Open Access Workshop DESY

AN INFORMATION SYSTEM FOR RESEARCH DATA IN MATERIAL SCIENCE

THORSTEN WUEST
Agenda

1. Introduction
2. Challenges and project goals
3. Use case and data model
4. Basic modeling approach
5. Web and mobile application
6. Open questions
7. Conclusion
Introduction

- Common agreement of the **benefits of continuous (open) access** of research data for different stakeholders
- Trend towards a disconnection of the data generating entity (person) and user of the research data
- Two groups of potential research data users **AFTER** data generation:
  - 1. **observation-oriented disciplines**: Usage of the research data after the generation plays an important role
  - 2. **experimental-oriented disciplines**: (detailed) Description of the process of great importance to allow the interpretation and usage of the results
InfoSys – Challenges

- **Large amount of research data** within material sciences
- Material sciences can be located within **second group (experimental)**
  - research data is generated through **wide variation of experiments**
  - **Metadata** is of great importance (e.g., machine used, process parameters, parameters of environment, etc.)
- As of today, there is just a **small degree of standardization** for archiving practiced
  - leads to **redundancies** and therefore avoidable cost
  - **documentation/guidelines** for archiving not adequate
  - **hinders exchange** (internal & external) & availability -> problematic for other, especially unaffiliated third persons
- Research data resulting from **publicly funded projects AND industrial research assignments**
InfoSys – Project goal

- **Large amount of research data** within material sciences
- Material sciences can be located within *second group (experimental)*
  - research data is generated through *wide variation of experiments*
  - **Metadata** is of great importance (e.g., machine used, process parameters, parameters of environment, etc.)

*The goal of the InfoSys project is to provide long term and sustainable access to material science research data for researchers and the research community*

- *Hinders exchange* (internal & external) & availability -> problematic for other, especially unaffiliated third persons
- Research data resulting from *publicly funded projects AND industrial research assignments*
InfoSys – Project goals (detailed)

- Transforming today’s heterogeneously and decentralized data management in an **efficient and centralized data infrastructure**
- Integration of **existing standards** (e.g., for experimental design) in standards for data management and archiving
- **provide long term and sustainable access** to material science research data
- Link the developed solution to **existing systems** (e.g., libraries, data bases, etc.)
- **Implementation** of the solution at the use case provider IWT
- Additionally implementation of developed solution at an **external institution** to prove the interoperability (exchange)
- Ensure the research data can be **found and referenced** (e.g., through DOI)
InfoSys – Project partners

- **Stiftung Institut für Werkstofftechnik (IWT)**
  - Provider of the use case as a material science institute
  - Knowledge of requirements and contact to community
  - Host of the research data infrastructure

- **Uni Bremen – Institut für integrierte Produktentwicklung (BIK)**
  - Experienced in requirements engineering and knowledge management
  - Experts in semantic data integration and data infrastructures
  - Long lasting cooperation with IWT (e.g., CRC 570)

- **Uni Bremen - Technologie-Zentrum Informatik und IT (TZI)**
  - Experts in development and design of web-based information systems
  - Information capturing, archiving and retrieval
  - Development and maintenance of ontology and semantic data handling
  - Experts in usability and user centric design
Tensile testing – first prototypical use case


http://info.admet.com/Portals/70514/images/machine.jpg

http://mee-inc.com/hamm/HAMM%202006%20300dpi_img_47.jpg
Data model from user perspective

**Project leader (researcher)**

Generate project
→ ID, name, project, authorization, etc.

- a) Define material
  - manufacturer
  - alloy
  - heat treatment
  - semi-finished part
  - purity
  - etc.

- b) Define experiment
  - testing method
  - metadata
  - basic conditions
  - dimensions
  - type of machine
  - etc.

- c) Enter results
  - Sample dimension
  - sample amount
  - force
  - tension
  - etc.

**Entry masks**
- „machine book“
- „experimental protocol“
- etc.

**User definable views**

- conclusions
  - experimental program
  - material
  - etc.

- links to:
  i) presentations; ii) publications;
In all „templates“ the possibility to add documents (word, pdf) is necessary

**State/quality**: e.g., raw data, preliminary test, published, checked

**Access authorization**: individually adjustable for persons and/or groups

**Access levels**:
- low
- high

Hamburg, 10. June 2013

Information system for research data in material science

Thorsten Wuest
Basic modeling

- **Requirements**
  - searchable and expandable databank schemata
  - formal model with higher inference possibilities

- **Goal:** databank AND formal ontology

---

**Foundational Ontology**

**Ontology of Plans**

**Descriptions & Situations**

**InfoSys: Ontology of Scientific Experiments (OSE)**

---

**Contact:**
Dr. Robert Porzel
porzel@tzi.de
Link to DOLCE

- DOLCE – Top-Level
A project is a proactively satisfied method. Additionally to a plan, a project includes at least one 'product' role to be played by some endurant (e.g. a material), or one 'result' role played by a perdurant with a definite participant (e.g. a changed state of a material).
Two main actors with different needs

**Researcher**
- Define material
  - manufacturer
  - alloy
  - heat treatment
  - semi-finished part
  - purity
  - etc.
- Define experiment
  - testing method
  - metadata
  - basic conditions
  - dimensions
  - type of machine
  - etc.
- Enter results
  - sample dimension
  - sample amount
  - force
  - tension
  - etc.

**Technician**

**Project leader (researcher)**

Generate project
→ ID, name, project, authorization, etc.

**User definable views**
- Entry masks
  - “machine book”
  - “experimental protocol”
  - etc.
- conclusions
  - experimental programme
  - material
  - etc.

**Links to:**
- Presentations; publications
- In all “templates” the possibility to add documents (word, pdf) is necessary

**State/quality:** e.g., raw data, preliminary test, published, checked

**Access authorization:** individually adjustable for persons and/or groups

**Access levels:**
- High
- Low
Development of front end

- **Researcher**
  - Input of large amount of data
  - Accessible from all computers
  - Transfer of data from/to other applications
  - **Web application on desktop PC**

- **Technician**
  - Today: machine book
  - Documentation of experiment design
    - Extensive input: text, photos, drawings, etc
  - Documentation of individual experiment
    - Fast and easy interaction
  - **Mobile application on tablet (All-in-one)**
Screen-shots - web application

dashboard

list of projects

new project

new material
Screen-shots - mobile application

new experiment

Specify results

send to database

English/German
Possibility to take picture of experimental design or drawings for documentation
Open access

- Project is reaching the end of the first phase
- Next phase the prototypical system will be implemented and *tested and continuously improved* with users from the material science community in their *day-to-day routine*
- Then system will be disseminated with external users at another material science institute to ensure the *interoperability*
- However, a lot of *questions*, especially concerning *open access*, have yet to be answered:
  - Quality check and open access
  - Open access and how to make sure sharing is mutual?
  - Industrial research assignments and open access
Quality check and open access

- Experiment results can determine how certain elements of e.g., a plane are designed, those parts being highly security relevant
- What happens when such a system fails due to flawed data?
- Who is responsible?
- How can the quality be ensured?
  - By declaration of the author?
  - Peer review for data?
  - External agencies (e.g., TÜV)?
  - By the users themselves? Maybe employing social features (e.g., like - facebook, review – amazon)
Open access - sharing mutual?

- Generating large amounts of data is very expensive in material sciences

- This is one of the main reasons why the public should profit from the data (at least if publicly funded)

  *But can that be a one way street?*

- With this issue we touch political topics, like the national economy and competitive advantage -> WHO is public?

- is open access for everyone the way to go or mutual agreements in groups which share information (e.g., through point system)?
Industrial research assignments and open access

- Most of the time a **disclosure agreement** has to be signed
- Legal repercussions are immanent if such research data is leaked

As a non-profit organization (like most institutes) that can cause serious **financial AND reputational trouble**

**Trusted system is needed** to protect some data while making other data publicly or selectively available
To this point, the project and the first results are accepted well within the material science community and created a lot of interest.

However, there are still many big challenges to be faced.

Some are specific for the material sciences (and engineering)
- Liability issues if data leads to failing products

Others concern the whole community of researchers
- Data security (medicine/social sciences?)
- Quality management for research data

Interdisciplinary exchange is needed!
Thank you for your attention!

Any questions?

Contact: wue@biba.uni-bremen.de

Acknowledgement

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen der Ausschreibung „Informationsinfrastrukturen für Forschungsdaten“ gefördert.

The authors gratefully acknowledge the financial support of the German Research Foundation (DFG) within the Call “Information Infrastructure for Research Data”.