Data standard for process mining

ICPM 2021- XES workshop
Elham Ramezani, PhD
KPMG Centre of Excellence for Process Mining
Current challenges with input data for process mining

Data sourcing
Data is scattered across multiple systems/sources.
Complex data structures without documentation.

Data extraction

Data processing
Example challenge: Workflow data not sufficient enough

Key Challenges:
- High process Cycle time, Missed SLA
- Lost business due to unsatisfied stakeholders
- High manual efforts
- Low standardization and automation
- Lack of visibility into skill-gap areas

Our Solution:
Enhanced approach combining Process Discovery, Process Analytics and NLP capabilities for Structured and Unstructured data analysis helped to uncover more granular inefficiencies and key root causes.

Key Considerations:
- Attrition - Loss of tacit knowledge
- Distributed work environment
- People don’t remember specifics
Current challenges with input data for process mining

Data sourcing
- Data is scattered across multiple systems/sources.
- Complex data structures without documentation.

Data extraction
- Big volume of data.
- Continuous data extraction/streaming data.
- Extracting data based on object dependencies.

Data processing
Example challenge: Extracting data on object dependencies

Warehouse Management

EU & NA

10+ systems

60 prod. sites

Warehouse Item

System B

Extract equivalent range

Join after extraction

System A

Transport Item

Extract a range of keys
Min. & Max. of the range

Potential challenges for continuous data ingestion

- Extract more/less data than needed
- Delay in data preparation
- Some manual steps need to be done
- Extensive reconciliation required to ensure completeness and accuracy of data
Current challenges with input data for process mining

Data sourcing
- Data is scattered across multiple systems/sources.
- Complex data structures without documentation.

Data extraction
- Big volume of data.
- Continuous data extraction/streaming data.
- Extracting data based on dependencies.

Data processing
- Distributed event log creation.
- Complex data pipelines.
- Inconsistent data formatting and logging mechanisms.
- Continuous data ingestion to PM tools.
- Missing information about relations between data points.
Example challenge: Complex data pipeline

- D&A platform 1
- PM platform
- D&A platform n
- Tenant A
- Tenant B
- Other source systems
- S/4 HANA CFIN
- SLT replication
- SAP ECC 1
- SAP ECC n
- Other source systems

© 2021 KPMG N.V., a Dutch limited liability company, is a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative (‘KPMG International’), a Swiss entity. All rights reserved.
Source systems currently do not generate XES

XES is hardly being used in industry unless for exchanging data between process mining tools.

Wide adoption of XES/ Event data standard

**Extensibility**
Tie event data standard with other relevant technologies for: i) distributed event data preparation ii) streaming data iii) querying event data.

**Expressivity**
Cover different event data: e.g., transaction log, message log, JSON files, etc.

**Flexibility**
i) Remove the strict notion of case. ii) Move from one big and heavy event log to ease the processing of big data volumes.

**Simplicity**
Speed up XES/ event data standard generation by developing low code data transformation tools.