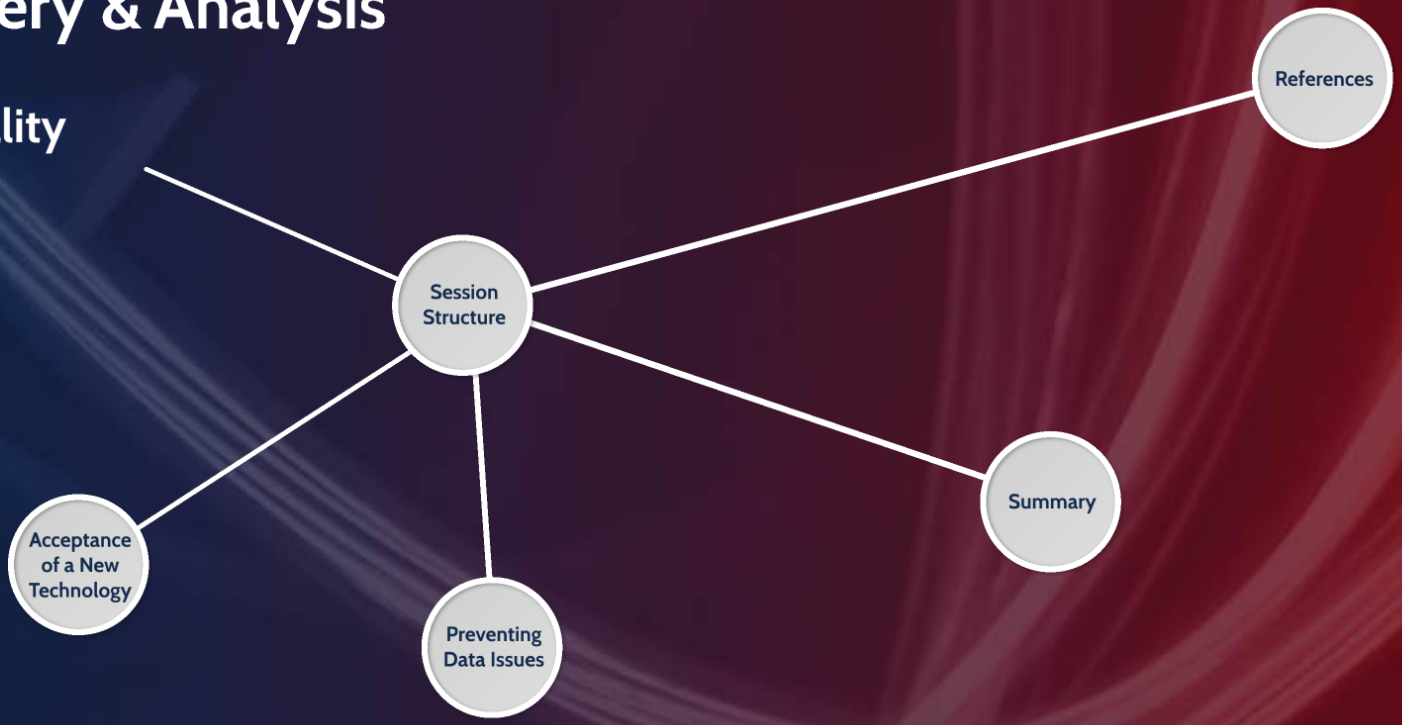
Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality



Friedrich-Alexander University, Chair of Digital Industrial Service Systems

Session Structure

Learning Objectives – Data quality

How quality of insights facilitates the acceptance of Process Mining in an organization

How data quality affects the quality of insights

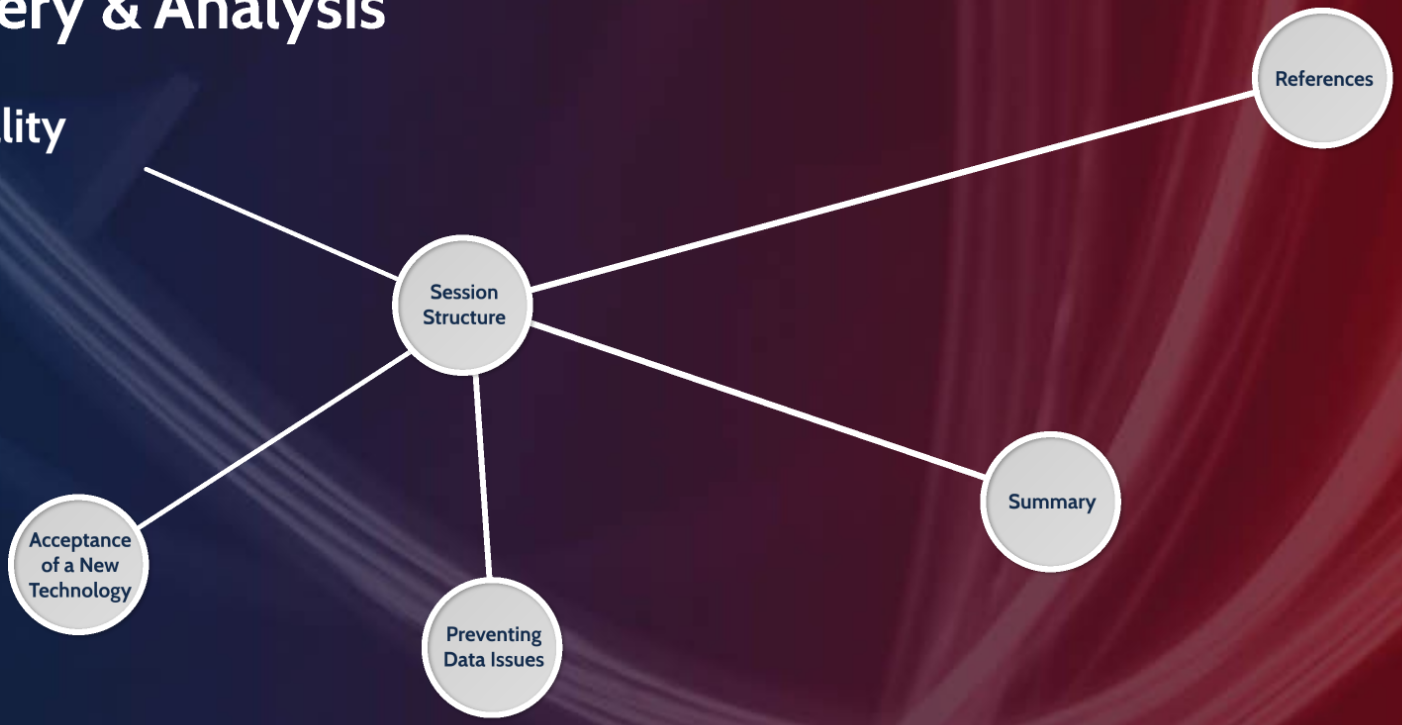
What are the common problems of data quality in the event log

How to prevent the data problems when preparing the event log

Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality

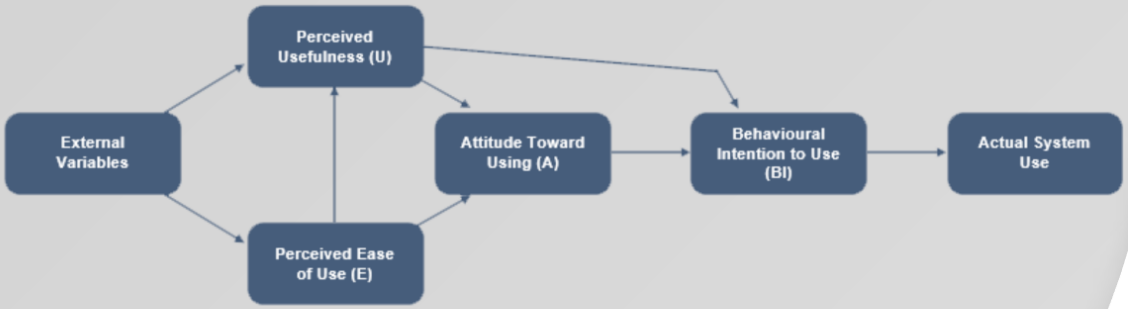


Friedrich-Alexander University, Chair of Digital Industrial Service Systems

Acceptance of a New Technology

Acceptance of a New Technology

Technology Acceptance Model (TAM) (Davis, 1987)



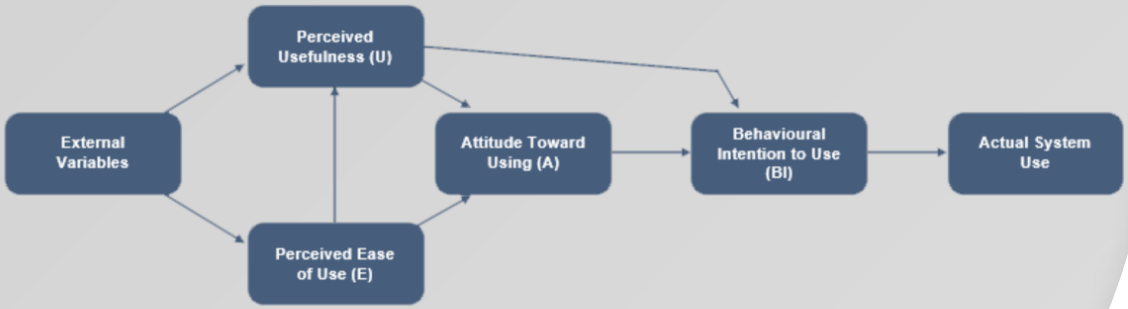
Acceptance of Process Mining

Influencing Factors of Analysis Quality

Acceptance of Process Mining

Acceptance of Process Mining

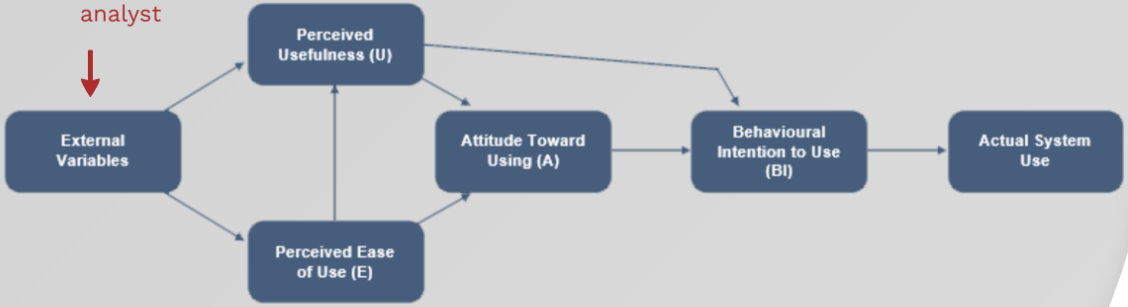
TAM in Process Mining



Acceptance of Process Mining

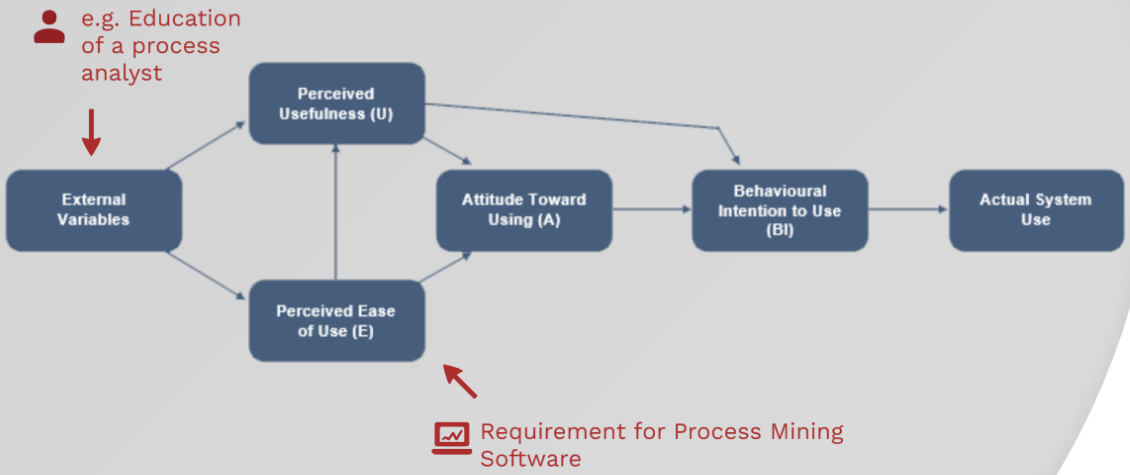
TAM in Process Mining

e.g. Education of a process analyst



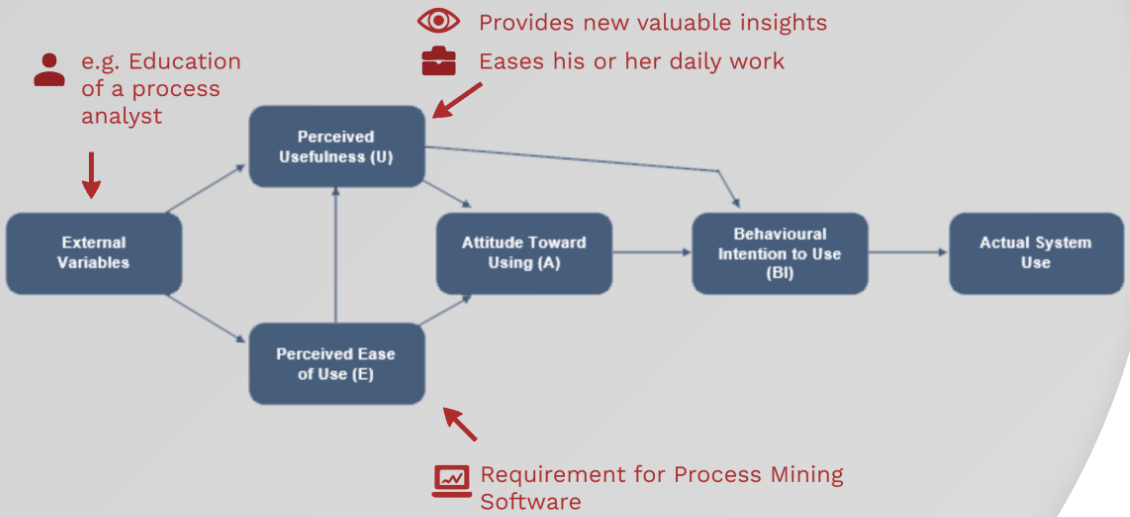
Acceptance of Process Mining

TAM in Process Mining



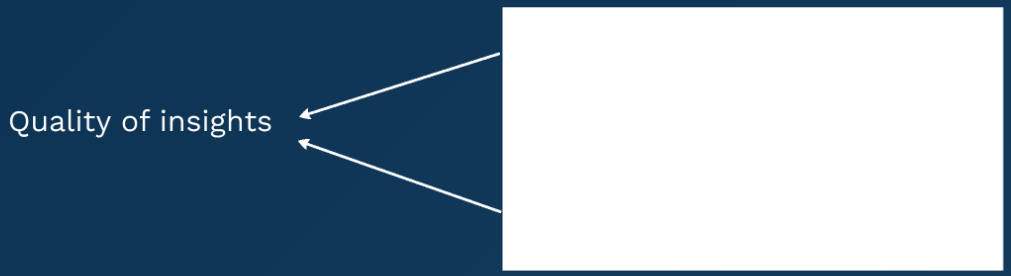
Acceptance of Process Mining

TAM in Process Mining



Influencing Factors of Analysis Quality

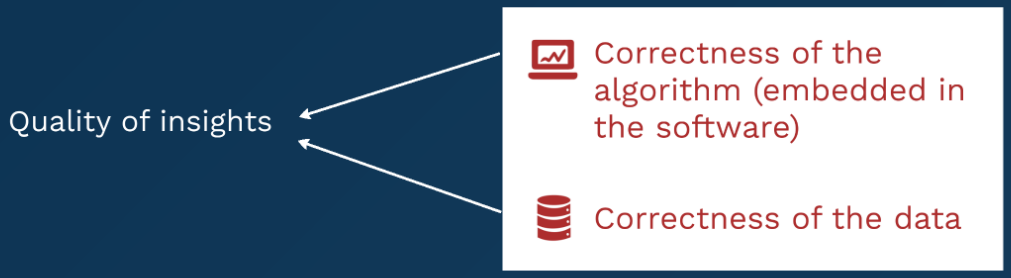
Influencing Factors of Analysis Quality



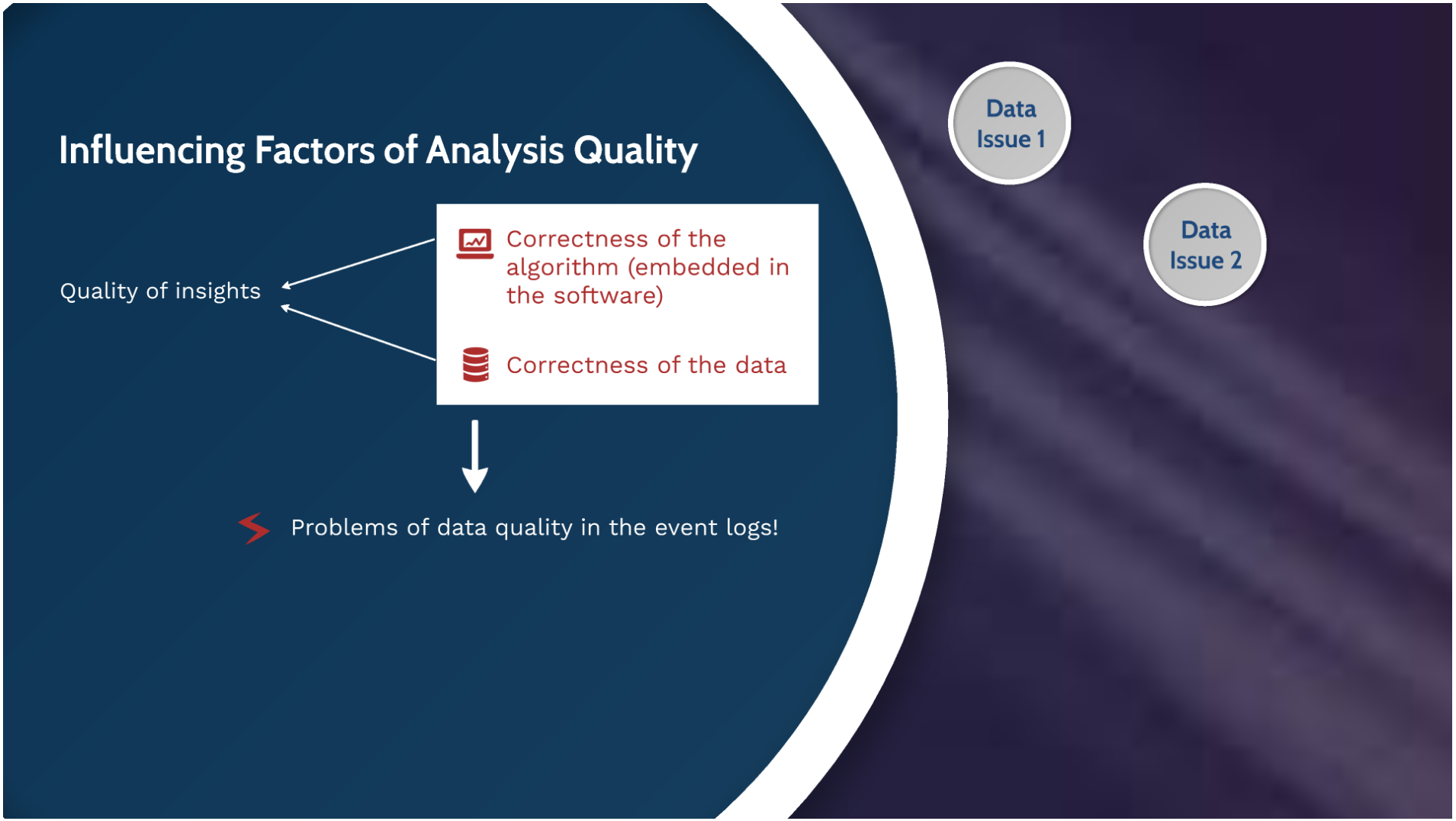
Data Issue 1

Data Issue 2

Influencing Factors of Analysis Quality



Influencing Factors of Analysis Quality



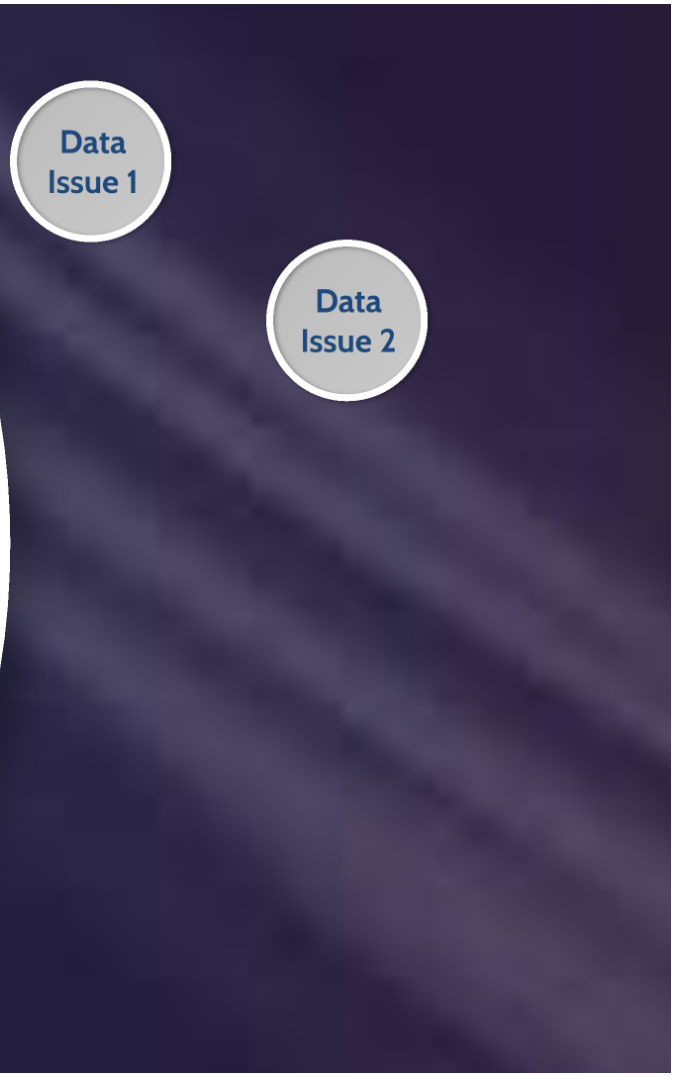
Influencing Factors of Analysis Quality

Quality of insights

- Correctness of the algorithm (embedded in the software)
- Correctness of the data

Problems of data quality in the event logs!

1. **Noise**
2. **Incompleteness**



Data Issue 1 - Noise (Van Der Aalst, 2016)

Data Issue 1 - Noise (Van Der Aalst, 2016)

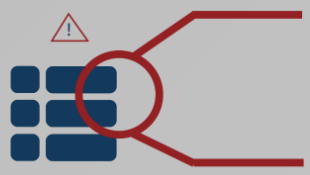
Rare and infrequent behavior

Data Issue 1 - Noise (Van Der Aalst, 2016)

Rare and infrequent behavior

Incorrect information

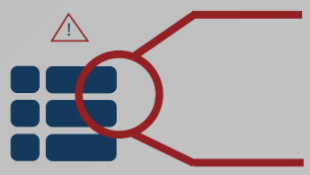
Data Issue 1 - Noise (Van Der Aalst, 2016)



Rare and infrequent behavior

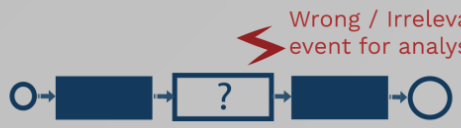
Incorrect information

Data Issue 1 - Noise (Van Der Aalst, 2016)



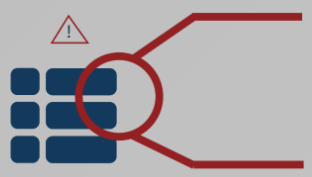
Rare and infrequent behavior

Incorrect information



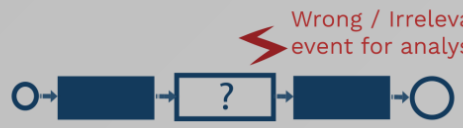
Less meaningful/
comprehensible
results

Data Issue 1 - Noise (Van Der Aalst, 2016)

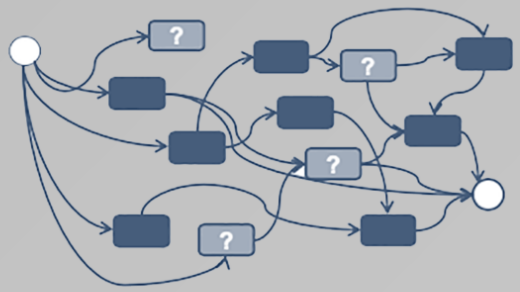


Rare and infrequent behavior

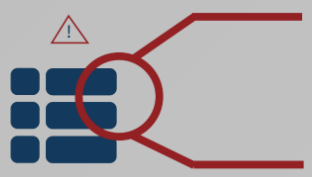
Incorrect information



Less meaningful/
comprehensible
results

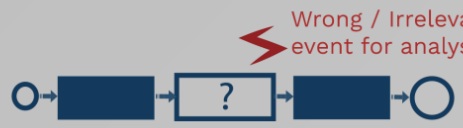


Data Issue 1 - Noise (Van Der Aalst, 2016)

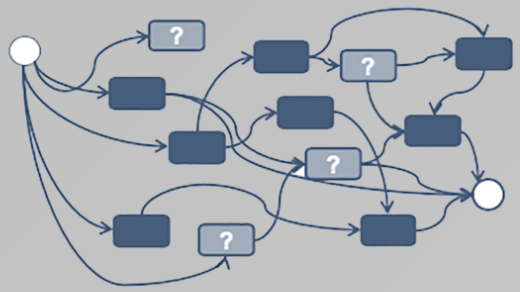


Rare and infrequent behavior

Incorrect information



Less meaningful/
comprehensible
results



Spaghetti diagram
(Van Der Aalst, 2011)

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Possible causes:

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Possible causes:

- parts of the process are executed in another IT system

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Possible causes:

- parts of the process are executed in another IT system
- system/data not included it into the event log

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Possible causes:

- parts of the process are executed in another IT system
- system/data not included it into the event log
- process (step) is not digitized

Data Issue 2 - Incompleteness (Van Der Aalst, 2016)

Not enough information

Possible causes:

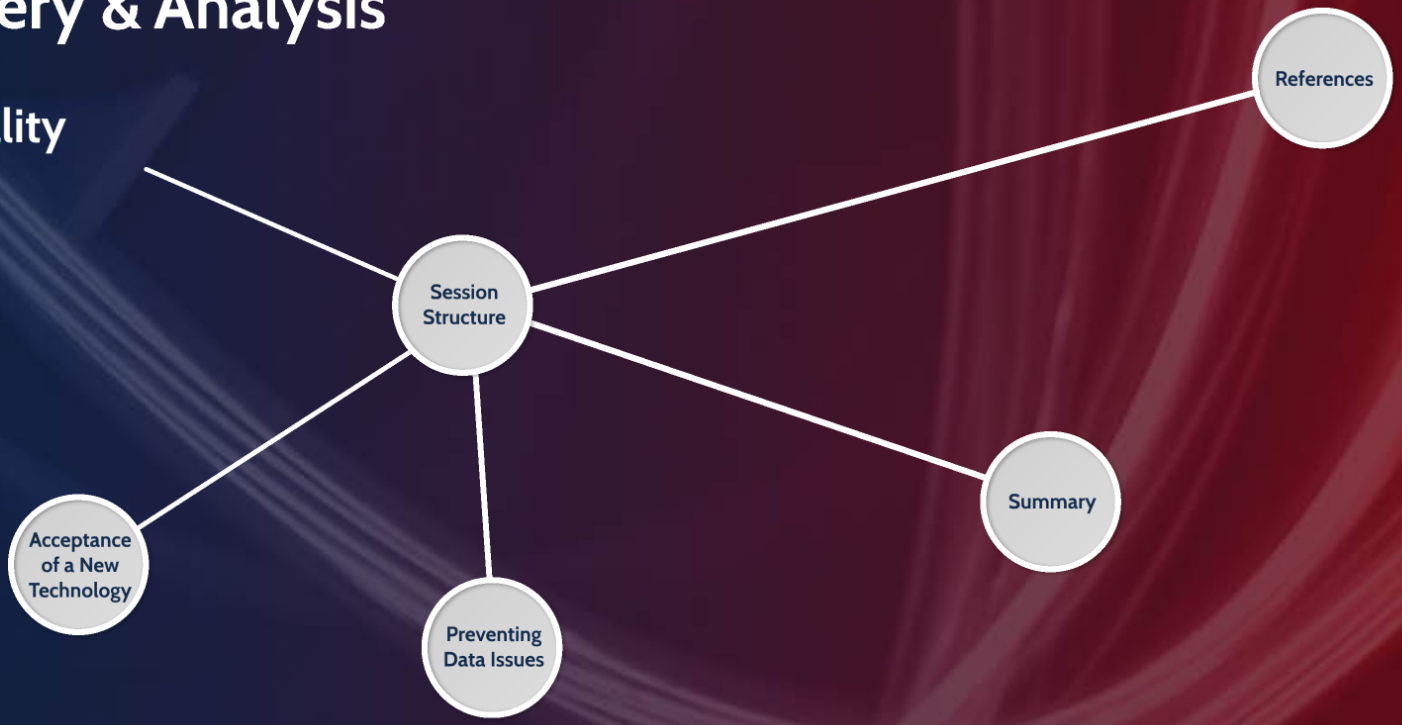
- parts of the process are executed in another IT system
- system/data not included it into the event log
- process (step) is not digitized

Disable a comprehensive data analysis

Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality



Friedrich-Alexander University, Chair of Digital Industrial Service Systems

Preventing Data Issues

Recap:

too much → **1 Noise**

too few → **2 Incompleteness**

Correctness of the algorithm (embedded in the software) ✓

Correctness of the data ✓

Preventing Data Issues

Recap:

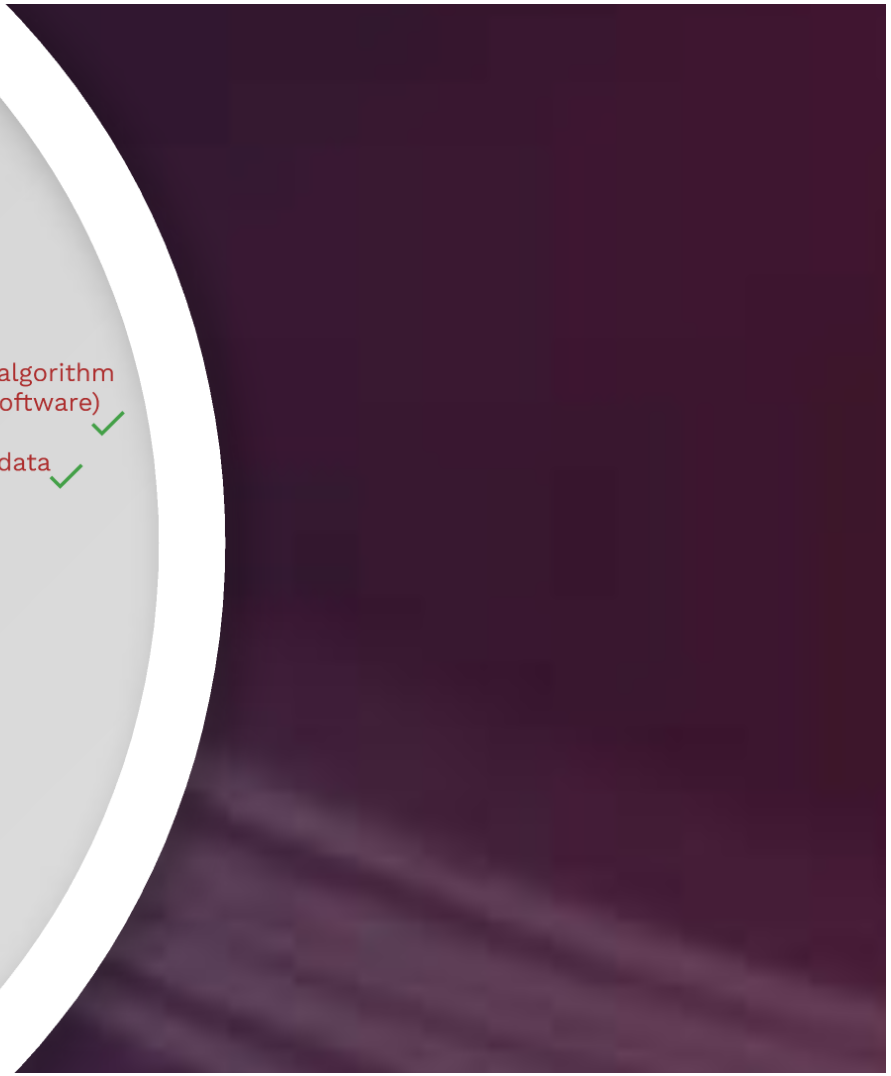
too much → **1 Noise**

too few → **2 Incompleteness**

Correctness of the algorithm (embedded in the software) ✓

Correctness of the data ✓

When preparing an event log



Preventing Data Issues

Recap:

too much → **1 Noise**

too few → **2 Incompleteness**

Correctness of the algorithm (embedded in the software) ✓

Correctness of the data ✓

When preparing an event log

(Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 1

Correlation (Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 1

Correlation (Van Der Aalst, 2016)

Focus: link all process steps for one case

Challenges for Preparing an Event Log - 1

Correlation (Van Der Aalst, 2016)

Focus: link all process steps for one case

E.g. Sales process: Sales Order

Challenges for Preparing an Event Log - 2

Timestamps (Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 2

Timestamps (Van Der Aalst, 2016)

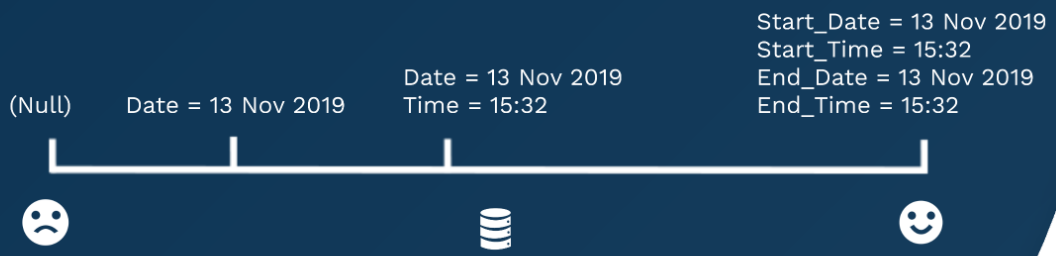
★ **The time of execution for each activity (start and end) is tracked in detail**

Challenges for Preparing an Event Log - 2

Timestamps (Van Der Aalst, 2016)

★ **The time of execution for each activity (start and end) is tracked in detail**

Different system stores different types of timestamps

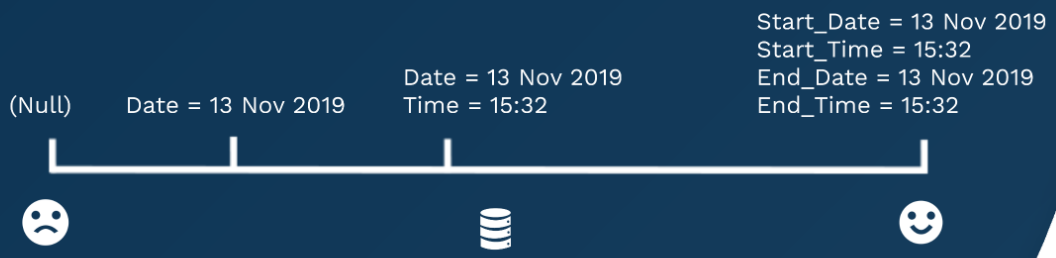


Challenges for Preparing an Event Log - 2

Timestamps (Van Der Aalst, 2016)

★ **The time of execution for each activity (start and end) is tracked in detail**

Different system stores different types of timestamps



⚠ Process Mining cannot be done without the timing & order of the activities

Challenges for Preparing an Event Log - 3

Snapshots (Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 3

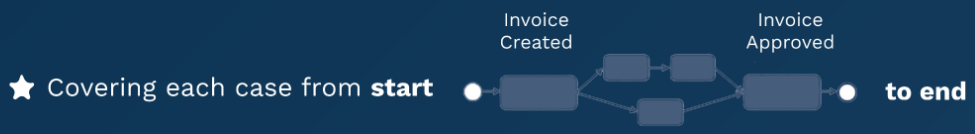
Snapshots (Van Der Aalst, 2016)

★ Covering each case from **start** 

to end

Challenges for Preparing an Event Log - 3

Snapshots (Van Der Aalst, 2016)



E.g. analysis period: 2019-01-01 – 2019-12-31

Challenges for Preparing an Event Log - 3

Snapshots (Van Der Aalst, 2016)

★ Covering each case from **start** **to end**

E.g. analysis period: 2019-01-01 – 2019-12-31

⚠ Missing a part of each case

Challenges for Preparing an Event Log - 4

Scoping (Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 4

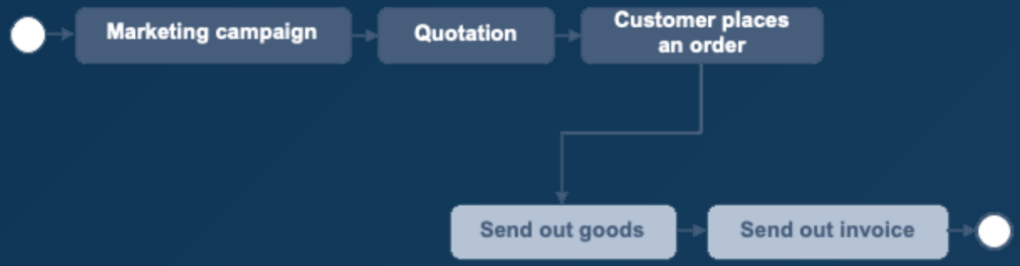
Scoping (Van Der Aalst, 2016)

★ **"start" & "end" of the process**

Challenges for Preparing an Event Log - 4

Scoping (Van Der Aalst, 2016)

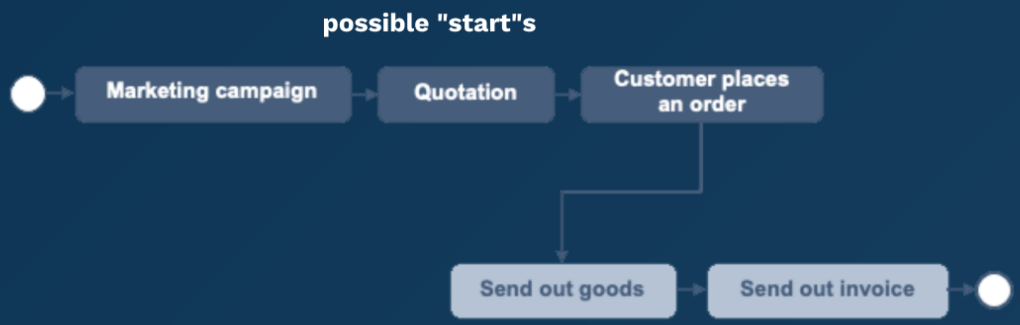
★ **"start" & "end" of the process**



Challenges for Preparing an Event Log - 4

Scoping (Van Der Aalst, 2016)

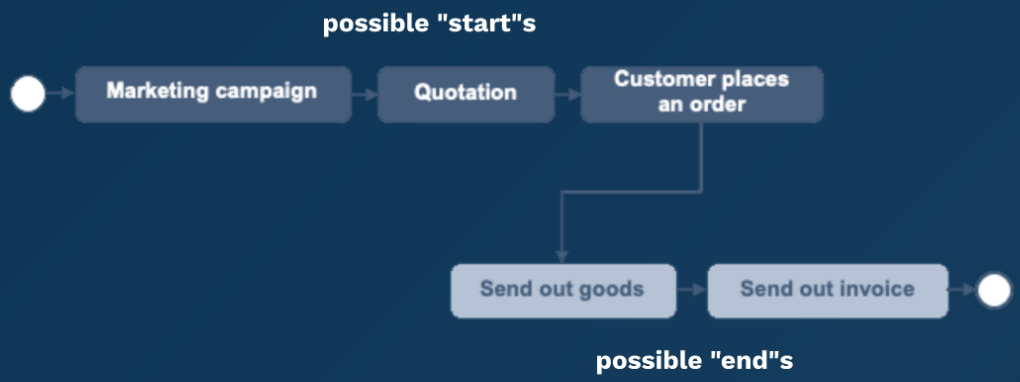
★ **"start" & "end" of the process**



Challenges for Preparing an Event Log - 4

Scoping (Van Der Aalst, 2016)

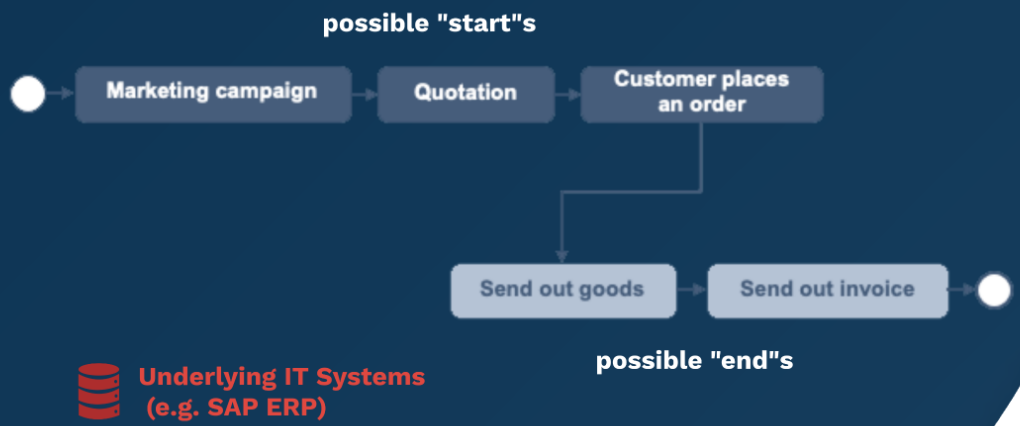
★ **"start" & "end" of the process**



Challenges for Preparing an Event Log - 4

Scoping (Van Der Aalst, 2016)

★ **"start" & "end" of the process**



Underlying IT Systems (e.g. SAP ERP)

Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Too generic

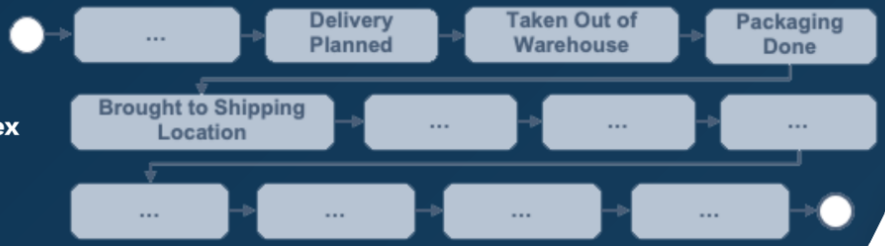
Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Too generic

Too complex



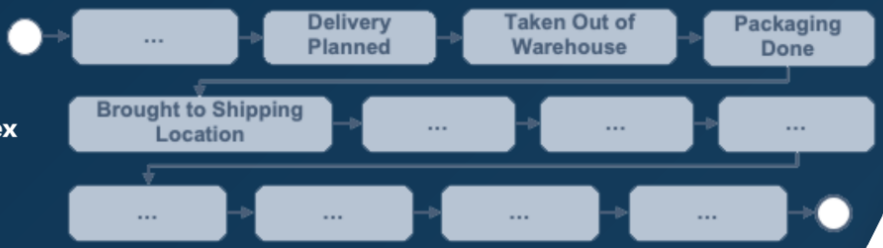
Challenges for Preparing an Event Log - 5

Granularity (Van Der Aalst, 2016)

★ **Right level of depth & detail**

Too generic

Too complex

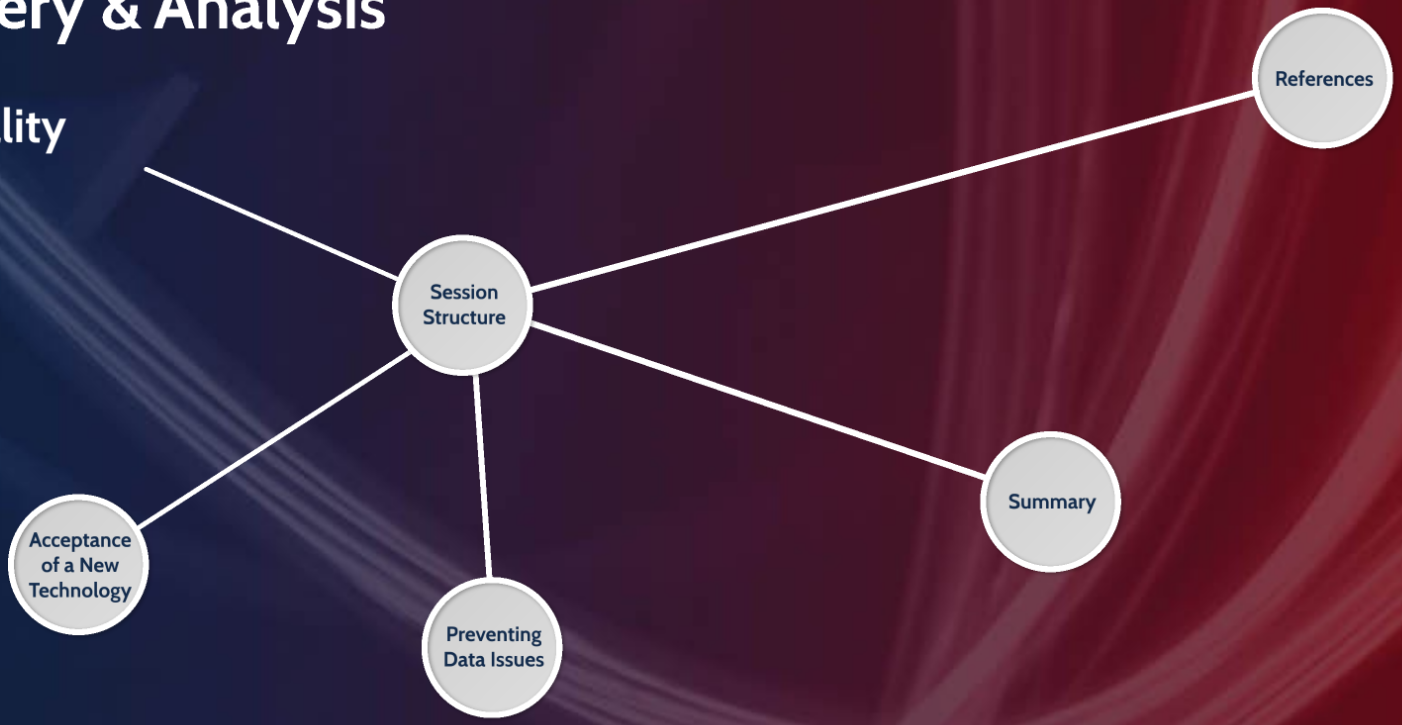


⚠ Balance of complexity and detail that yields insights

Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality



Friedrich-Alexander University, Chair of Digital Industrial Service Systems

Summing Up

Summing Up

2 **common problems in data sets**

Summing Up

2 common problems in data sets

- Noise
- Incompleteness

Summing Up

2 common problems in data sets

- Noise
- Incompleteness

5 challenges of ensuring data quality

Summing Up

2 **common problems in data sets**

- Noise
- Incompleteness

5 **challenges of ensuring data quality**

Summing Up

2 common problems in data sets

- Noise
- Incompleteness

5 challenges of ensuring data quality

★ High-quality event log = Meaningful results

Summing Up

2 common problems in data sets

- Noise
- Incompleteness

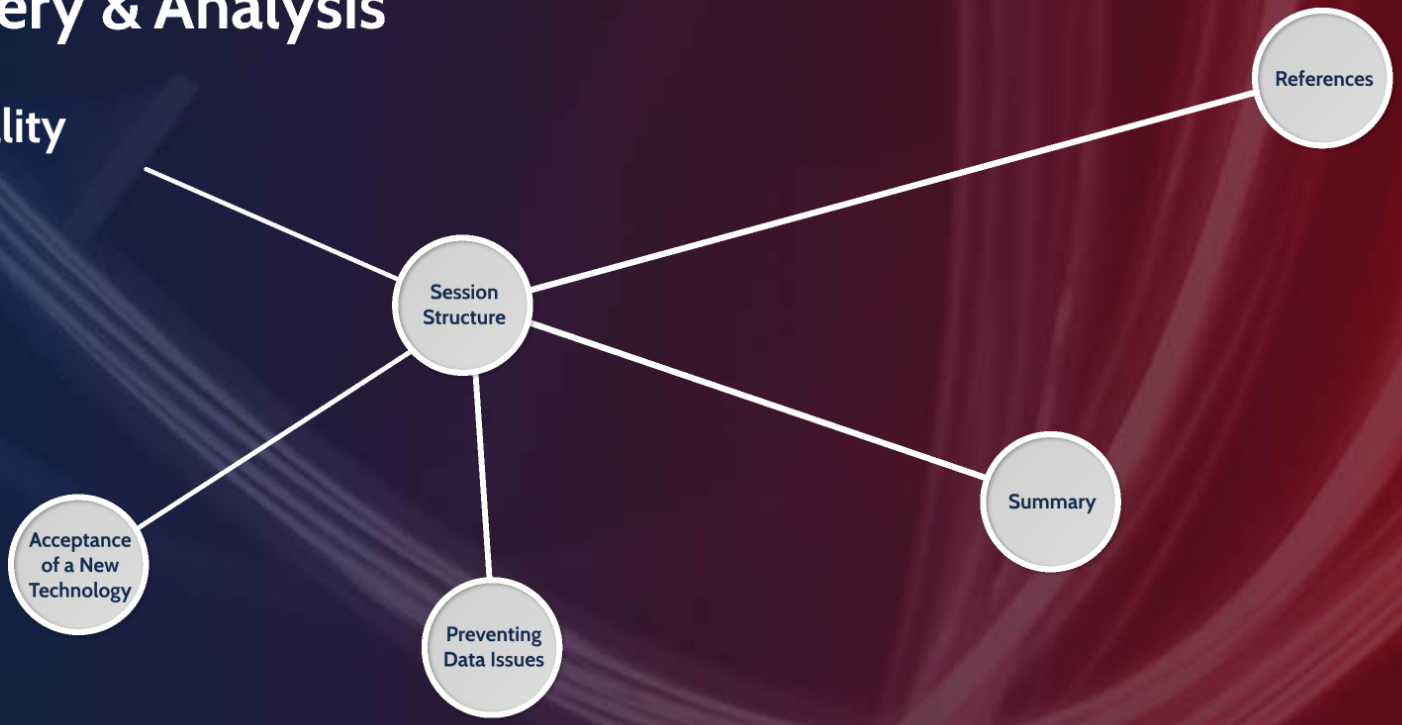
5 challenges of ensuring data quality

- ★ High-quality event log = Meaningful results
- ★ Correct & meaningful results = Trust in your solution

Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality



Friedrich-Alexander University, Chair of Digital Industrial Service Systems

References

1. Davis, F. (1987). User Acceptance of Information Systems: The Technology Acceptance Model (TAM). 1st ed.

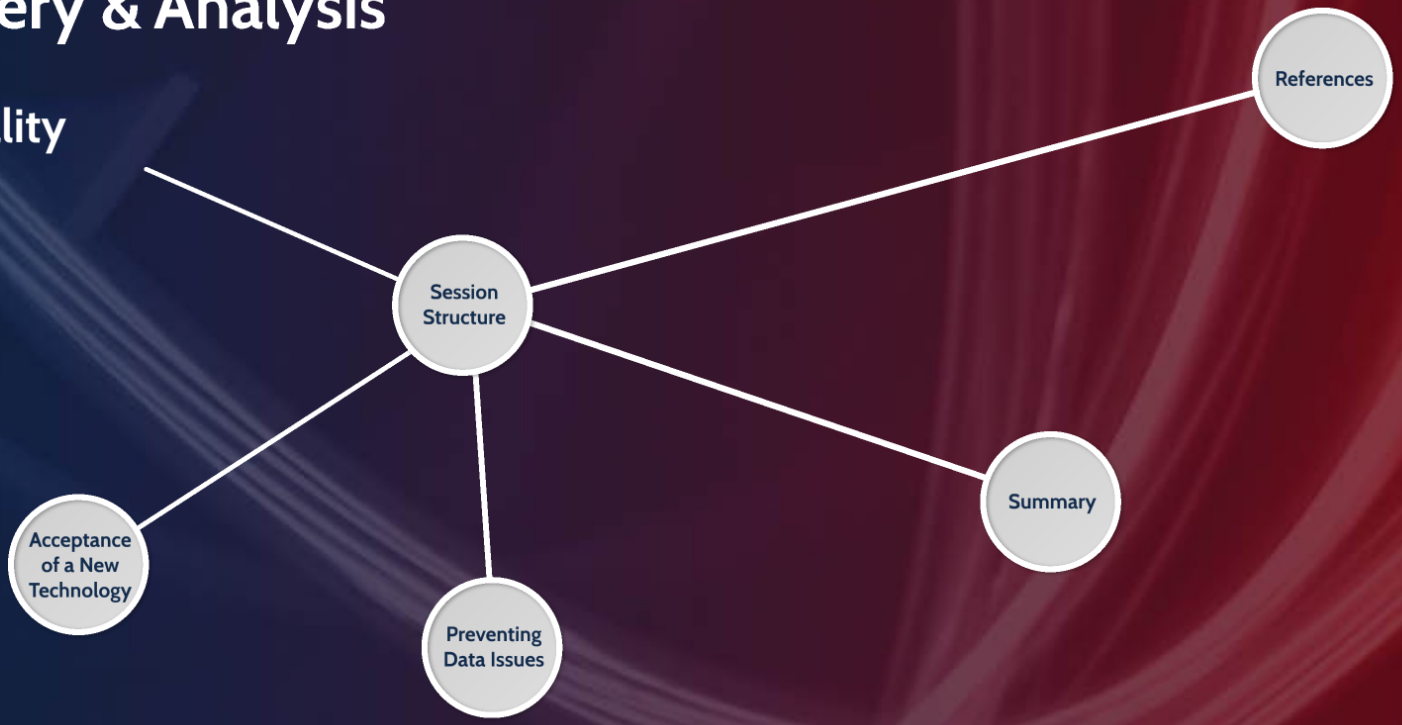
2. Van Der Aalst, W. (2016). Process Mining: Data Science in Action (2nd ed.). Heidelberg: Springer.

3. Van Der Aalst, W. M. (2011, April). Process mining: discovering and improving Spaghetti and Lasagna processes. In 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) (pp. 1-7). IEEE.

Process Analytics: Turning Data Into Insights

Discovery & Analysis

Data Quality



Friedrich-Alexander University, Chair of Digital Industrial Service Systems