

<b>Name:</b> Paulo Rúben LOPES QUEIRÓS <b>Level 2:</b> Mathias VANNESTE <b>Path:</b> <a href="#">L1_PauloLopesQueiros/L0_AliZORA</a>
--

## Summary:

In business process modeling (BPM), there has been significant progress in enhancing data integration within models, driven by both academic research and industry demands, reflected in tools such as RAW-SYS, or other adaptive object-oriented (AO-BPM) methods.

## Example:

RAW-SYS is a framework for formalizing processes, actions, and data models in computer programs, enabling rigorous verification of system behavior, particularly focusing on complex data interactions and program interconnections (cf. runtime behavior). RAW-SYS formalizes how processes work, what actions they take, and the data they use. It identifies three types of bounded information: global data store size, local data store size, and the number of concurrent cases.

RAW-SYS exemplifies an AO-BPM approach. When it comes to implementing this method, it requires companies to adapt existing structures, which is most of the time the main challenge as changes have impact on all the operations. Previous studies analyzed the importance of improving BPM modularization but lacked empirical examination.

To facilitate the adaptation onto AO-BPM characterized as sets of interconnected activities aimed at transforming inputs into outputs, explainable artificial intelligence (XAI) models have been developed. These models enable human users to comprehend and place trust in the outcomes generated by machine learning algorithms. XAI models are applied in tasks such as process mining (PM), which for example can involve binary classification, utilizing datasets for analysis and decision-making purposes.

## Example:

A hospital dataset who contains predominantly duplicate data, presents a challenge typical of predictive maintenance (PM) datasets with high duplication rates. This lack of diversity within the dataset complicates machine learning tasks. Machine learning algorithms rely on diverse data to learn patterns effectively. When a dataset is dominated by duplicates, it limits the algorithm's ability to generalize and make accurate predictions.

From a AO-BPM perspective, to address this issue natural language processing (NLP) models could employ contiguous sequences of  $n$  (n-grams) items, typically words or characters, extracted from a text or dataset. They are used to analyze patterns and relationships within the data to transform event sequences into vectors suitable for conventional machine learning models. Techniques such as data cleaning and deduplication can be applied to remove redundant or duplicate entries from the dataset, thereby increasing its diversity and improving the effectiveness of machine learning algorithms.

Other than practical solutions in implementing data integration within models, theoretical knowledge management solutions have also been developed for firms and their processes. The one referenced in our documents is a structured system comprising five interconnected submodels addressing various aspects of business activities and processes, with a focus on formalizing rules and/or formulas: business activities and processes are defined, and an example application demonstrates the practical implementation of the framework using formal specifications.