## On the Complexity of Process Preservation: A Case Studyon an E-Science Experiment



## **Rudolf Mayer, Stephan Strodl and Andreas Rauber**

Secure Business Austria Vienna, Austria Institute for Software Technology and Interactive Systems Information and Software Engineering Group Vienna University of Technology, Austria



- Preservation of processes emerging topic in Digital Preservation
- Business or scientific processes
- Motivation: need to re-run process in the future
- e.g. to demonstrate correct execution, to verify results...
- Preservation of processes goes beyond current approaches
- Processes are complex objects
- depend on heterogenous IT systems
- embedded in organisational and socio-economic context
- Need to describe and archive complete context

## **Music Classification Experiment**

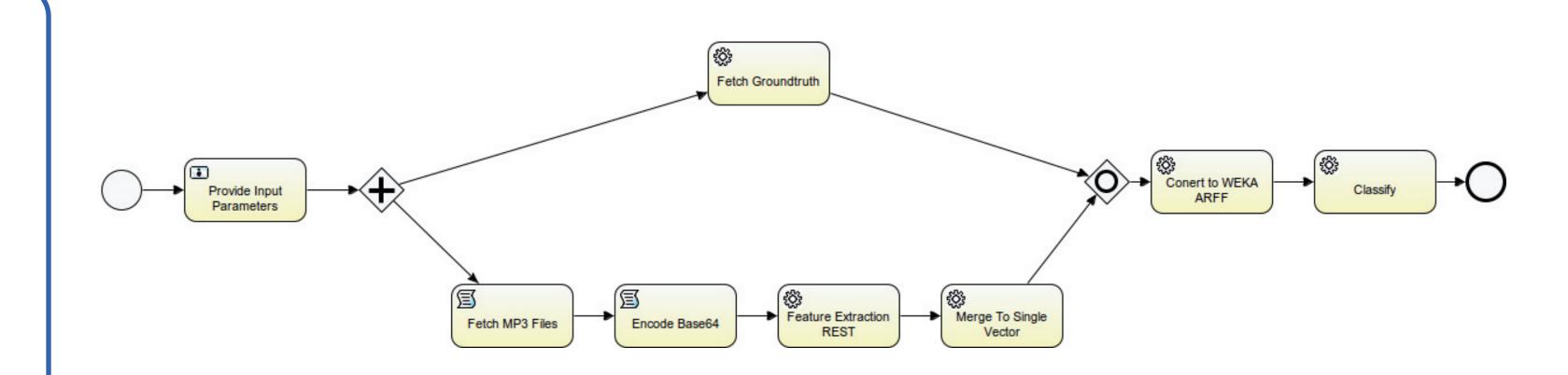
- Scientific experiment from the machine learning / IR domain
- Classification of music into predefined set of genres
- Learns a machine-learning model from given training data
- (i.e. data with manually assigned class/genre)
- Predicts genre for previously unseen data
- Useful e.g. for online music store, recommendation services, etc
- Input:
- Music (e.g. MP3 format)
- Ground truth/gold standard
- Output: Classification of music, e.g. into genres
- Intermediate steps
- Extract numeric description (features) from music

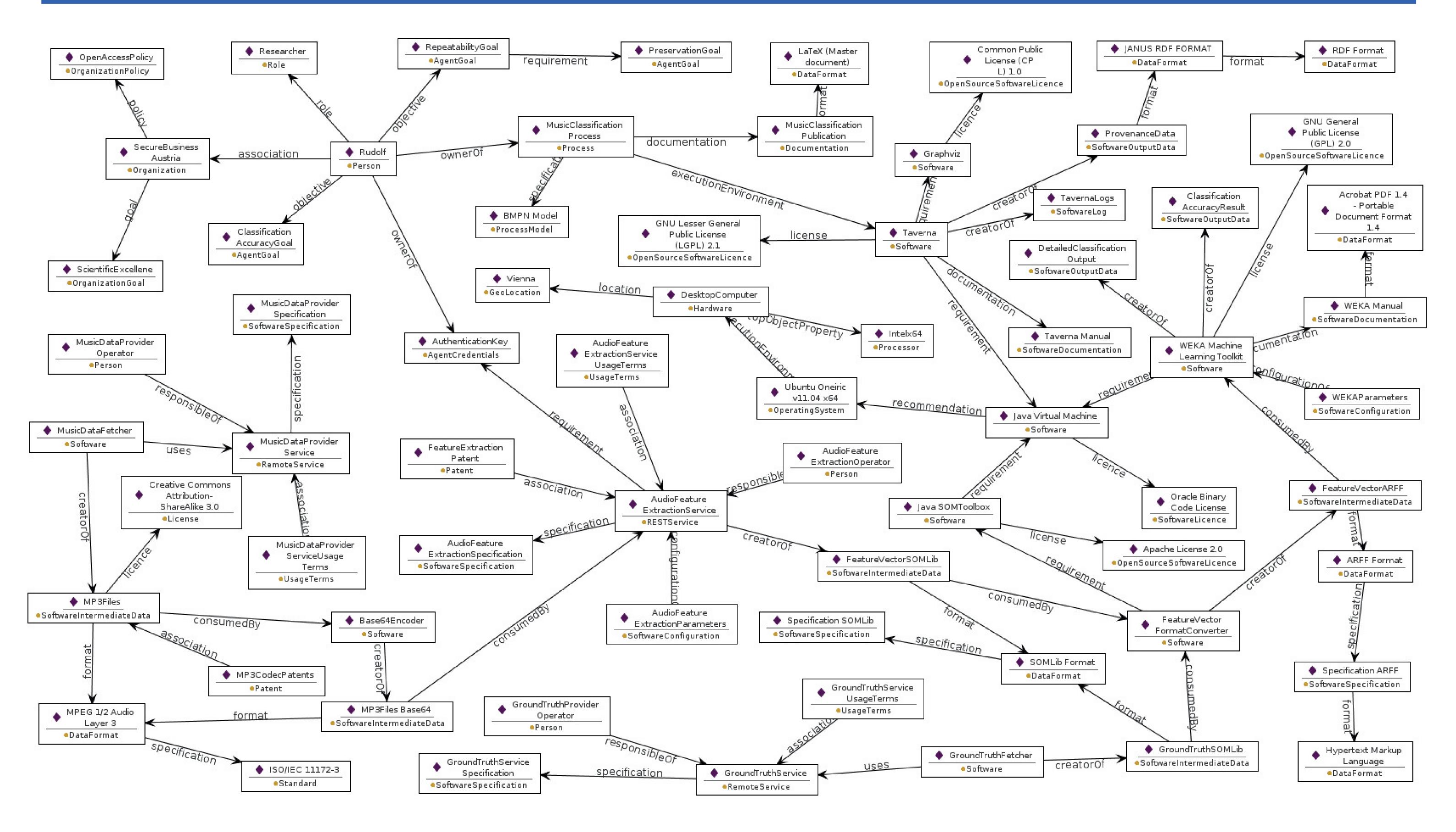
## **Important Aspects**

- Software environment including configuration (machine learning software, operating system)
- External services: feature extraction, ground truth, ...
- Hardware (e.g. computation on GPUs)
- Licenses & web-service access keys
- Experiment parameters

. . .

• Input data, intermediate data created in the process





9th International Conference on Digital Preservation, October 1 - 5 2012, Toronto, Canada