Towards a Decision Support Architecture for Digital Preservation of Business Processes

iPRES 2012 - Toronto, Canada

Alex Neumann\textsuperscript{1}, Hossein Miri\textsuperscript{1}, John Thomson\textsuperscript{2}, Goncalo Antunes\textsuperscript{3}, Rudolf Mayer\textsuperscript{4}, Michael Beigl\textsuperscript{1}

\textsuperscript{1}Karlsruhe Institute of Technology, Karlsruhe, Germany
\textsuperscript{2}Caixa Magica Software, Lisbon, Portugal
\textsuperscript{3}INESC ID, Lisbon, Portugal
\textsuperscript{4}Secure Business Austria, Vienna, Austria
Agenda

• Motivation and Goal

Introduction

• Architecture

Decision Support

• Business Processes and their Environments

Model

• What to preserve?

Use Case

• Summary and Outlook

Conclusion
Introduction

Motivation and Goal
- Business Continuity (means long-term)
  - Business depends on many processes
    - Administration, Management, Engineering, Science, ...
    - Processes describe the behavior of entities in domains
      - Define the context in which digital objects are generated, manipulated, or executed.
  - Processes are at risk of going down

- Time-Resilient Business Processes
  - Enabled by digital preservation of processes
    - Preserve all relevant processes and their relevant context
      - Technical infrastructure and standards
      - People (roles, concerns, responsibilities, ...)
      - ...
Preservation Process

Process Preservation Decision Events
• When to preserve, and why?

Process Preservation Planning
• What to preserve, and why?

Process Preservation Execution

Process Re-Deployment
• What to re-deploy, and why?
- Involved arguments need to be preserved besides decisions
  - Effective tool in business
- Answering problems is (in worst case) process-specific
  - What to preserve?
    - Different process dependencies are relevant
- Answer problems as automated as possible for cost-efficiency
  - Problems are frequently evaluated
One Generic Tool

- Interpret challenges as (generic logic-based) reasoning problems
  - Process-specific formulation of problems
  - Logically explainable solutions using a generic, sound and complete logic calculus

- One tool, which is easy to preserve
  - Operates on a well-documented knowledge representation and according to a well-documented calculus
  - Future will benefit from our system being able to explain its (process-specific) reasoning on a step-by-step base
    - e.g. based on inference rules, as a forward chainer would
Decision Support

Architecture
Decision Support Architecture

- Context Model (Instance)
  - Describes business processes and their context
  - OWL 2 DL proposed for potentially required expressiveness
- Formulate problems on these models
  - OWL 2 DL
  - SPARQL
- Apply off-the-shelf reasoners (and solvers) to solve problems
  - Pellet
  - APT-BPO
Model

Business Processes and their Environment
Context Model

- Describes business processes and their context
- Processes
  - Time Condition/Event Structures (Time Petri Nets)
    - Models causal flow and temporal constraints
    - Aligned with other models, such as WF4Ever (wrt causal flow)
    - Aligned to process models in provenance models, such as SHAMAN Context Model and PREMIS (wrt causal flow)
- Context
  - Ontology (proposed OWL 2 DL)
    - Model classes, individuals, relations and rules as generic process context framework
    - Aligned with other models, such as SHAMAN CM
- Get training and test data
- Get ground truth
- Extract features
- Build classifier and perform classification
Process Context (1)

- Process Specification
  - Inputs and Outputs
  - URLs, Files, Documents, Streams, Constants, Classifier, Classifications
- Services (purple/blue)
- Software
  - Platform (Taverna)
  - Libraries (WEKA, SOMLib)
- Specifications
  - ARFF, REST, HTTP, HTML, MP3, Algorithms
Process Context (2)

Processes

Services

Software

Hardware

...Other aspects...

SOMLib

Taverna

JVM

WEKA
Use Case

What to preserve?
- Process-specific notion of what is required by a process to be preserved and successfully re-deployed

- Requirements scopes
  - Entire domain of process preservation
    - Based on a repository
  - Sub-domains of process preservation
    - Based on a repository, or a digital preservation engineer

- Requirements types
  - **Required conditions**
    - What dependencies are at least **required** for re-deployment?
  - **Completeness conditions**
    - What dependencies need better to be **complete** for re-deployment?
Music Classification Process

- Model instance built using
  - Data extractors/crawlers
  - Expert knowledge
- Process preservation requirements
  - Need causal behavior equivalence only
    - **Required conditions**
      - Preserve anything but HW
    - **Completeness conditions**
      - All related software is preserved
What is required to be preserved?
- Preserve anything but hardware
  - Software, Specifications, Services, Inputs/Outputs

Is the model complete?
- All related software is preserved
  - SW completeness condition not yet satisfied
• Populate the context model
  • Extraction tools
  • Expert knowledge

• Specify requirements of preservation setting
  • Required conditions
  • Completeness conditions

• Use reasoning engine
  • Requirements satisfied?
  • What parts of the model are required to be preserved?
  • Are these parts complete?

• Inspect and adapt the model to requirements

• Address required conditions
• Address completeness conditions, and sign off model parts

Workflow
Conclusion

Summary and Outlook
Generic architecture to assist in digital preservation of business processes
- Based on off-the-shelf reasoners (and solvers) that operate according to generic logic calculi
  - Logically explainable answers to problems
  - Tools and arguments are easily preservable
- Addressed three decision support problems (using this architecture)
  - Talk
    - What to preserve? (Pellet reasoner, N2EXPTIME-complete)
  - Paper
    - When to preserve? (Pellet reasoner, N2EXPTIME-complete)
    - What to re-deploy? (APT-BPO solver, NP-complete)
Problem: Expressiveness of our ontology language (OWL 2 DL) poses high computational complexity in ontology reasoning
  - Satisfiability and querying is N2EXPTIME-complete

Future: Evaluation of the architecture in use cases of the TIMBUS project will determine
  - Practical feasibility of OWL 2 DL in the domain of business process preservation
  - Whether we can restrict expressiveness of used language
Towards a Decision Support Architecture for Digital Preservation of Business Processes

iPRES 2012 - Toronto, Canada

Alex Neumann¹, Hossein Miri¹, John Thomson², Goncalo Antunes³, Rudolf Mayer⁴, Michael Beigl¹

¹Karlsruhe Institute of Technology, Karlsruhe, Germany
²Caixa Magica Software, Lisbon, Portugal
³INESC ID, Lisbon, Portugal
⁴Secure Business Austria, Vienna, Austria