

KG4EL PP FPM – Knowledge Graph Artefact to construct a federated event log

I would like to gather your opinion on my artefact I developed throughout my thesis for the construction of a federated event log given the theoretical model type I proposed by Van der Aalst (2021). The concept of FPM I explained in our last meeting and I can explain it again.

During my thesis I first interviewed you experts and based on your statements, I elicited requirements to support a process mining practitioner for the construction of a federated event log for inter-organizational comparative process mining. Here is the overview of the requirements I gathered to support the PM practitioner:

	Business Perspective	Function	Goal
R1	Increase shared vision of what attributes the EL shall contain	Overview of the EL class diagram with the target EL data model	Make sure all participants share domain knowledge which processes are to be analyzed
R2	Increase shared domain knowledge	Semantic enrichment to specify useful domain knowledge about the process and the PM project	Enable self-explanatory understanding of the target EL
R3	Integrate a Smart Data Layer for PM practitioner to run queries	Understand through an ontology the EL challenges and data quality issues	Visualize EL challenges and data quality issues through a taxonomy graph
	Technical Perspective	Function	Goal
R4	Enhance data quality by identifying missing in log (MIL)	Script functions to identify MIL	Identify MIL
R5	Enhance data quality by identifying duplicates in log	Script functions to identify duplicates	Identify duplicates in log
R6	Enhance data quality by identifying interoperability problems with the data formats	Script functions to identify data formats differences between ELs	Identify interoperability problems in the log (count the number of characters)
R7	Enhance data quality by identifying blanks and NULLs	Script functions to identify blanks and NULLs in data fields	Identify blanks and NULLs in the log
R8	Create a Python preprocessing script which can be run locally to capture the state of the event log containing R4 – R7	Create a DataFrame which captures the state of the event log	Create a script which can be run (locally) which enabling event logs to be compared
R9	Integrate a Smart Data Layer through a Knowledge Graph (Ontology/Taxonomies)	Integrate an ontology referring to the EL challenges and data quality issues	Visualize and connect the imported ELs to the data quality and EL issues captured
R10	Integrate a Smart Data Layer by adding an organizational graph to the event log class diagram	Integrate important information for the PM practitioner where to locate a human or technical resource	Contain further information on the resources linked to a specific event log
R11	Identify if the event logs fit the target data format	Query the different data structures of the event logs through the relationships	Check how many data entities fit the target data type

R12	Identify if the event logs can build a common data model	Query the different data structures of the event logs for a common data model through the relationships and properties	Identify which datatype is most common between all event logs to determine the target data type (tentative event log construction)
-----	--	--	--

The use case is similar to the intra-organizational comparative process mining initiatives at Siemens to calculate the lead time of the supply chain in different locations, because the ERP and source systems all had different configurations. Here is the idea of my developed concept:

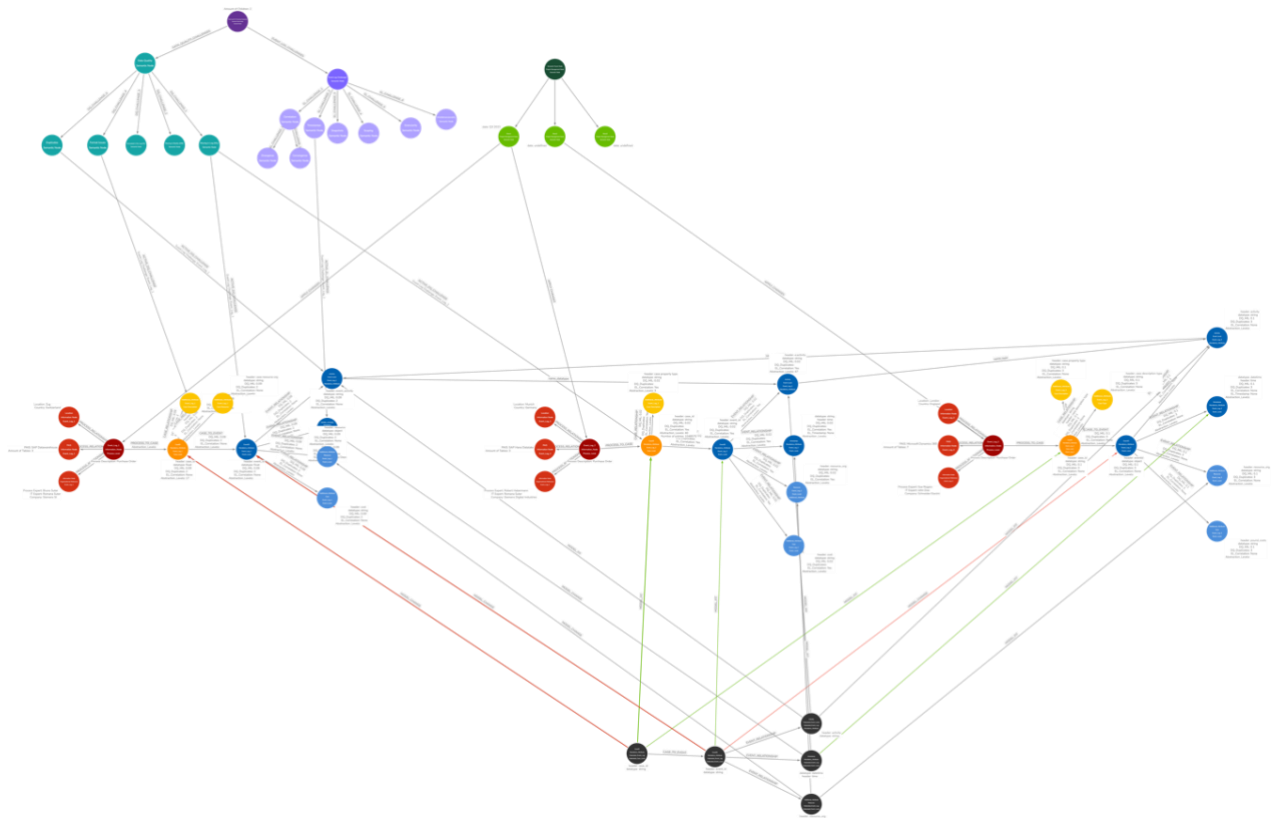


Figure 1: Overview describing the KG4EL PP FPM concept

A key aspect of my artefact was the development of a script to have an understanding of the state of a given event log in Python. The event log is then visualized in a graph model as a knowledge graph.

After this demonstration I would like to ask you to give me feedback regarding the artefact.