A Modeling Methodology for Scientific Processes

Stefan Jablonski, Bernhard Volz, M. Abdul Rehman
Applications in Geology

- **Main characteristics**
  - Many, heterogeneous systems involved
  - Data of different quality and format must be exchanged
  - Scientists are part of the applications

- **Key points:** Integration of data and human actors

- **Aspects in integration**
  - Technical: Data extraction and exchange (format, protocols)
  - Semantical: Different terminologies and ontologies of data
  - Organizational: Roles and rights of human agents

- **Existing Systems (Kepler, Taverna etc.)**
  - Working, but: Too specific, hard to extend
  - Missing: manual tasks, use of external models (e.g. data model)
  - We criticize the method that stands behind them, not their purpose or use!

**Contribution:** Not just another Information System but also a **Structured Method**

- Model the application as a process (abstract & concrete)
  - Processes cope with complex application scenarios and can be easily adjusted to changing requirements
  - Processes set up a structural framework that offers possibilities to introduce aspects of data and agent integration
- Automatically derive applications from these models which execute a process
- Leverage from existing techniques and methodologies!
PDA is a structural framework for
- Development and customization of modeling methods
- Development and customization of tool chain (modeling and execution infrastructure)
Building blocks of a process: Perspectives - Modularization modeling constructs and model

- Functional perspective - structural composition of a process
- Behavioral perspective - flow of control in between process steps
- Data & data flow perspective - data within the process
- Organizational perspective - responsibilities and roles in the process
- Operational perspective - tools and applications

Our modeling tool i<PM implements the POPM approach; advantages

- New modeling elements can be integrated easily; highly customizable
- Integration of external models (data, organizational, operational) and manual process steps
- Data integration based on ontologies, etc.

Foundation: Perspective Oriented Process Modeling (POPM)

This list is far from being complete/fixed; it can be extended or restricted depending on the individual needs of an application!
- Simple process involving human actors

Color can have a meaning:
- Type / kind of data,
- Special tasks etc.

Integration of external models:
- Definition of data (e.g. ER),
- Definition of invocation semantics,
- Organizational mapping etc.
Perspectives in the process model (Overview)

Operational perspective → Microsoft Excel

Functional perspective

Organizational perspective

Behavioral perspective (causal dependencies)

Data source

Data and Data Flow perspective

Data sink

Measure volume

Research Assistant
Data Integration: How does it work?

- Two different "modes": automatic and semi-automatic mapping
  - Depending on the ontologies used
  - Format conversions (not shown)

Automatic mapping possible if common element found

Semi-Automatic mapping if no common element found

Manually defined correspondence
Global Scenario: Process Driven Information System for Geology

Process Driven Information System for Geology

Process Library
- Analysis & Manipulation
- Im- & Export Processes
- Data Conversion
- Queries

Client Software
- i<PM Process Modeling Solution
- Customized Clients (GIS etc.)
- Reporting

i<PE Process Enactment Engine
- Service (Invocation) Library
- Ontology Transformation

Data Storage
- Scientific Databases
- Meta Repository
- Knowledge Database

Data Sources
- External & Internal

29.05.2007 © 2007 · Lehrstuhl für Angewandte Informatik IV · Bernhard Volz
Summary

- Existing scientific workflow systems provide capabilities to integrate data, but they are too specific
  - Often bound to an application domain
  - Hard to adapt/extend: Special functionality which is often needed cannot be integrated cleanly

- PDA together with POPM provides this flexibility and extensibility
  - Small amount of actual coding needs to be done, mostly "configuration"
  - Ability to extend the methodology, modeling language and the software tools is part of the method (PDA) and not a feature of one special (sub-)system
  - Example: Modeling construct for classifying data items

29.05.2007 · Lehrstuhl für Angewandte Informatik IV · Bernhard Volz