Comparing Application Profiles and Ontologies for describing Scientific Data

XATA 2011
Contents

• Introduction
• Diversity - a challenge for Data Curation
• The Pollutant Analysis workflow
• Representing Experimental Data
  ▶ Using Application Profiles
  ▶ Using Ontologies
• Conclusions & Future Work
Introduction

• E-Science generates large datasets
• Stronger investment on data production
• Research grant providers now require publication of base data (e.g. NSF)

Increasing need for Scientific Data Curation
Challenges for Data Curation

Social

Technical

Data Diversity

Ethical

Political
The pollutant analysis workflow

<table>
<thead>
<tr>
<th>Step</th>
<th>Resulting Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gather Samples</td>
<td>Methodology description</td>
</tr>
<tr>
<td>2. Experimental Analysis</td>
<td>Instrument Parameters</td>
</tr>
<tr>
<td>3. Build Result Spreadsheet</td>
<td>Data</td>
</tr>
<tr>
<td>4. Write Report on Results</td>
<td>Conclusions Report</td>
</tr>
</tbody>
</table>
Representations
Application Profile

WPAEDML

run

experiment

results

result

dc:coverage

related_things

dcterms:references

dcterms:isReferencedBy

tested_molecules

cml:molecule

methodology

dc:identifier

formats

dcterms:format

dc:elementsGroup

This group is included as a convenience for schema authors who need to refer to all the elements.
Conclusions & Future Work

- Ontologies capture more representation information, since they preserve not only the data but also its meaning.
- XML Schemas are easier to discover and reuse than Ontologies.
Conclusions & Future Work

- Future Work includes
  - Offering an XML representation for this data on a repository solution using XML Schemas and possibly Ontologies
  - Using this XML representation to provide better querying methods for the preserved data
  - Investigating broader data models to represent datasets from different domains