Why this Implementation Network (IN)?

Peter Mutschke (GESIS)

GO Inter Kick-Off Workshop, October 1, 2019, GESIS, Cologne
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

“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?

Sealing of Soils at Respondent’s Dwelling

Coordinates: 50.919721, 6.967379

http://www.sora-projekt.de/
### Key Challenges of Implementing Interoperability

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

[Manola/Mutschke/Scherp/Tochtermann/Wittenburg (2019): Implementing FAIR Data Infrastructures. Dagstuhl Perspectives Workshop 18472]
Further issues that have motivated the IN

- Lack of reference infrastructures
  - that fully implement the “I” of the FAIR principles
  - that may serve as a guideline for data provider to make their data FAIR

- Disparate communities:
  - open data world (W3C) vs research data world (RDA, DDI)
  - GO Inter as a contribution to close the gap

- Lack of measures to assess interoperability
  - new IN coming up: FAIRwizard/metrics

- Lack of cross-cutting INs:
  - Most INs domain-specific
  - cross-disciplinary perspective strongly required
Objectives of the IN

- 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 use 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
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
Guiding (Research) Questions

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?
Goal of the workshop

- Review of the Manifesto
- Clarify open issues
- Identify action points
- Identify WGs on action points
- Specify an agenda until end of 2020 (and beyond)
- Specify the mode of operation of the IN
- Identify next steps (until end of the year)
Thank you!