

AIIV
UNIVERSITÄT
BAYREUTH


Universität Bayreuth
Lehrstuhl für Angewandte Informatik IV
Datenbanken und Informationssysteme
Prof. Dr.-Ing. Jablonski

Process and Ontology Based Data Integration

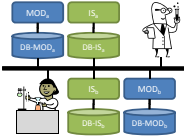
Stefan Jablonski, Bernhard Volz, M. Abdul Rehman

Lehrstuhl für Angewandte Informatik IV
Universität Bayreuth
Fakultät für Mathematik und Physik
D-95440 Bayreuth

Fon +49 921 35-7629
Fax +49 921 35-7622
email Bernhard.Volz@uni-bayreuth.de
www http://www.ai4.uni-bayreuth.de



(Distributed) Applications




- **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
 - Manual tasks, use of external models (e.g. data model) often not (well) supported
 - We criticize the method that stands behind them, not their purpose or use!

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

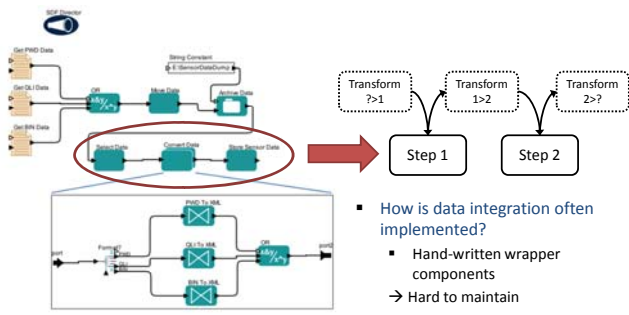
- Model the application as a process (abstract & concrete) → Perspective Oriented Process Modeling
 - Processes cope with complex application scenarios and can be easily adjusted to changing requirements
 - Processes set up a structural framework that offers possibilities to introduce aspects of data and agent integration
- Enrich processes with information necessary for data integration
 - Ontologies, mappings etc.
 - **Requirement: This information should not disguise the function/purpose of a process**

28.10.2008
© 2008 - Lehrstuhl für Angewandte Informatik IV · Bernhard Volz
2




Data Integration - Status quo?

- **Integration steps are modeled explicitly**
 - Principle: Insert an integration step in between two normal steps of a process
 - Real purpose of the process is disguised, communication is constricted



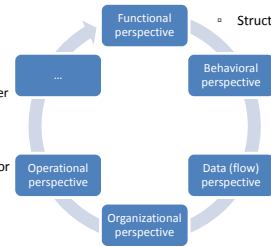
- **How is data integration often implemented?**
 - Hand-written wrapper components
 - Hard to maintain

28.10.2008
© 2008 - Lehrstuhl für Angewandte Informatik IV · Bernhard Volz
3



Perspective Oriented Process Modeling (POPM)

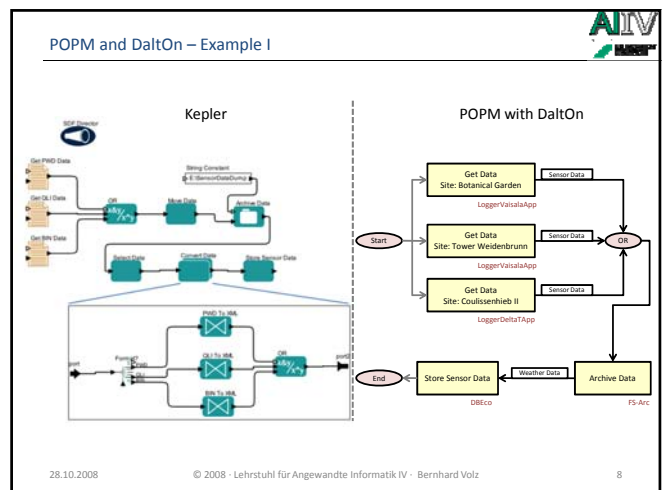
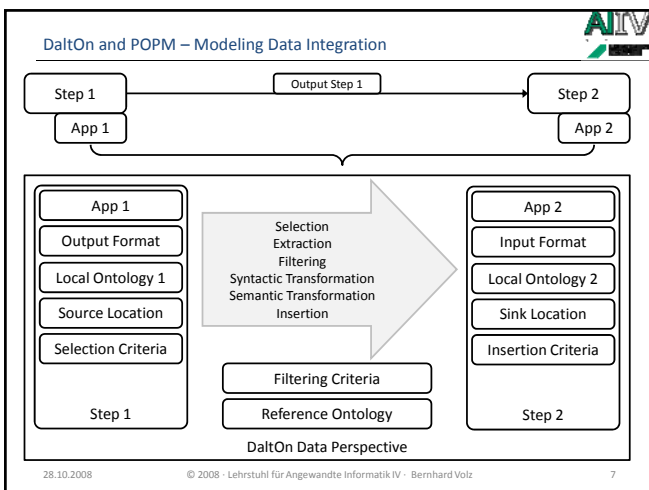
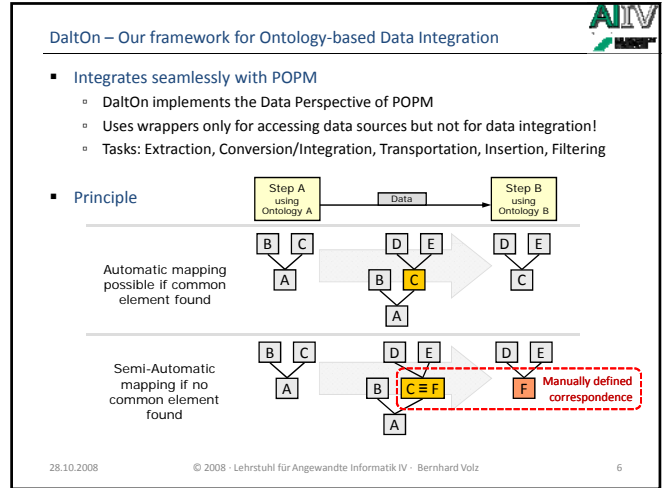
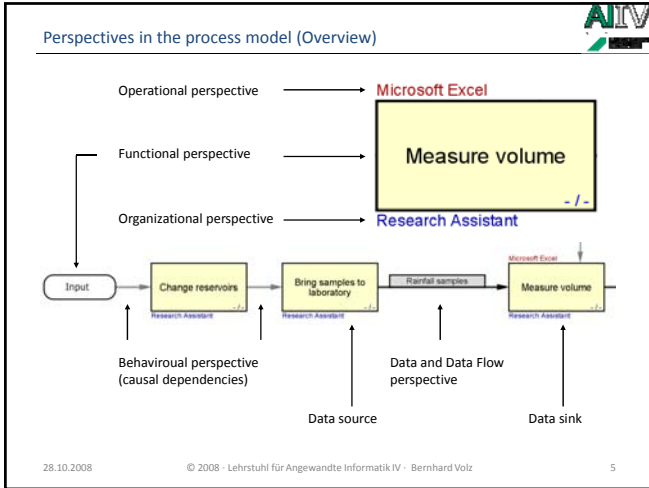
- **Perspectives are the building blocks of each modeling construct in POPM**
 - Orthogonality of perspectives – each perspective contributes information/meaning which is not yet contained in another perspective
 - Facilitates extensions greatly

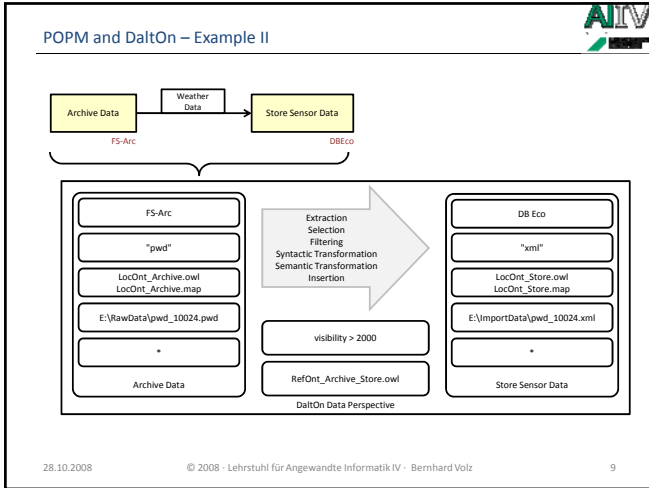


- The list is far from being complete as different application domains require additional or other perspectives
- Tools and applications for the execution of a work step

- **Functional perspective**
 - Structural composition of a process
- **Behavioral perspective**
 - Flow of control in between the work steps of a process
- **Data (flow) perspective**
 - Data within the process
 - Producer / Consumer, etc.
 - Flow of data in a process
- **Organizational perspective**
 - Responsibilities and roles

10/28/2008
© 2008 - Lehrstuhl für Angewandte Informatik IV · Bernhard Volz
4





DaltOn – Example II – Consolidation of Excel Sheets

This image shows a consolidation of Excel sheets. It features several spreadsheets with data tables and formulas. Key elements include:

- A_TOC.XLS**: A table of contents or index sheet.
- A_AES_FIA.XLS**: A sheet with data columns like 'MPDC mg/l', 'DB mg/l', 'Ca mg/l', etc.
- A_DXA.XLS**: A sheet with data columns like 'MPDC mg/l', 'DB mg/l', 'Ca mg/l', etc.
- A_ZIN.XLS**: A sheet with data columns like 'MPDC mg/l', 'DB mg/l', 'Ca mg/l', etc.

 The spreadsheets contain numerical data and formulas, with some cells highlighted in red and blue. The bottom of the image shows a summary table with columns for 'Sample', 'MPDC mg/l', 'DB mg/l', 'Ca mg/l', 'Mg mg/l', 'K mg/l', 'Na mg/l', 'Fe mg/l', 'Mn mg/l', 'Si mg/l', 'S mg/l', 'Br mg/l', 'Cl', 'NO3', 'PO4', and 'SO4'.

28.10.2008 © 2008 - Lehrstuhl für Angewandte Informatik IV - Bernhard Volz 10

- ### Summary
- Process models provide capabilities to describe complex scenarios
 - Easy to understand – No need to read code
 - Communication is supported
 - Existing scientific workflow systems provide capabilities to integrate data, but they are too specific
 - Data integration usually disguises the purpose of a process (integration = insertion of at least one special work step)
 - Often bound to an application domain
 - POPM
 - Clear separation of concerns via the five basic perspectives
 - How these perspectives are presented is up to the modeling environment; unimportant things can be easily hidden (nevertheless they are present)
 - Data integration information is part of one perspective
 - DaltOn
 - Performs multiple integration tasks
 - Generic framework – based on ontological description of data; no hand-coded integration wrapper!
 - Can also be used independent from a process
- 28.10.2008 © 2008 - Lehrstuhl für Angewandte Informatik IV - Bernhard Volz 11