CROSS-DOMAIN INTEROPERABILITY OF HETEROGENEOUS RESEARCH DATA (GO INTER)
Peter Mutschke (GESIS)
BACKGROUND: DISCONNECTED DATA SILOS

- Fragmentation of research data landscape:
  - highly disciplinary and disconnected
  - heterogeneous data storage
  - heterogeneous and often unstandardized metadata and vocabularies
  - lack of cross-disciplinary and interconnected systems

- Problematic when it comes to searching and linking research data across community borders in the context of multidisciplinary research

- Cross-domain interoperability: key element to facilitate data integration across community/repository boundaries

“When data silos are bad news for everyone.”
[https://www.information-age.com/breaking-down-data-silos-123481841/]
USE CASE

Linking survey data to spatial data (DFG project SoRa)

How satisfied are you with the accessibility of green areas?

Coordinates: 50.919721, 6.967379

Sealing of Soils at Respondent’s Dwelling

Respondent’s Dwelling

Respondent’s Address

http://www.sora-projekt.de/
MAIN PURPOSE AND OBJECTIVES

- To provide a cross-domain interoperability framework consisting of methods, tools and guidelines for implementing and assessing semantic interoperability of research data across discipline borders (by building upon existing standards)
- To develop and evaluate reference implementations of interoperability for real-world cross-domain research uses cases by broadly applying existing standards, vocabularies and semantics technologies
- To engage with other GO FAIR Implementation Networks and related initiatives to disseminate and exchange best practice solutions for cross-domain interoperability
- Overall goal: to contribute to the establishment of cross-domain infrastructures that build on open standards
**SEMANTIC INTEROPERABILITY: CHALLENGES**

Dagstuhl Perspectives Workshop on „Implementing FAIR Data Infrastructures“ (2018)

<table>
<thead>
<tr>
<th>Challenge</th>
<th>How to address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many different layers of interoperability: ranging from encoding up to structural and semantic specification of data</td>
<td>(Cross-)domain specific use cases exploring interoperability could help to better understand these differences</td>
</tr>
<tr>
<td>Various metadata standards, data formats, data encoding methods, representation languages; no format validation</td>
<td>Use of standards, semantic technologies for data transformation; registries of schemas and vocabularies; Digital Object concept (RDA) for data organization</td>
</tr>
<tr>
<td>Different vocabularies to describe data; semantic interpretation varies substantially</td>
<td>Ontologies lookup services, ontology crosswalks, smart ontology mapping, tools for semantic annotations</td>
</tr>
<tr>
<td>Lack of means to link digital objects with operations suitable for their type</td>
<td>Mechanisms to link types of Digital Objects with operations (RDA data type registry)</td>
</tr>
</tbody>
</table>

[Manola/Mutschke/Scherp/Tochtermann/Wittenburg (2019): Implementing FAIR Data Infrastructures. Dagstuhl Perspectives Workshop 18472]
SEMANTIC INTEROPERABILITY: IDEAS, METHODS, SOLUTIONS

WG: Interoperability in Theory

- Challenges:
  - many repositories for vocabularies; built for different reasons; different metadata to describe semantic artifacts; not all resources available to manage ontologies
  - ontologies come in all kinds of flavours, ranging from domain specific solutions up to ‘foundational ontologies’; no information literacy of non-specialists as regards ontology engineering

- Action points:
  - lookup list of currently produced resources in projects
  - document on ‘ontology engineering for dummies’ (in coop with FAIRsFAIR)

WG: Interoperability in Practice

- Challenge: lack of best practice document to make the process of creating interoperable and FAIR data as simple as possible for the researcher

- Action points:
  - making a practical tool that researchers can use
  - providing an example of an end-to-end workflow for how researchers could make their data more interoperable (SoRa as reference use case)
  - executable data management plan, incl potential costs and benefits at each step in the process
GO INTER CONSORTIUM

- Peter Mutschke, Felix Bensmann, Benjamin Zapilko, GESIS – Leibniz Institute for the Social Sciences (GESIS), Cologne, Germany
- Abt Leit, Leibniz Information Centre for Economics (ZBW), Kiel, Germany
- Marc Ritterberg, Paul Lübbrecht, Daniel Schiffner, German Institute for International Educational Research (DIPF), Frankfurt, Germany
- Gotthard Meinel, Suji KumarSlider, Leibniz Institute of Ecological Urban and Regional Development (IOER), Dresden, Germany
- Michael Bonjak, Roland Ronthum, Tanja Burgard, Leibniz Institute for Psychology Information (ZPID), Trier, Germany
- York Sure-Vetter, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
- Andrea Scharnhorst, Kathleen Gregory, Gerard Coen, Ronald Siebes, Herbert van de Sompel, Data Archiving and Networked Services (DANS), Den Haag, The Netherlands
- Richard P. Smiraglia, Institute for Knowledge Organization and Structure, Lake Oswego, Oregon, USA
- Robert Perigi, Czech Technical University in Prague (CTU), Czech Republic
- Giancarlo Guizzardi, Free University of Bozen-Bolzano, Italy
- Tiago Prince Sales – University of Trento, Italy
- Michel Dumontier, Ricardo de Miranda Azevedo, Vincent Emenet, Chris Evelo, Maastricht University, The Netherlands
- Sören Auer, Markus Stokker, Oliver Koepper, TIB Hannover, Germany
- Oscar Corcho, María Poveda, Universidad Politécnica de Madrid, Spain
- Stefan Dietze, Heinrich-Hildebrandt-University Düsseldorf, GESIS – Leibniz Institute for the Social Sciences (GESIS), Germany
- Mathieu Daguin, Insight Centre for Data Analytics, National University of Ireland, Galway, Ireland
- John Domingue, Knowledge Media Institute, The Open University, Milton Keynes, UK
- Javad Chamouna, L3S Research Center, Leibniz University Hannover, Germany
- Paul Groth, University of Amsterdam, The Netherlands
- Julia Lane, New York University, USA
- Wouter Beck, Triply BV & guest researcher at VU University Amsterdam, The Netherlands
- Paolo Manghi, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo” – CNR, Pisa, Italy
- Tobias Kuhn, VU University Amsterdam, The Netherlands
- Peter McQuilton, Oxford, UK
- Caroline Goble, The University of Manchester, UK
- John Graybeal, Stanford University, USA
- Ruth Duerr, Ronin Institute for Independent Scholarship
- Milan Opletal, University of Maribor, Slovenia
- Barbara Magagna, Environment Agency Austria
Thank you!
MAIN TASKS

- Define execution plan & roadmap
- Review of existing technologies and standards
- Explore cross-domain use cases to better understand interoperability
- Provide mapping services that guide data providers in bringing (meta)data into common representation formats and in mapping their data to existing vocabularies
- Provide ontologies lookup services as gatekeepers across different standards and vocabularies
- Provide methods for qualified linking and annotating by broadly applying semantic technologies (ontology crosswalks, cross-ontology linking, use of foundational ontologies)
- Create semantically rich cross-domain research knowledge graphs supporting cross-community data search and analysis
- Apply existing or provide novel measures and gradational maturity models for assessing cross-domain interoperability (see fairmetrics.org and fairsfair.eu)
- Develop and evaluate reference implementations for real-world use cases
- Publish guidelines for implementing and assessing cross-domain interoperability
How to operationalize interoperability in a way that facilitates data integration across community boundaries (while at the same time capturing the meaning of data)?

- How can a semantic layer look like that tam the semantic disconnects?
  - Metadata templates (see CEDAR), Mapping services, Ontology services

- How can a data access layer look like that encapsulate the specific data organization level?
  - Digital Object Interface Protocol (RDA)

- How can linked data best be represented to finally integrate knowledge delivery?
  - Open Research Knowledge Graphs

- How to measure interoperability?
- How should future RDIs look like?