Digital Libraries 2014

Creating lightweight ontologies for dataset description:

Practical applications in a cross-domain research data management workflow

João Aguiar Castro **joaoaguiarcastro@gmail.com**
João Rocha da Silva **joaorosilva@gmail.com**
Cristina Ribeiro **mcr@fe.up.pt**

London, September 2014

CONTENT

- Research Data Management
- Lightweight ontologies
- Case Studies
- Modelling Process

RESEARCH DATA MANAGEMENT

Research Data Management comprise many challenges…

Resource description is a particularly difficult one

- Domain diversity
- Massive data creation
- Datasets produced by countless small research groups - lack of resources
- Absence of timely description can yield vague descriptions

RESEARCH DATA MANAGEMENT

Metadata schemas are complex which make them hard to adopt

The description process it too time-consuming and divert researchers from their main activities

Application Profiles are self-contained

For datasets in a fast-paced, multi-domain research environment, a more incremental approach is desirable

LIGHTWEIGHT ONTOLOGIES

Lightweight ontologies as a solution for dataset description

- ✓ Capable of representing the semantics of each research domain
- ✓ Evolve asynchronously
- ✓ Avoid complexity
- ✓ Easily manageable by data curators
- ✓ Easily processable by machines

The owl version of Dublin Core and Friend of a Friend are examples of two widely used lightweight ontologies

CASE STUDIES

MECHANICAL ENGINEERING – FRACTURE MECHANICS

Double Cantilever Beam experiments that study the resistance of a given specimen subjected to pressure.

ANALYTICAL CHEMESTRY – POLLUTANT ANALYSIS

Routine analysies regarding the concentration of certain pollutants in a sample, collected at a given time and place, named *Runs*.

MODELLING PROCESS

Providing researchers with a set of meaningful descriptors

Meetings between

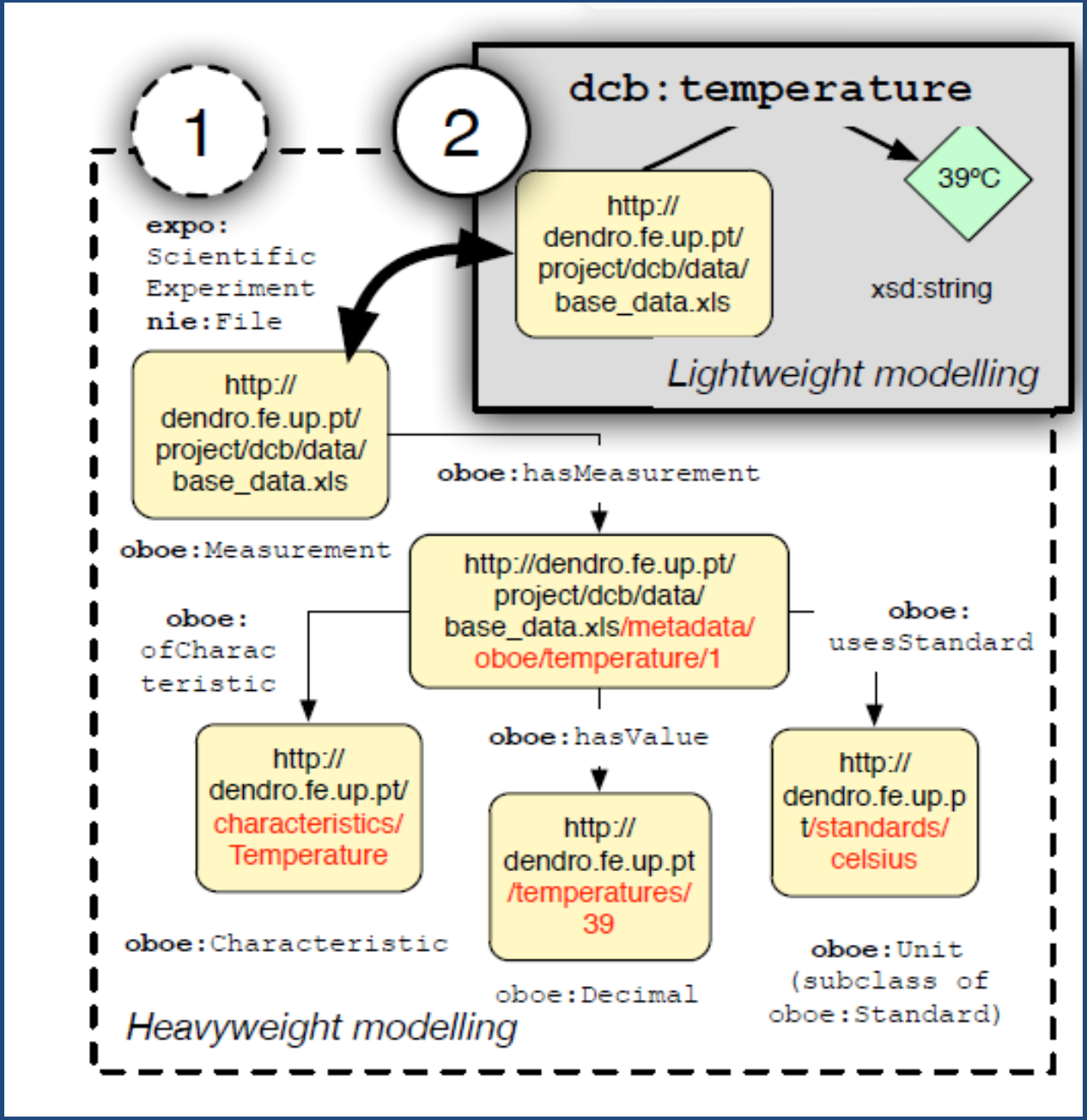
the curator (data expert) and data creators (domain experts)

- 2 formal meetings, about 1h30m each;
- Introduce basic data management concepts – “metadata”, “descriptor”;
- Suggestion of descriptors from different metadata schemas;
- Identification of domain-specific concepts;
- Application Profile design.

Assess data management practices – Script adapted from the Data Curation Tool Kit

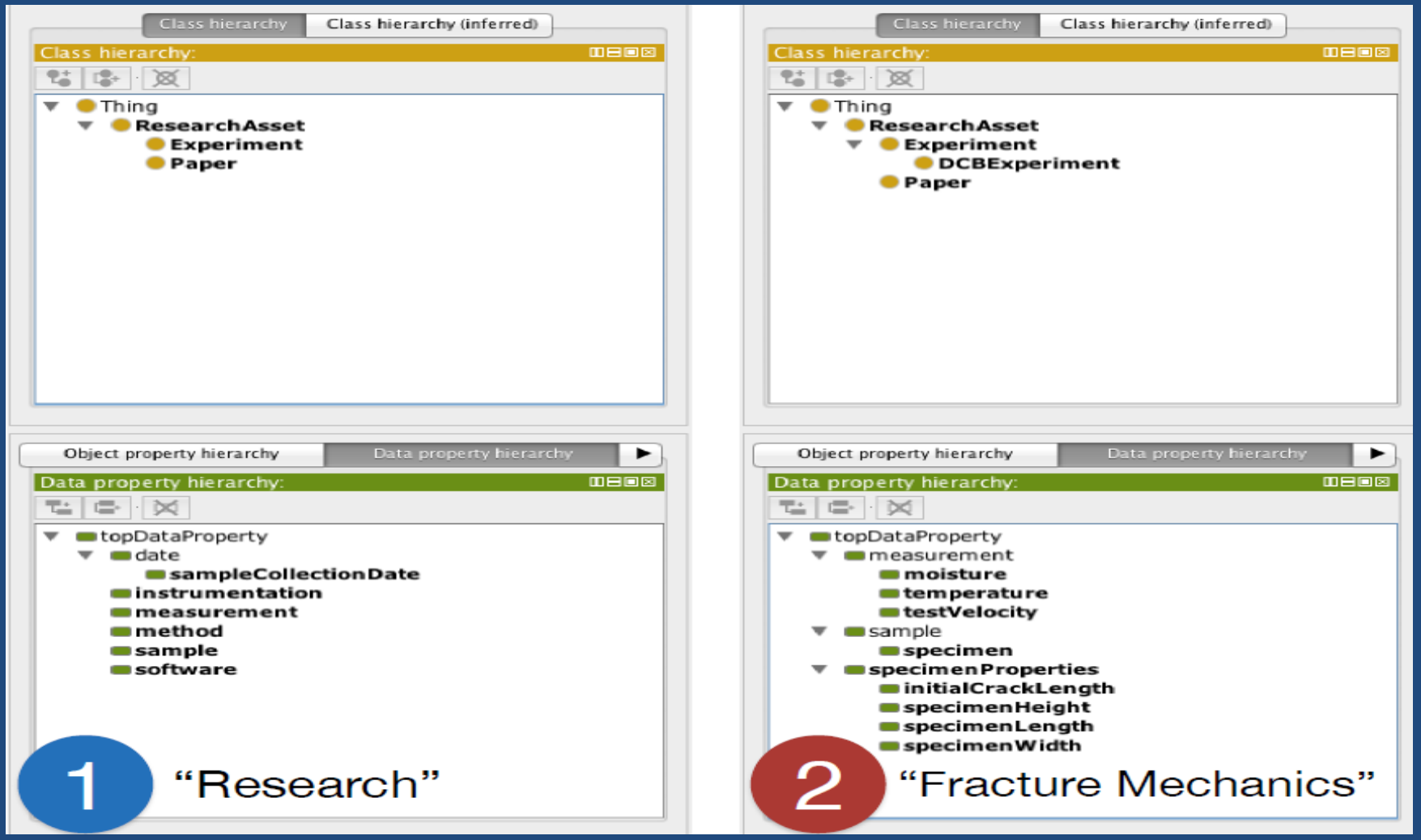
Pardue University & University of Illionais at Urbana-Champaign

MODELLING PROCESS



Both approaches can co-exist, via ontologies relations, if the need should arise

MODELLING PROCESS

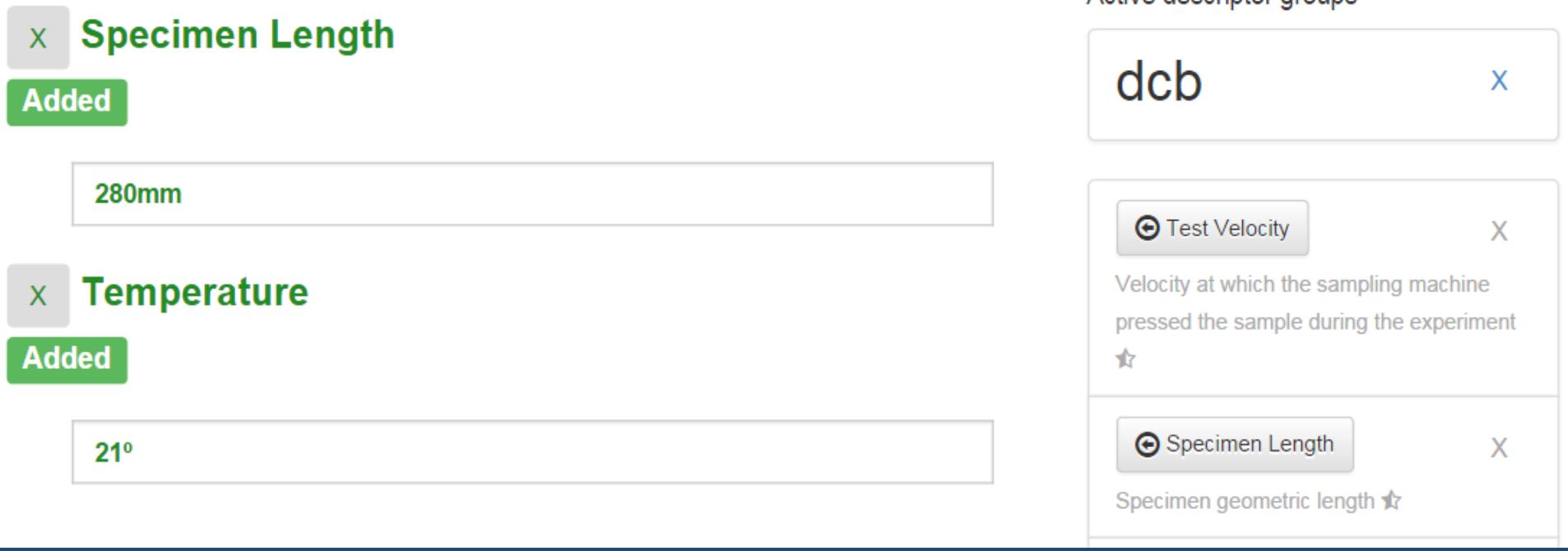


Research – a generic lightweight ontology

- Match the *File-Folder* representation of datasets
- Extension point from which other domain-specific ontologies can be derived

Our ontology-based Research Data Management Platform

The properties defined in our ontologies were loaded and used as a source of descriptors to demonstrate its applicability to a scenario where researchers describe their own datasets



João Aguiar Castro joaoaguiarcastro@gmail.com
João Rocha da Silva joaorosilva@gmail.com
Cristina Ribeiro mcr@fe.up.pt

End