GO FAIR Implementation Network Manifesto: Cross-Domain Interoperability of Heterogeneous Research Data (Go Inter)

Introduction

[Short paragraph, context and justification]

A crucial obstacle to Open Science lies in the proliferation of domain-specific, disconnected "data silos". Despite the existence of widely accepted standards for data representation and linking (see the W3C[1]/RDA[2] standards), the data in these silos are often described using heterogeneous and often unstandardized metadata and vocabularies which cannot be easily linked with each other, making the interoperability, discovery and reuse of research data across community borders challenging tasks. A key challenge of a cross-disciplinary implementation of the FAIR principles lies in the complexities of **interoperability**, whose different layers, ranging from encoding up to structural and semantic specifications of data, are yet to be fully understood. In the broad debate about the FAIR principles, guidelines and recommendations are only just starting to discuss how to operationalize FAIR in research data infrastructures[3] or to address the FAIRness[4] of a given data set. As a consequence, efforts to implement the FAIR principles are not currently linked to the wider context of so-called reference models for research infrastructures[5]. Reference models that guide data providers in how best to represent their data in ways that capture the meaning of the data while ensuring interoperability without information loss are extremely rare. There is also a lack of understanding about how best to navigate between different levels of granularity or detail provided in domain specific data documentation schemes[6] and how to map between different Knowledge Organisation Systems. To address these challenges of interoperability, reference models need to be generic enough to be adaptable to different scientific domains. This is especially needed when it comes to linking data from different communities, one of the engines behind data-driven discoveries in interdisciplinary research.

---

[1] https://www.w3.org/, see also https://5stardata.info/en/

[2] https://rd-alliance.org/

[3] See, e.g., the PARTHENOS Guidelines to FAIRify data management and make data reusable (http://www.parthenos-project.eu/portal/policies_guidelines) and FAIRsFAIR (https://www.fairsfair.eu/)

[4] See as an example fairmetrics.org

[5] See, e.g., the technical reference model of CESSDA, as discussed recently here: https://euriseworkshop.sciencesconf.org/program

[6] See, for instance, Gueret at al. (2013): Genericity versus expressivity - an exercise in semantic interoperable research information systems for Web Science, preprint, https://arxiv.org/abs/1304.5743

The proposed network focuses on the "*I*" in FAIR, interoperability. While we recognize that data interoperability depends on both human, institutional and technical interoperability[7], our focus is on addressing the specific question of **semantic interoperability**. The proposed Implementation Network aims at applying, developing and evaluating methods, tools and guidelines for implementing and assessing the semantic interoperability of heterogeneous research (meta)data across discipline borders. This will be done in collaboration with similar initiatives such as the H2020 project FAIRsFAIR[8] and related RDA Interest Groups[9]. The work of the implementation network will take into account both domain-specific as well as common Web standards, and it will be research-driven from the beginning, i.e. it will be based on cross-community research use cases for which reference applications will be implemented.

Purpose of the Implementation Network

[A few sentences with clear contact and alignment with the FAIR principles]
The proposed implementation network aims at applying, developing and evaluating methods, tools and guidelines for implementing and assessing **semantic interoperability of heterogeneous research data across discipline borders**. By this cross-domain interoperability framework the implementation network intends to foster data sharing and discovery across traditional discipline boundaries.

Overarching Principle of Operation

"We commit to comply with the Rules of Engagement of GO FAIR Implementation Networks" *[required text]*.

Targeted Objectives for the Internet of FAIR Data and Services

1. To provide a cross-domain interoperability framework consisting of methods, tools and guidelines for implementing and assessing semantic interoperability of heterogeneous research data across discipline borders

[7] See Goldstein, E., Gasser, U., and Budish, B. (2018): Data Commons Version 1.0: A Framework to Build Toward AI for Good. Available at: https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-73414d7e72be
[8] https://www.fairsfair.eu/
[9] Such as the RDA Vocabulary and Semantic Service Interest Group and any related RDA groups on domain-specific semantics

2. To develop and evaluate reference implementations of interoperability for real-world cross-domain research uses case by broadly applying existing standards, vocabularies and semantics technologies
3. To engage with other GO FAIR implementation networks and related initiatives to disseminate and exchange best practice solutions for cross-domain interoperability

Primary Tasks

- Complete the execution plan & roadmap as part of the process becoming a GO FAIR Implementation Network (within in 6 months after approval)
- Review of existing technologies and standards as well as past and ongoing initiatives and projects which address the interoperability aspect of the FAIR principles[10]
- Explore cross-domain use cases to better understand differences of interoperability, in particular the different layers of interoperability
- Identify the main hurdles for concerted and coordinated actions to implement measures to enhance interoperability
- Derive requirements for cross-domain interoperability from real-world cross-domain research use cases
- Provide assistance services that guide data providers in bringing (meta)data into common representation formats and schemes (such as schema.org, DCAT[11]), in mapping their data to existing vocabularies and in making data available via standard protocols
- Provide an ontologies lookup service[12] that work as a gatekeeper across different standards and domains and overcomes incongruences between different vocabularies
- Provide models and methods for qualified linking and annotating cross-domain research data by broadly applying existing technologies from the Semantic Web community, such as ontology crosswalks (e.g. LOV[13]), smart ontology mapping, ontology alignment and existing Linked Data compliant semantic annotation services such as B2NOTE[14]
- Create semantically rich cross-domain research knowledge graphs that may better support cross-community data search and analysis
- Explore the use of foundational ontologies (namely Unified Foundational Ontology) for providing deep semantic meaning of data and thus improved means for interoperability.
- Apply existing or provide novel measures and gradational maturity models for assessing cross-domain interoperability (see fairmetrics.org and fairsfair.eu)

---

[10] Such as FAIRsFAIR (https://www.fairsfair.eu/), Data without Boundaries (http://www.dwbproject.org/), Open Research Knowledge Graph (https://zenodo.org/record/1157185), SoRa (https://www.gesis.org/en/research/external-funding-projects/overview-external-funding-projects/sora/)
[11] https://www.w3.org/TR/vocab-dcat/
[12] See, e.g., Goldfarb D. and Le Franc Y. (2017): Enhancing the Discoverability and Interoperability of Multi-disciplinary Semantic Repositories, http://ceur-ws.org/Vol-1933/paper-7.pdf
[13] https://lov.linkeddata.es/dataset/lov/
[14] https://b2note.eudat.eu

- Develop and evaluate reference implementations for real-world use cases that link data from different communities, by applying common Web (W3C) standards and technologies as well as solutions proposed by initiatives like RDA, such as the Digital Object Interface Protocol (DOIP)[15]
- Publish guidelines for implementing and assessing cross-domain interoperability
- Publish documents that summarise insights of the network, including requirements, specifications, descriptions of models, methods and implementations, evaluation results as well as gaps and weaknesses identified
- Share results with the GO-FAIR community and other related initiatives and networks through common workshops.

Use Cases

The concepts to be developed should be based on real-world cross-domain research use cases:
https://docs.google.com/document/d/1qwYj2SuptvpCq9bRsC2tL_O6r_BtjaMUR3q-vMBJieU/edit#

Membership list

We consider this Manifesto to be one way by which the undersigned stakeholders can **speak with one voice** on a number of critical issues that are of generic importance to the objectives of FAIR, and on which we feel we have reached consensus. *[required text]*

- Peter Mutschke, GESIS – Leibniz Institute for the Social Sciences (GESIS), Cologne, Germany (peter.mutschke@gesis.org), coordinator
- Atif Latif, Leibniz Information Centre for Economics (ZBW), Kiel, Germany (A.Latif@zbw.eu) [confirmed]
- Marc Rittberger, German Institute for International Educational Research (DIPF), Frankfurt, Germany (rittberger@dipf.de) [confirmed]
- Gotthard Meinel, Leibniz Institute of Ecological Urban and Regional Development (IOER), Dresden, Germany (g.meinel@ioer.de) [confirmed]
- Michael Bosnjak, Leibniz Institute for Psychology Information (ZPID), Trier, Germany (mb@leibniz-psychology.org) [confirmed]
- York Sure-Vetter, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany (york.sure-vetter@kit.edu) [confirmed]
- Andrea Scharnhorst, Kathleen Gregory, Data Archiving and Networked Services (DANS), Den Haag, The Netherlands (andrea.scharnhorst@dans.knaw.nl) [confirmed]

[15] P. Wittenburg, G. Strawn, B. Mons, L. Bonino, E. Schultes (2018): Digital Objects as Drivers towards Convergence in Data Infrastructures.
http://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11

- Richard P. Smiraglia, University of Wisconsin-Milwaukee, USA (richard.smiraglia@dans.knaw.nl) [confirmed]
- Robert Pergl, Czech Technical University in Prague (CTU), Czech Republic (perglr@fit.cvut.cz) [confirmed]
- Yann le Franc, e-Science Data Factory, Paris, France (ylefranc@esciencefactory.com) [confirmed]
- Giancarlo Guizzardi, Free University of Bozen-Bolzano, Italy, (Giancarlo.Guizzardi@unibz.it) [confirmed]
- Tiago Prince Sales - University of Trento, Italy (tgoprince@gmail.com) [confirmed]
- Michel Dumontier, Ricardo de Miranda Azevedo, Vincent Emonet, Maastricht University, The Netherlands (michel.dumontier@maastrichtuniversity.nl) [confirmed]
- Sören Auer, TIB Hannover, Germany (Auer@tib.eu) [confirmed]
- Oscar Corcho, Maria Poveda, Madrid University, Spain (ocorcho|mpoveda@fi.upm.es) [confirmed]
- Stefan Dietze, Heinrich-Heine-University Düsseldorf, Germany [confirmed]