

Publishing open data and services for the Flemish Research Information Space

Christophe Debruyne, Pieter De Leenheer, Peter Spyns, Geert van Grootel, Stijn Christiaens



Vrije Universiteit Brussel

Outline

- Introduction
- Problems
- Approach: Method and Tool
- Demo
- Results
- Conclusions and Future Work

Introduction

- A good overview of a country or region's science and technology base is crucial to stimulate research and innovation
- The Department of Economy, Science and Innovation (EWI) therefore launched the Flanders Research Information Space (FRIS) program



Introduction: Goals of FRIS

- A virtual research information space covering all Flemish players in the field of Economy, Science and Innovation

The current portal
researchportal.be

Researchers

Organizations

Projects

Publications

The screenshot shows the FRIS Onderzoekspitaal website. The browser address bar displays "http://www.researchportal.be/person/robert-meersman". The page header includes the logo "EVI FRIS Onderzoekspitaal" and navigation links for "Nederlands" and "English". A search bar is visible with the text "robert meersman" fris. The main content area features a profile for "ROBERT MEERSMAN" at "Vrije Universiteit Brussel". Below the profile, there are tabs for "Overzicht", "Samenwerkingsverbanden", "Kaart", "Onderzoeksprojecten", and "Organisaties". A network diagram is displayed, showing Robert Meersman at the center, connected to other researchers: Jos PIERSON, WOLFGANG DE MEUTER, OLGA DE TROYER, and Peter SPYNS. The diagram also shows connections to organizations like "Informatie en Toegepaste Info..." and "Onderzoek naar Software Techno...".

Introduction: Goals of FRIS

- Reduce the current administrative burden of knowledge institutions
 - Repeatedly report the same information to different institutions in different formats
- Services could automate this process if data would be accessible in a uniform way

Problem

➤ Two problems:

1. What does the data mean? → Capturing the semantics of the domain in an ontology
2. How do we appropriately annotate the data (from heterogeneous) sources?

Problem

- Heterogeneous data sources
 - Knowledge institutions store information in autonomously developed information systems
 - European Commission furthermore asks knowledge institutions to report according the Common European Research Interchange Format (CERIF)
 - Knowledge “scattered” across different communities

Problem

➤ CERIF

- Created with Entity-relationship model for an almost unlimited flexibility on roles and classifications used with entities
- Show limitations when it comes to communicating the knowledge stored with that model

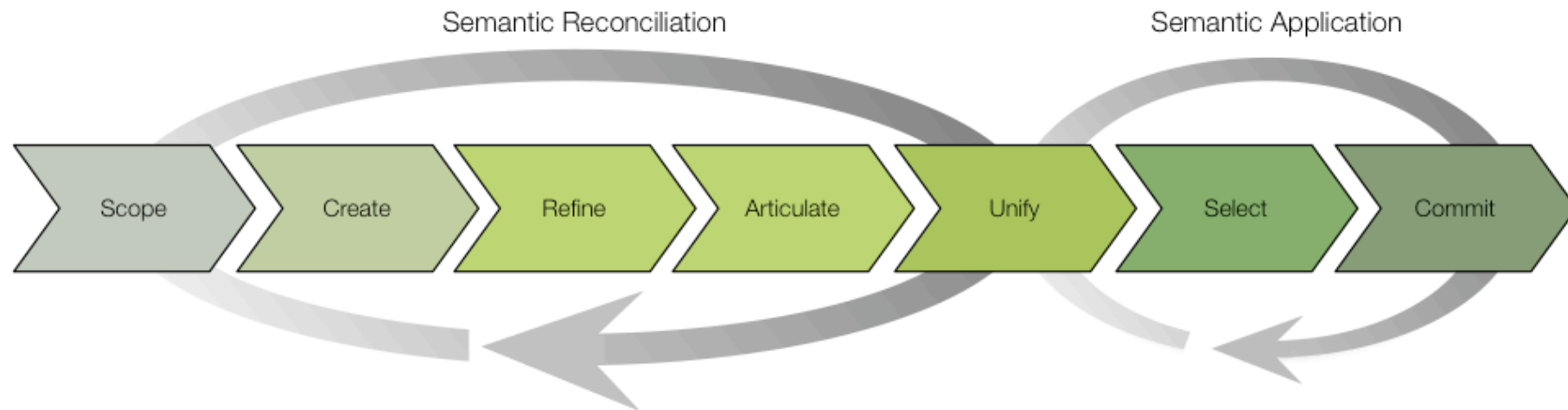
Problem

- To populate the FRIS portal with delivered CERIF and other data sources, we need
 1. A consensus amongst the involved parties – forming communities – on a common conceptual model (ontology)
 2. An easy, repeatable process for validating and integrating data
 3. A means to use that data in a generic way so services could be built

Approach: Method and Tool

- Business Semantics Management
 - Community driven ontology engineering
 - Adopts Semantics of Business Vocabulary and Rules (SBVR)
 - Fact-oriented (like NIAM, ORM)
 - Communicating (binary) facts
 - Decision attribute/relation not at design time, but at implementation time.

Approach: Method and Tool



De Leenheer, P. (2009) On Community-based Ontology Evolution:
Foundations of Business Semantics Management, PhD Thesis, Vrije Universiteit Brussel

Approach: Method and Tool

- **Business Semantics Glossary**
 - Supports Semantic Reconciliation
 - Semantic Communities → A group of stakeholders who want to achieve interoperability
 - Speech Communities → Subsets of a semantic community employing a certain terminology
 - Vocabularies → Sets of facts expressed in a terminology of a speech community + constraints capturing the intended universe of discourse

 **Project**

Flanders Research Information Space > CERIF > Project

Attributes |
 Names |
 Categorization Scheme

Definitions

Planned set of interrelated **tasks** to be **executed** over a fixed **period** and within certain **cost** and other **limitations**.

Examples

Large Hadron Collider

Fact Types

<input checked="" type="checkbox"/> Project has Title	(Candidate)			
<input checked="" type="checkbox"/> Project funded by Funding Programme	(Candidate)			
<input checked="" type="checkbox"/> Project executed by Organisation	(Candidate)			
<input checked="" type="checkbox"/> Project has Budget	(Candidate)			
<input checked="" type="checkbox"/> Project has Duration	(Candidate)			
<input checked="" type="checkbox"/> Project ends on Date	(Candidate)			
<input checked="" type="checkbox"/> Project starts on Date	(Candidate)			
<input checked="" type="checkbox"/> Project described by Discipline Code	(Candidate)			

General Rule Set


Project described by at most 3 **Discipline Code**.
 Project executed by at least 1 **Organisation**.
 Project ends on exactly 1 **Date**.
 Project has at least 1 **Title**.
 Project starts on exactly 1 **Date**.

Information

Status: **Candidate** 
 Concept Type: **Object Type** 
 Steward:  **Pieter De Leenheer** 

37.5%
 articulated

Attributes





Relations

Taxonomy



Stakeholders

 **Felix Van De Maele**
 **Administrator**
 **Geert Van Grootel**

Approach: Method and Tool

➤ Exporting OWL



Approach: Method and Tool

➤ Exporting OWL

```
- <owl:Class rdf:about="Http://labs.collibra.com/bsgtrunk/bin/view/Project/Project">
  - <Project:ProjectHasBudget>
    <owl:Class rdf:about="Http://labs.collibra.com/bsgtrunk/bin/view/Project/Budget"/>
  </Project:ProjectHasBudget>
  <dc:identifier rdf:datatype="http://www.w3.org/2001/XMLSchema#string">1f5c7dcd-9942-4313-950d-d26e7bd140c6</dc:identifier>
- <rdfs:subClassOf>
  - <owl:Restriction>
    <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</owl:minCardinality>
  - <owl:onProperty>
    <owl:ObjectProperty rdf:about="Http://labs.collibra.com/bsgtrunk/bin/view/Project/ProjectHasTitle"/>
  </owl:onProperty>
  </owl:Restriction>
</rdfs:subClassOf>
```


Approach: Method and Tool

- Annotating existing data sources
 - We used D2R Server¹, generates a “skeleton” vocabulary based on a relational database schema
 - Semantics of the exported ontology are added to the skeleton file
 - Result online → <http://starpc20.vub.ac.be:2020/>

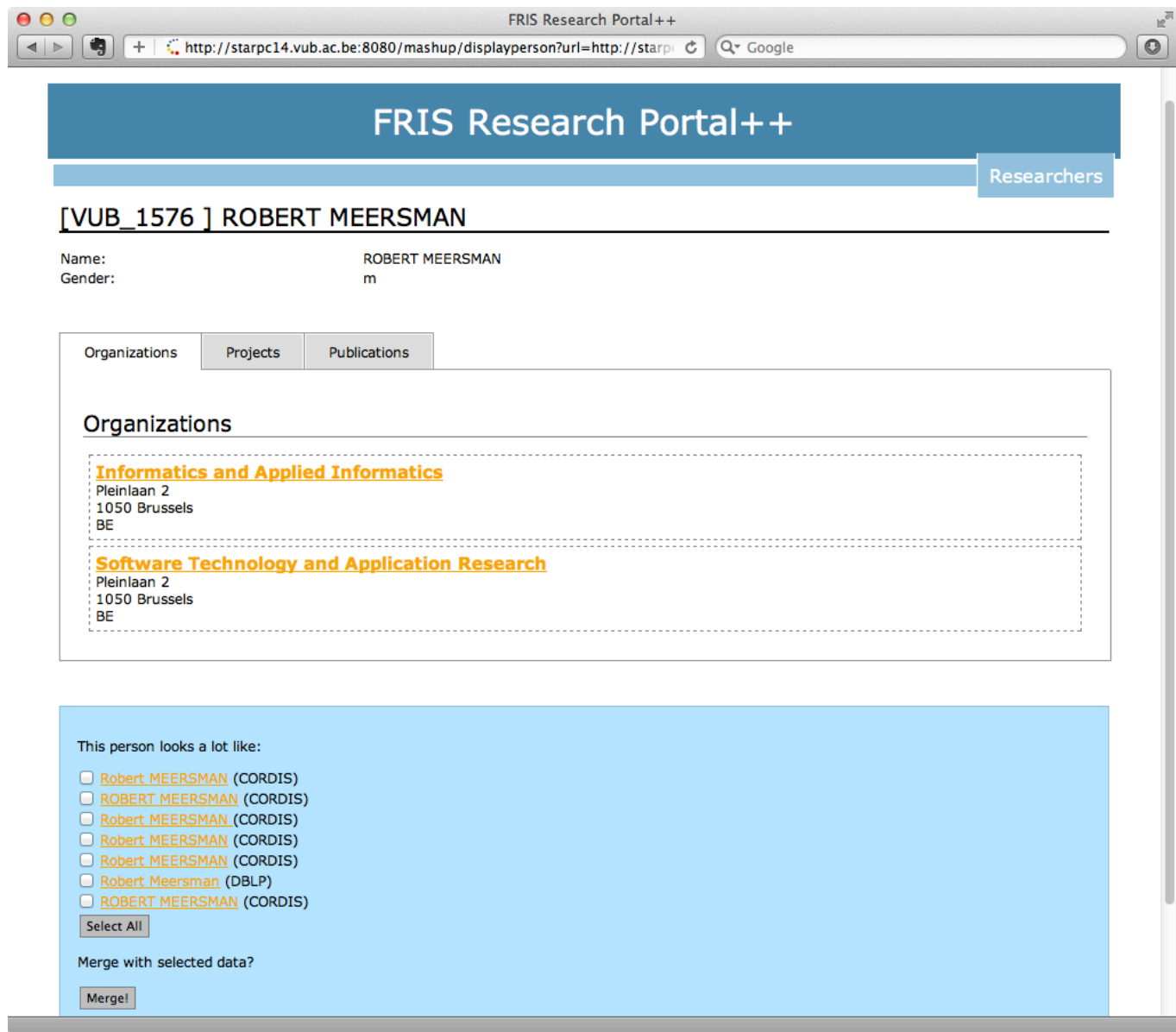
¹<http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/>

Approach: Method and Tool

- Demo of Business Semantics Glossary

Results

- SPARQL endpoint is running
Feel free to test and use it!
- Services built on top of that data
 - Results submitted elsewhere
 - For instance,
<http://starpc14.vub.ac.be:8080/mashup/person>



Conclusions

- A teachable/Repeatable process for capturing knowledge in an ontology and annotating data
- Heterogeneous data sources uniformly accessible through one ontology
- Development of services automating (reporting) tasks now possible

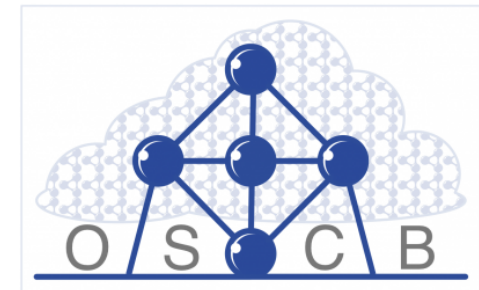
Future work

- Facilitating the annotation process. E.g., transforming Ω -RIDL annotations into D2R mappings

- Lifecycle management
<http://www.acsi-project.eu/>



- Hybrid Ontology Engineering for Linked Data
<http://www.oscb.be/>



Questions?