Comparing Application Profiles and Ontologies for describing Scientific Data

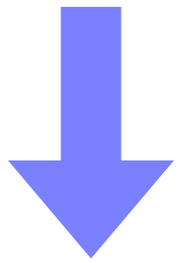
XATA 2011

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)

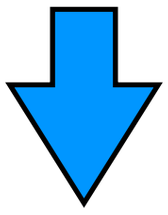


Increasing need for Scientific Data Curation

Challenges for Data Curation

Social

Technical



Data Diversity

Ethical

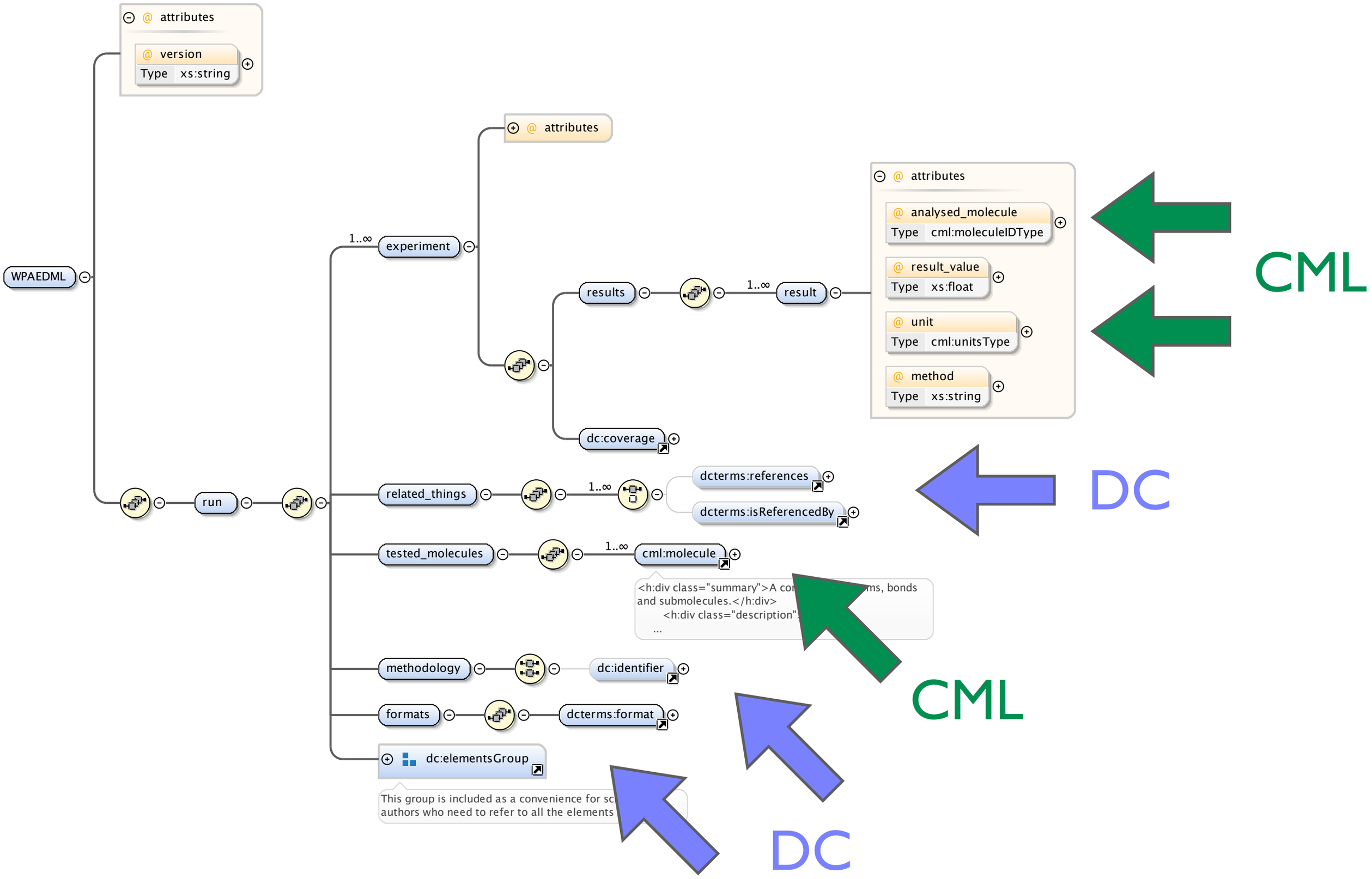
Political

The pollutant analysis workflow

Step	Resulting Asset
1. Gather Samples	Methodology description
2. Experimental Analysis	Instrument Parameters
3. Build Result Spreadsheet	Data
4. Write Report on Results	Conclusions Report

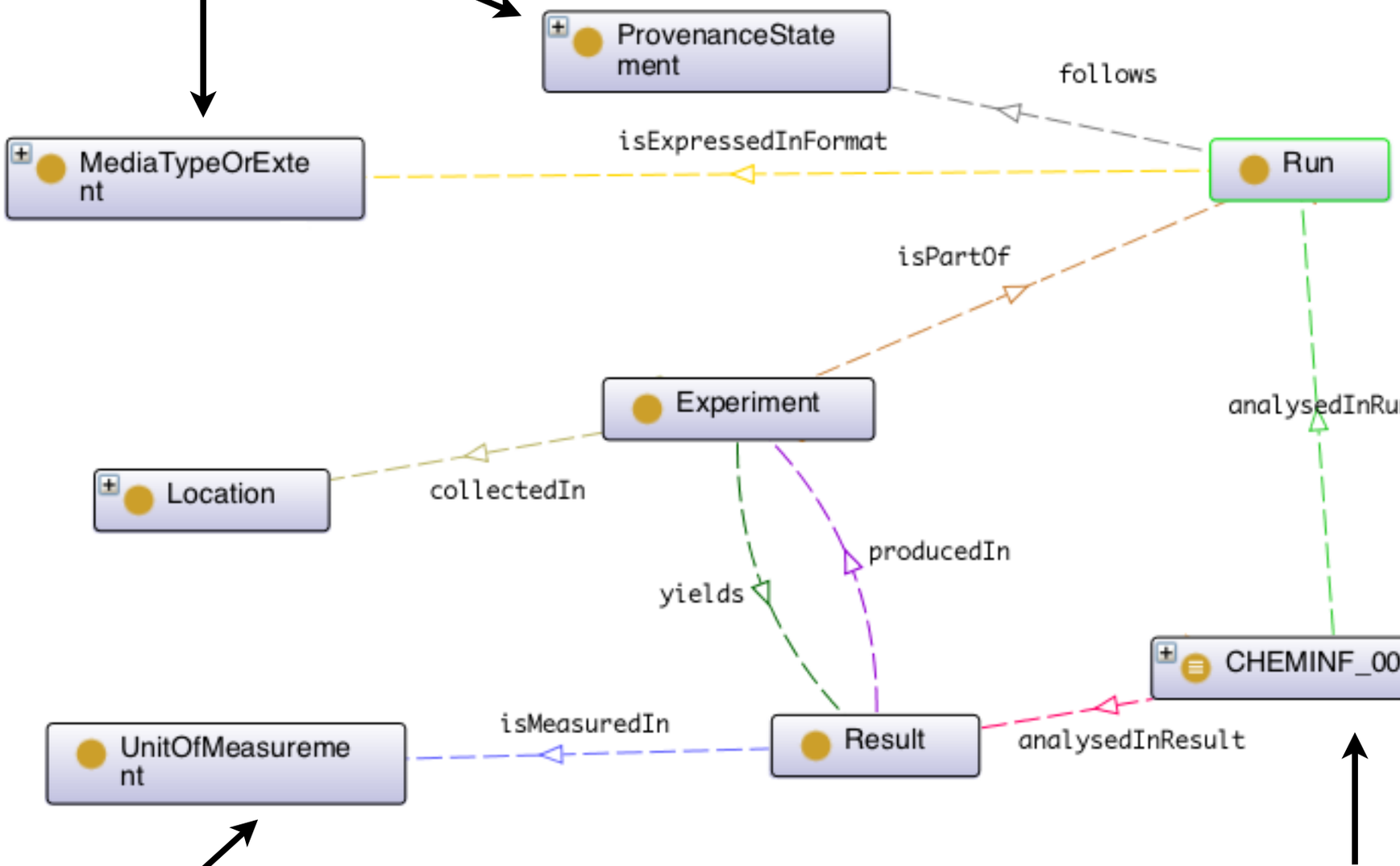
Representations

Application Profile



Ontology

DC RDF



MU Ontology

Semantic Chemistry Ontology

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