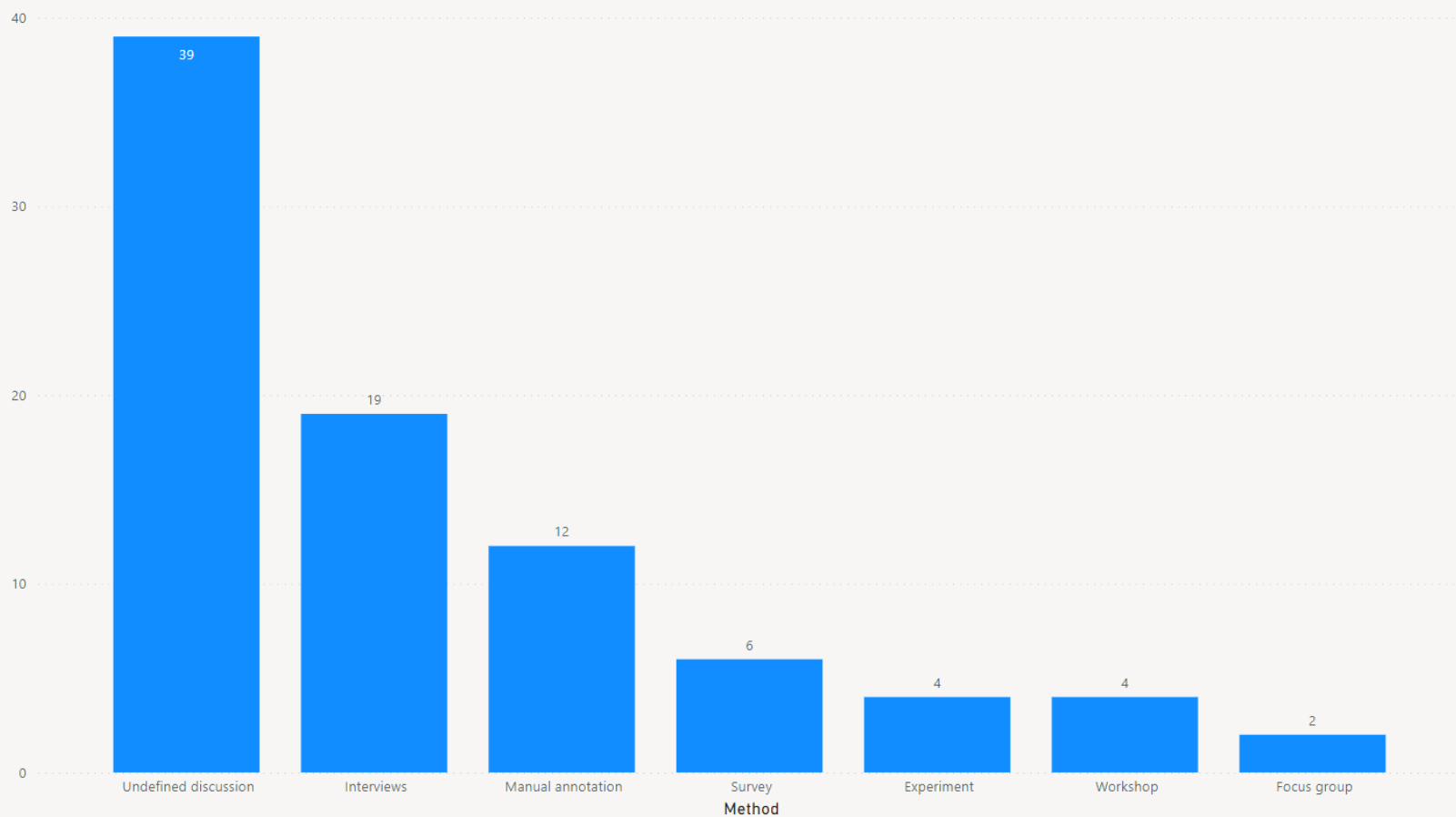


Koorn, J. J., Beerepoot, I., Dani, V. S., Lu, X., Van de Weerd, I., Leopold, H., & Reijers, H. A. (2021, October). Bringing rigor to the qualitative evaluation of process mining findings: an analysis and a proposal. In *2021 3rd International Conference on Process Mining (ICPM)* (pp. 120-127). IEEE.

# Structuring Evaluation Sessions



Beerepoot, I., Martin, N., & Koorn, J. J. (2023). From Insights to INTEL: Evaluating Process Mining Insights with Healthcare Professionals. *HICSS 2023*.

# First Evaluation

*Surprising*

- Number of times prijsvraag/competition occurs across projects

- Definitive design > preliminary design

- Handovers: Ved > Bra and Sob > Klo surprisingly often, Zie > Bra and Sob > Kar not that much
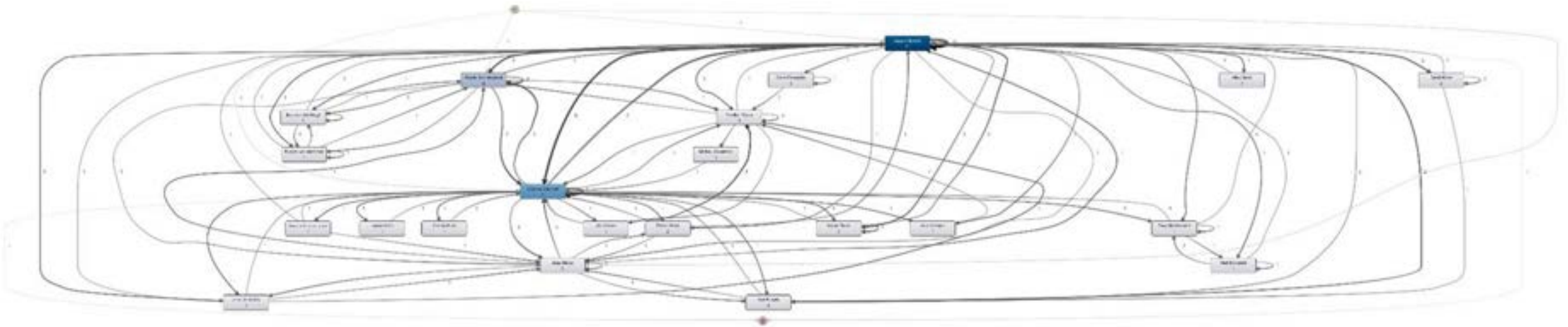
*Emerging questions*

- To what extent are people involved in projects from start to end?

- What are the times between hours spent on the project, the time of registration, the time of approval, and the billing date?

- What is the relation between hours spent and billed amount?
  - not a PM question
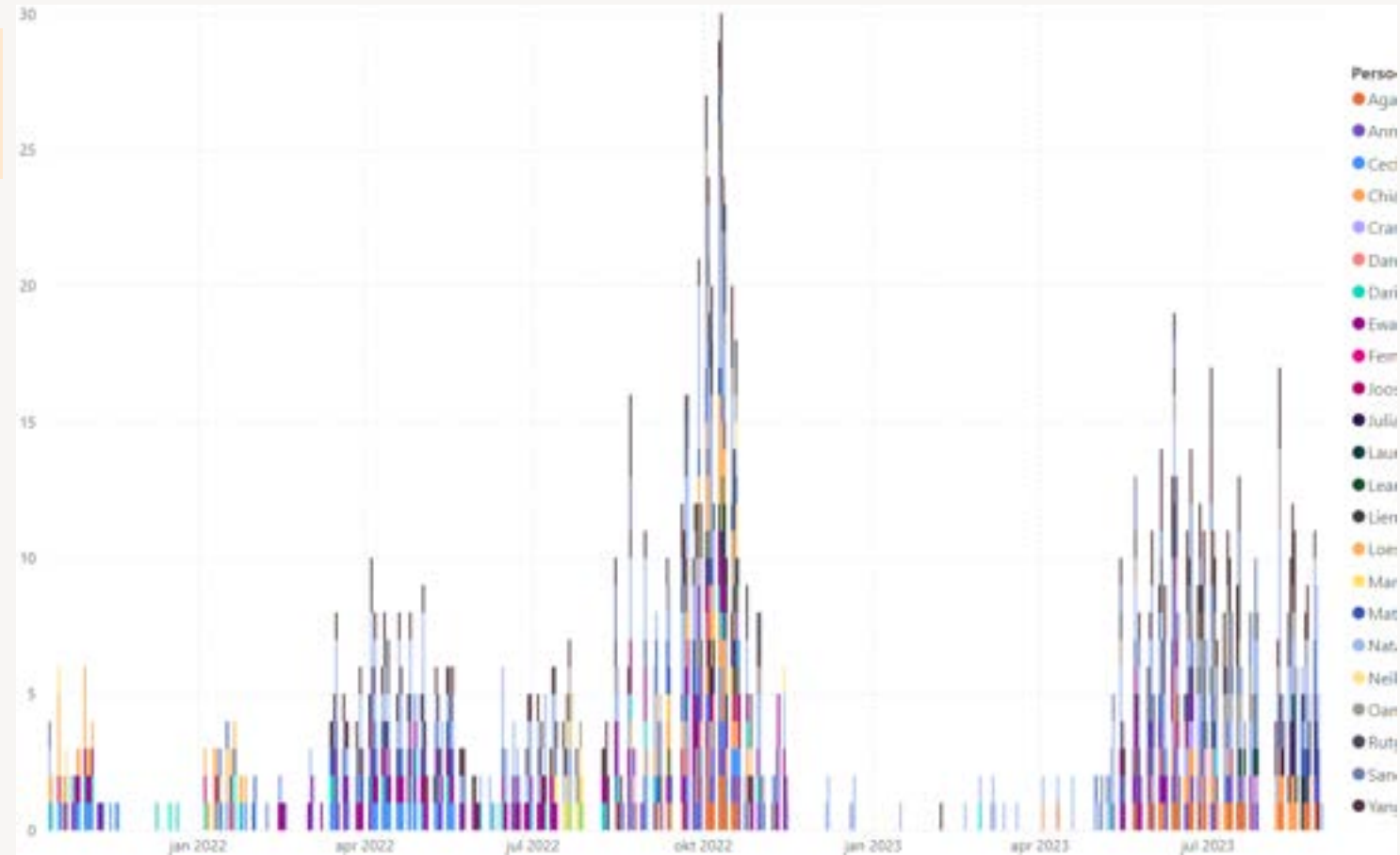
# Social Network Analysis

Q4. Which resources are involved and when?

- Project lead involved at the start and end
- One very central designer

# Social network analysis

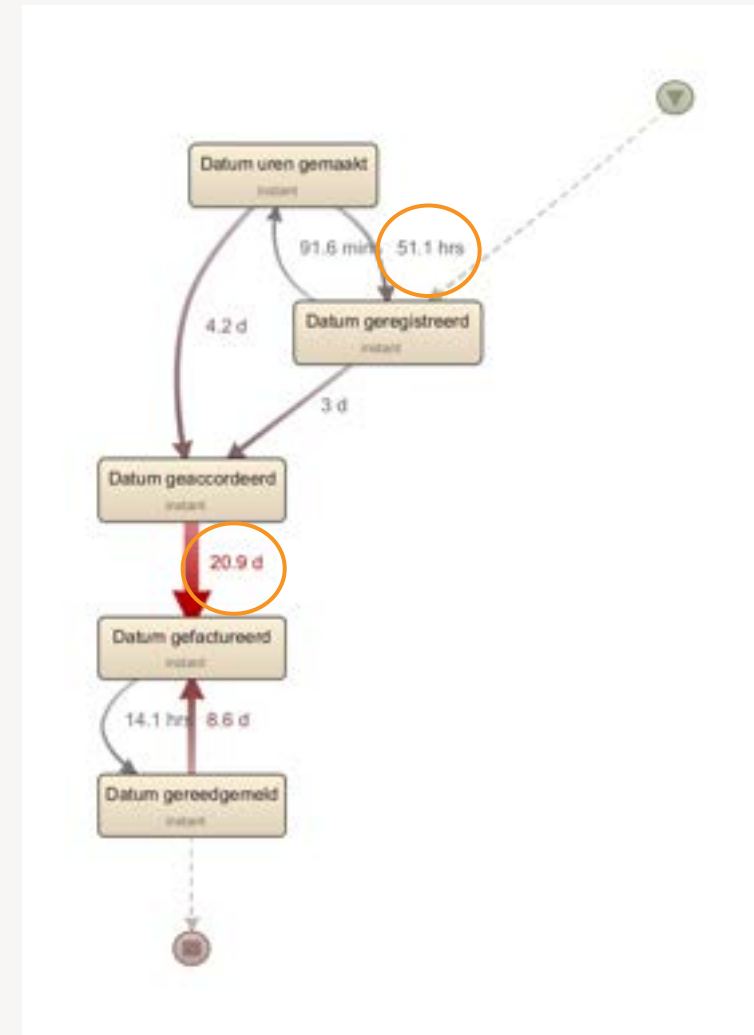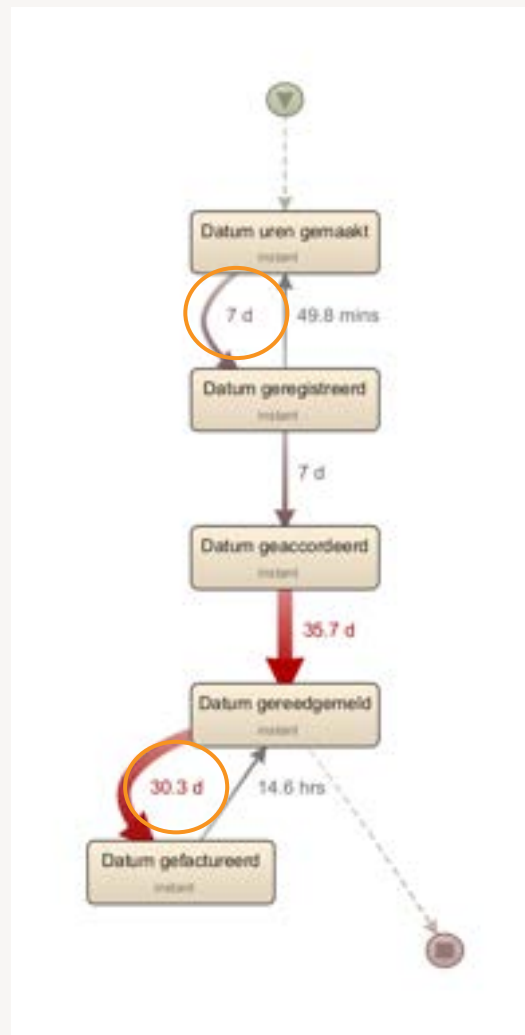Q4. Which resources are involved and when?
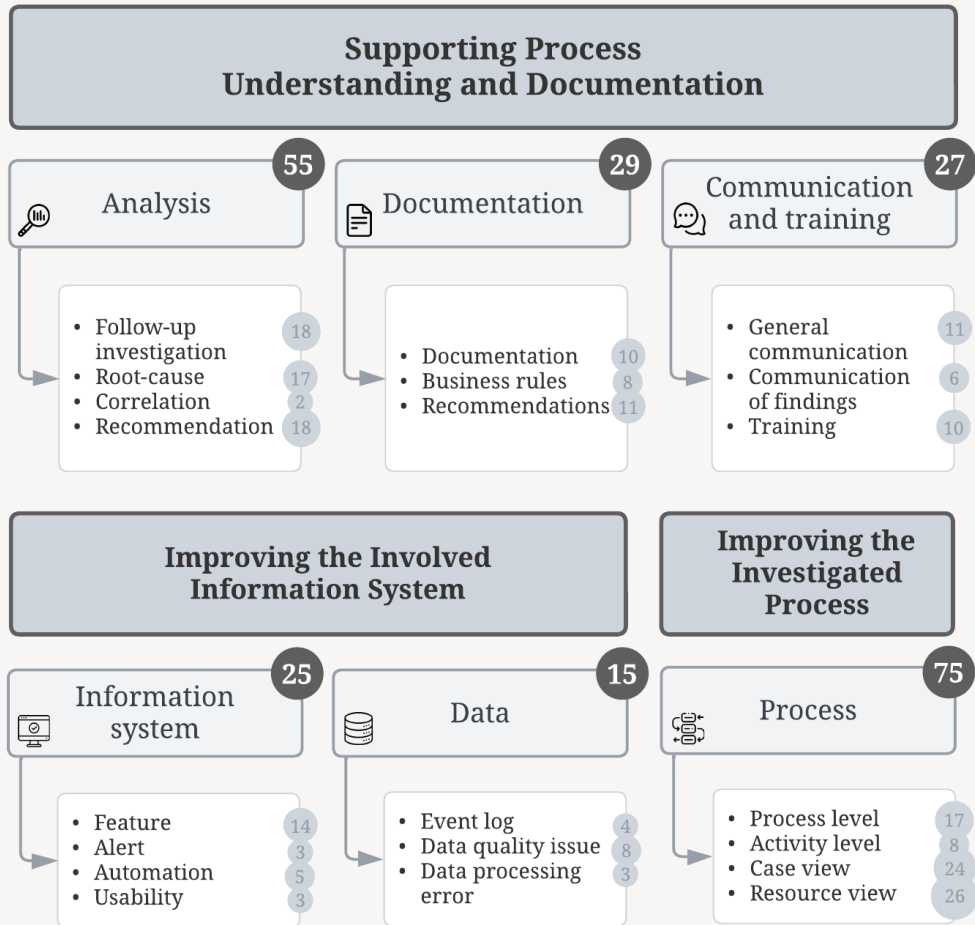


- Deadline halfway October

# Temporal Analysis

Q5. What are the times between hours spent on the project, the time of registration, the time of approval, and the billing date?

- Many people rely on hours being registered
- If late: deferral of billing (or even unable to be billed!)

- Best performer often registers hours ahead of time
- Time until billing also much shorter

# Second Evaluation

| | | |
|---|---|---|
| **Supporting Process Understanding and Documentation** | | |

| Analysis — 55 | Documentation — 29 | Communication and training — 27 |
|---|---|---|
| • Follow-up investigation  18<br>• Root-cause  17<br>• Correlation  2<br>• Recommendation  18 | • Documentation  10<br>• Business rules  8<br>• Recommendations  11 | • General communication  11<br>• Communication of findings  6<br>• Training  10 |

| | |
|---|---|
| **Improving the Involved Information System** | **Improving the Investigated Process** |

| Information system — 25 | Data — 15 | Process — 75 |
|---|---|---|
| • Feature  14<br>• Alert  3<br>• Automation  5<br>• Usability  3 | • Event log  4<br>• Data quality issue  8<br>• Data processing error  3 | • Process level  17<br>• Activity level  8<br>• Case view  24<br>• Resource view  26 |

*Actions chosen*

1. Follow-up investigations financial data (compare profitable and loss-making projects)
2. Monitoring dashboards for project leads (resource perspective)
   a. support for project meetings
   b. check reliability on single resources
   c. organize teams for projects
3. Improve activity naming (for some) projects to enable better analyses
4. Awarding resources that register their hours best

Stein Dani, V., Leopold, H., van der Werf, J. M. E., Beerepoot, I.M., & Reijers, H. A. (2023). From Process Mining Insights to Process Improvement: All Talk and No Action? In International Conference on Cooperative Information Systems (CoopIS).

# Strategies from Qualitative Methods

| Strategy | Effect | Practical guideline |
| --- | --- | --- |
| Engage with the field of research | An open and honest evaluation | Carefully select domain experts; include data quality issues in presentation of results |
| Triangulation | Completeness and consistency of the results | Use multiple quantitative (e.g. simulations) and qualitative evaluation methods (e.g. interviews) |
| Peer review or external audit | Credibility of the analysis and interpretation | Plan peer reviews to reflect on research design, approach, and results on a regular basis and keep notes of these meetings |
| Refine work hypothesis | Transparency and soundness of the results | Keep detailed notes on hypotheses, how they are tested, and the final results and use these to guide the evaluation with domain experts |
| Clarify and normalize bias | Transparency and reliability of the results | Discuss different types of biases in the evaluation or limitation section |
| Perform member checking | Credibility of the results | Ask interviewees to check the correctness and authenticity of a summarized report of the interview results and interpretation |

Koorn, J. J., Beerepoot, I., Dani, V. S., Lu, X., Van de Weerd, I., Leopold, H., & Reijers, H. A. (2021, October). Bringing rigor to the qualitative evaluation of process mining findings: an analysis and a proposal. In *2021 3rd International Conference on Process Mining (ICPM)* (pp. 120-127). IEEE.
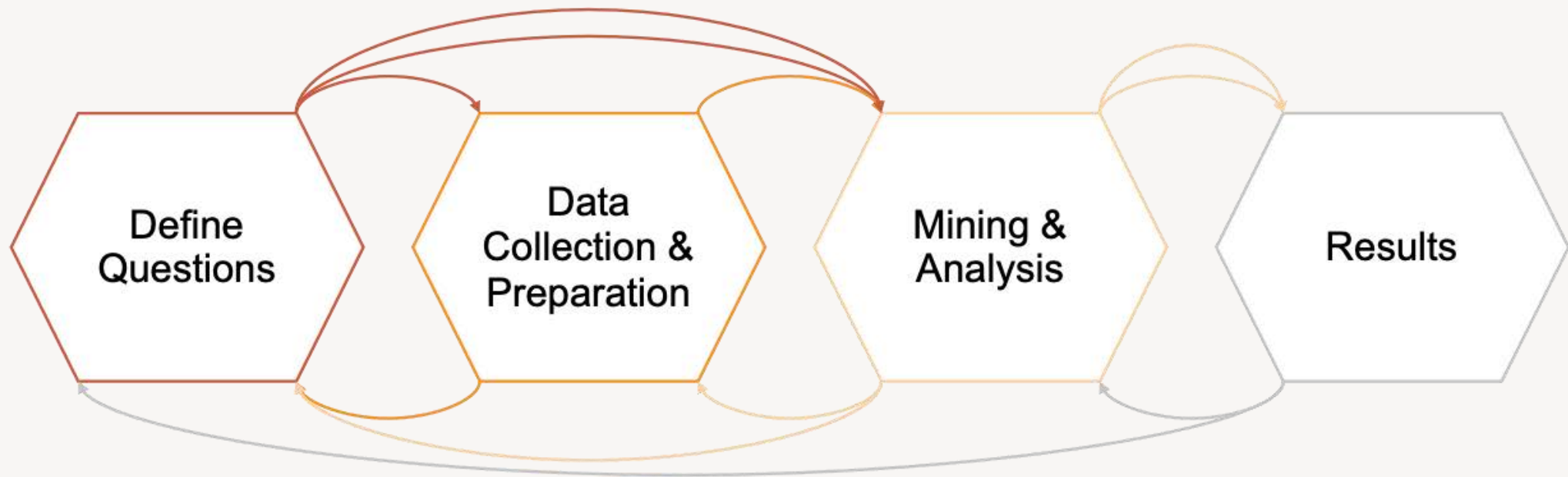
# Closing
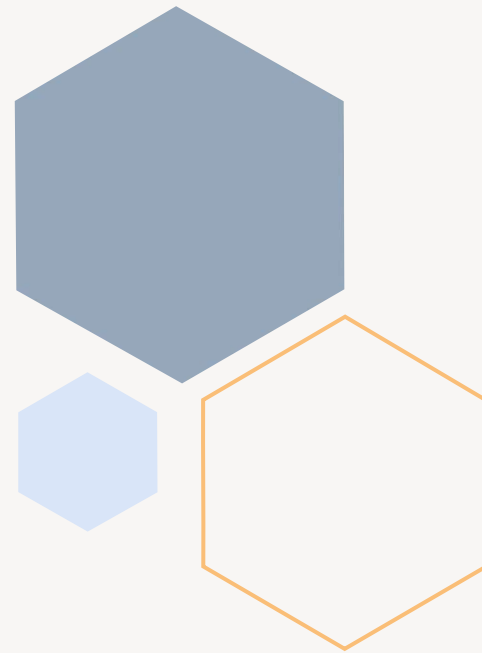
# Supporting Process Mining Tasks



Data →

Define Questions

Data Collection & Preparation

Mining & Analysis

Results

# The Process View



To what extent are people involved in projects from start to end?

# Literature

**Define Questions**
- Zerbato, F., Koorn, J. J., Beerepoot, I., Weber, B., & Reijers, H. A. (2022, September). On the Origin of Questions in Process Mining Projects. In *International Conference on Enterprise Design, Operations, and Computing* (pp. 165-181). Cham: Springer International Publishing.

**Data Collection and Preparation**
- Stein Dani, V., Leopold, H., van der Werf, J. M. E., & Reijers, H. A. (2022, September). Supporting Event Log Extraction Based on Matching. In *International Conference on Business Process Management* (pp. 322-333). Cham: Springer International Publishing.
- Stein Dani, V., Leopold, H., van der Werf, J. M. E., Lu, X., Beerepoot, I., Koorn, J. J., & Reijers, H. A. (2021, September). Towards understanding the role of the human in event log extraction. In *International Conference on Business Process Management* (pp. 86-98). Cham: Springer International Publishing.

**Mining and Analysis**
- Zerbato, F., Soffer, P., & Weber, B. (2022, September). Process mining practices: evidence from interviews. In *International Conference on Business Process Management* (pp. 268-285). Cham: Springer International Publishing.

**Results**
- Beerepoot, I., Martin, N., & Koorn, J. J. (2023). From Insights to INTEL: Evaluating Process Mining Insights with Healthcare Professionals. *HICSS 2023.*
- Koorn, J. J., Beerepoot, I., Dani, V. S., Lu, X., Van de Weerd, I., Leopold, H., & Reijers, H. A. (2021, October). Bringing rigor to the qualitative evaluation of process mining findings: an analysis and a proposal. In *2021 3rd International Conference on Process Mining (ICPM)* (pp. 120-127). IEEE.
- Stein Dani, V., Leopold, H., van der Werf, J. M. E., Beerepoot, I.M., & Reijers, H. A. (2023). From Process Mining Insights to Process Improvement: All Talk and No Action? In *International Conference on Cooperative Information Systems* (CoopIS).

**Challenges (All phases)**
- Zimmermann, L., Zerbato, F., & Weber, B. (2022, May). Process mining challenges perceived by analysts: An interview study. In *International Conference on Business Process Modeling, Development and Support* (pp. 3-17). Cham: Springer International Publishing.

# Thank You

Iris Beerepoot

Francesca Zerbato

Barbara Weber

Pnina Soffer

**Get in touch!**

i.m.beerepoot@uu.nl

francesca.zerbato@unisg.ch

**Swiss National Science Foundation**
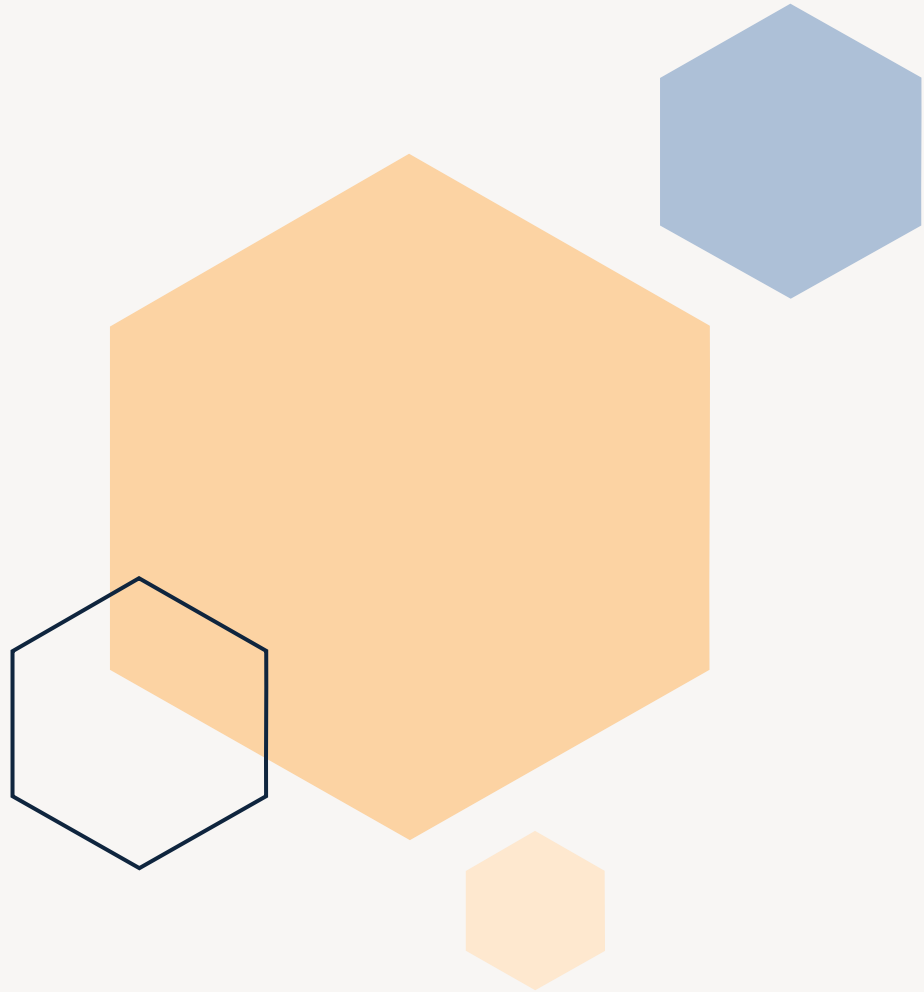
# Questions?

Let's discuss!

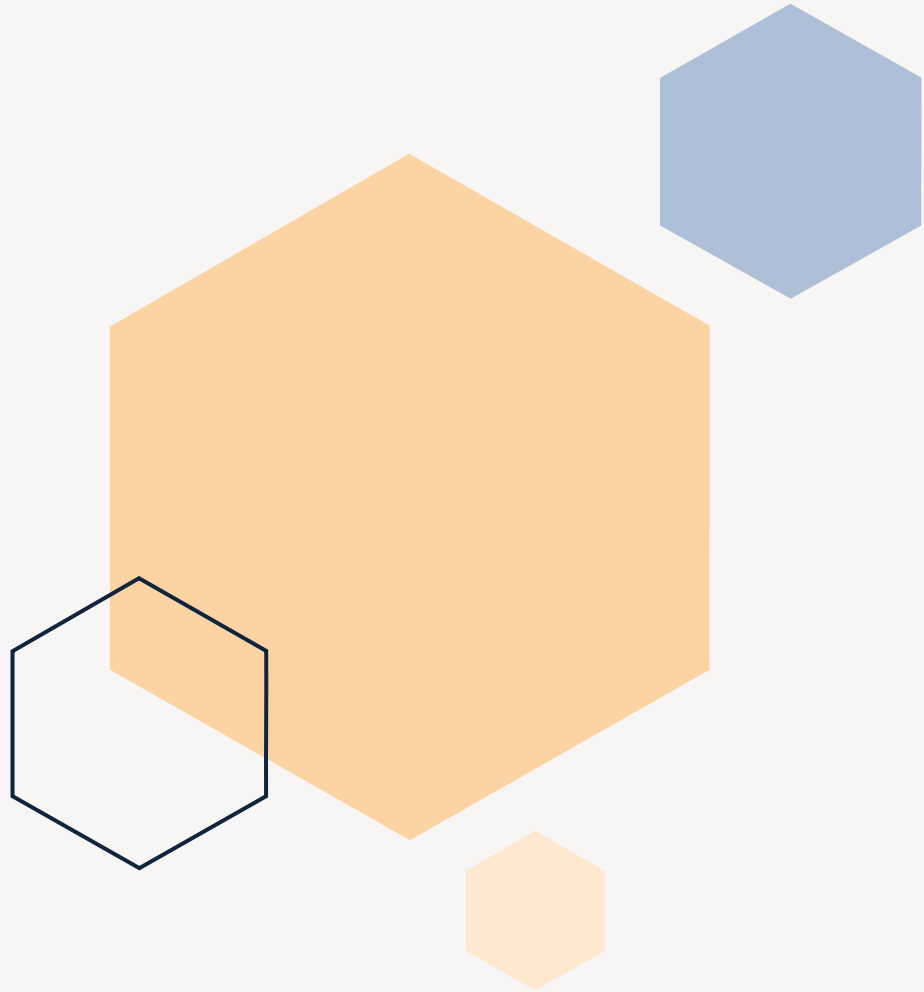# What makes an analysis question a "good" one?

Define Questions

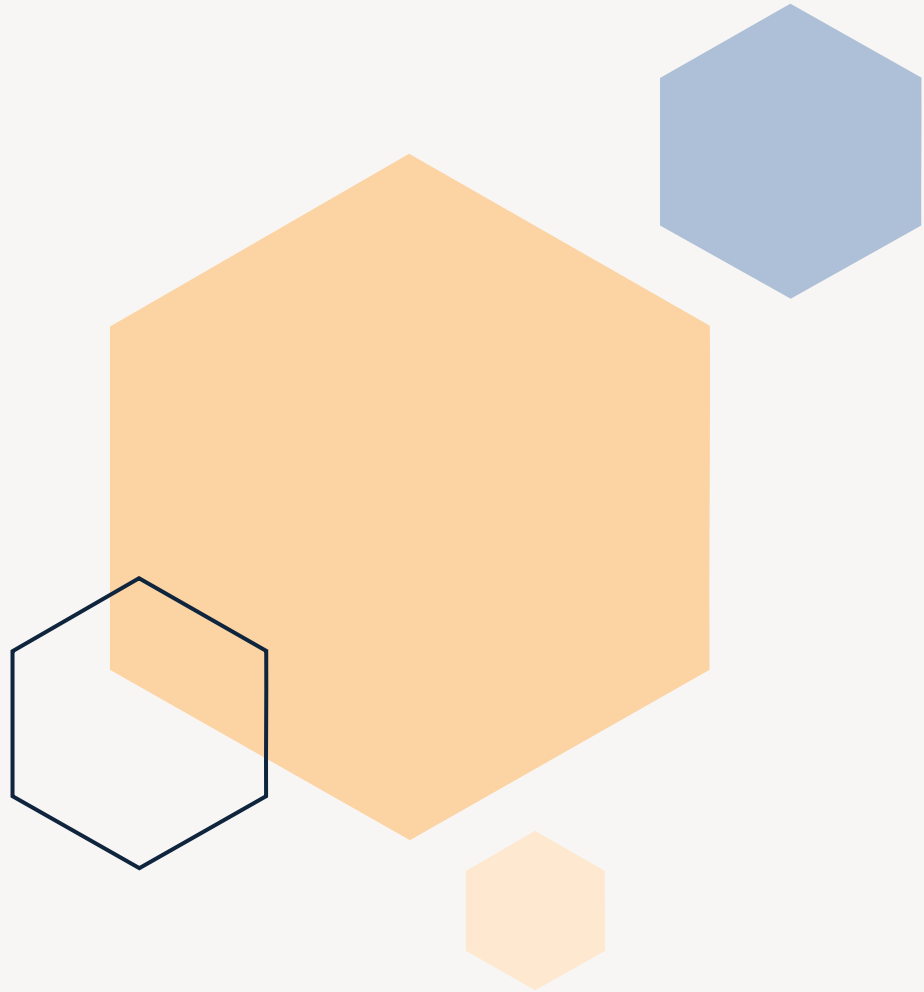# For what questions should Process Mining not be used?

Define Questions

# Is the effort of data collection and preparation ever too high to outweigh the benefits of PM?

Data Collection & Preparation

# What are crucial factors that determine how to steer the analysis?

Mining & Analysis

# How 'bad' is it to not have a structured validation of insights?

Results

# Thank you

Iris Beerepoot
Francesca Zerbato
Barbara Weber
Pnina Soffer

**Get in touch!**
i.m.beerepoot@uu.nl
francesca.zerbato@unisg.ch

55