

## **Results of the Kick-off Workshop of the GO FAIR Implementation Network GO INTER**

Posted on 4 November 2019

On October 1st and 2nd, 2019, the **kick-off meeting of the GO FAIR implementation network “Cross-Domain Interoperability of Heterogeneous Research Data (GO Inter)”** with 20 participants from 13 organizations took place at GESIS in Cologne. The network aims at bringing together experts from the Semantic Web community, data providers and use case partners from different domains in order to apply, develop and evaluate methods, tools and guidelines for implementing and assessing the (semantic) interoperability of heterogeneous research (meta)data across discipline borders. Thus, the network focuses on the “I” in FAIR, interoperability. This is seen as one of the main obstacles for the combination of research data from different scientific disciplines in the context of multidisciplinary research.

The participants were **welcomed by Peter Mutschke** (GESIS – Leibniz Institute for the Social Sciences) who founded and coordinates the IN. In his [introductory talk](#) he presented the motivation for and the objectives and tasks of the IN. As an example for the use case driven approach of the IN he pointed to the DFG project SoRa which aims at combining research data from the social sciences and the spatial sciences to enable multidisciplinary research on topics like environmental justice, which deals with the relationship between the different environmental burdens of different social groups, their perception and local environmental conditions.

Afterwards **Catharina Wasner (GO FAIR, Hamburg)** gave a [presentation of the GO FAIR initiative](#), followed by a dense program of inspiring **lectures** covering a broad range of relevant issues such as interoperability of ontologies, open research knowledge graphs, data sharing and searching:

### **Ignition Talks:**

- [FAIR Semantics: From common ontology design practices to a cross-disciplinary semantic ecosystem](#) (Yann Le Franc, e-Science Data Factory, France)
- [How did information flows change in the digital era?](#) (Sören Auer, TIB Hannover, Germany)
- [The Social-Spatial Research Data Infrastructure SoRa](#) (York Sure-Vetter, Karlsruhe Institute of Technology (KIT), Germany; Felix Bensmann, GESIS – Leibniz Institute for the Social Sciences, Germany; Sujit Sikder, Leibniz Institute of Ecological Urban and Regional Development (IOER), Germany)
- [Semantic Data Interoperability](#) (Luiz Bonino, GO FAIR, The Netherlands)
- [Ontology-Driven Conceptual Modelling in the Services of Interoperability](#) (Robert Pergl, Czech Technical University in Prague (CTU), Czech Republic)
- [TriplyDB: FAIR Data as the New Default](#) (Wouter Beek, Triply B.V. & VU University Amsterdam, The Netherlands)

### **Use Case Presentations:**

- [How to deal with evolving vocabulary when it comes to linking](#) (Gerard Coen / Andrea Scharnhorst / Ronald Siebes, Data Archiving and Networked Services (DANS), The Netherlands)

- [Shared Access for Items, Scripts and Streaming-Data](#) (Paul Libbrecht / Daniel Schiffner, German Institute for International Educational Research (DIPF), Germany)
- [Seeking data: A use \(user\) case](#) (Kathleen Gregory, Data Archiving and Networked Services (DANS), The Netherlands; Paul Groth, University of Amsterdam, The Netherlands)

#### Short Presentations:

- [I-ADOPT. An RDA WG](#) (Javad Chamanara, L3S Research Center, Leibniz University Hanover, Germany)
- [Ontology Engineering: Tools and Catalogues](#) (Oscar Corcho, Universidad Politécnica de Madrid, Spain)
- [Hints about Research Data Sharing](#) (Paul Libbrecht / Daniel Schiffner, German Institute for International Educational Research (DIPF), Germany)
- [Making Neural Networks FAIR](#) (York Sure-Vetter, Karlsruhe Institute of Technology (KIT), Germany)

Based on these talks and the overall discussion of the main challenges concerning cross-domain interoperability the participants split into **two working groups to discuss two major areas** identified as particularly relevant for the future discussion: One working group addressed the topic of **interoperability of ontologies and knowledge organization systems in data repositories**. Another group focused on the question of **how the process of creating interoperable research data can be made as simple as possible for the researcher**. The participants agreed to continue their work on the aforementioned issues in the form of permanent working groups. The initial results of the two working groups are described in the following. The IN envisions a document on best practices as regards interoperability.

#### Working Group: Interoperability in theory

##### Members:

- Oscar Corcho, UPM, Spain
- Paul Libbrecht, DIPF, Germany
- Luiz Bonino, GO FAIR office, The Netherlands
- Ronald Siebes, DANS, The Netherlands
- Robert Pergl, CTU, Czech Republic
- Andrea Scharnhorst, DANS, The Netherlands
- Javad Chamanara, Leibniz University Hannover, Germany
- Yann Le Franc, eScience Data Factory, France

##### Theoretical foundations of interoperability

The group started from the idea to do something around ontologies, Knowledge Organisation Systems (KOS) or, as an even broader umbrella term, “semantic artifacts”. We observe the existence of many repositories for vocabularies or KOS relevant for Linked (Open) Open data publishing, such as LOV, Bartoc (not only LD), DANS KOSo (classified KOS). Not all of them are guaranteed to be sustained. They all have been built for different reasons and hence come with different metadata to

describe a semantic artifact as a ‘system’ or entity. We also observed during the workshop that – due to the innovative pace in the field of semantic web technologies – even among the specialists, not all resources available to build, evaluate, and manage ontologies are known: neither the tools nor the practice around the tools. As a consequence, to create **a lookup list of currently produced resources** in projects is one action point of this group.

Ontologies come in all kinds of flavours – depending on the domains: on one end of the spectrum there exist attempts to create ‘foundational ontologies’ as general upper models to guide development of more specific ontologies – but these recent attempts are not that much informed by historic attempts to ‘classify the world’ (see e.g. Paul Otlet, Universal Decimal Classification). On the other end of the spectrum, the semantic web specialists engage with different ‘users’ or domain experts, and for that process, the non-specialists need to have some knowledge of the nitty-gritty ontology engineering to be able to really engage with the tech-savvies. This inspired the **idea for an ‘ontology engineering for dummies’** (informed by publications such as Kendall, E. F., & McGuinness, D. L. (2019). *Ontology engineering*. Morgan and Claypool Publishers).

#### **Actions:**

1. As currently the project FAIRsFAIR is about to produce a document on FAIRsemantics, we propose to contribute to this document and to organise the text around the above-mentioned lookup list of existing resources (together with the FAIRsemantics group). Deadline end of 2019 (led by Yann Le Franc).
2. Write an introduction for non-experts “Ontology engineering demystified” (led by DANS) 2020.

Both documents will inform the envisioned Best Practice document of the GO INTER IN.

### **Working Group: Interoperability in Practice**

#### **Members:**

- Wouter Beek, TriplyDB, The Netherlands
- Gerard Coen, DANS, The Netherlands
- Kathleen Gregory, DANS, The Netherlands
- Paul Groth, University of Amsterdam, The Netherlands
- Daniel Schiffner, DIPF, Germany
- Sujit Kumar Sikder, IOER, Germany
- York Sure-Vetter, KIT, Germany

#### **End-to-end interoperable (FAIR) data in practice**

The goal of this working group is to make the process of creating interoperable and FAIR data as simple as possible for the researcher. In addition to documenting best practices in interoperability, we aim to do this by:

1. making a practical tool that researchers can use,

2. providing an example of how this tool could be implemented to create interoperable and FAIR data, and
3. addressing the costs (i.e. required resources and challenges) and benefits (incentives) for both researchers and other decision-makers involved in making data more interoperable.

We envision **creating a package** (e.g. for R, Python), **that implements existing functions and makes use of existing ontologies, metadata schemas** (e.g. that developed by Datacite) **and infrastructures**. We also seek to build on existing packages which may have a similar end result of making data FAIR. We believe that it is important to build on tools that are already a part of researchers' workflows. In this way, we also contribute to making research performed in VREs such as Jupyter Notebooks FAIR.

Although the tool that we develop will have the end effect of "FAIRifying data" our focus is on increasing the **interoperability** of data specifically. We plan to do this by developing a series of functions which will harmonize data produced in VREs so that concepts and classes can be easily reconciled with external data sources.

For example, based on the Hunspell package functions `hunspell_check` & `hunspell_suggest` we envisage extending functionality of a 'check' function to also ensure terms are aligned with the appropriate ontology for the specified research domain and a variable is added which would facilitate linking with other datasets (using PIDs). A 'suggest' function could revise attribute labels to follow core metadata for observational data (see link to [RDA i-ADOPT WG](#)). There is also scope for automatically producing basic descriptive metadata (in the form of descriptive statistics) which could later be extracted or viewed as a "data summary" to reduce time spent by researchers when searching for datasets which meet their needs.

We plan on using the tool we develop to detail an example end-to-end workflow for how researchers could make their data more interoperable. We recognize that the process of making data FAIR in general and interoperable in particular is not a simple one. It involves many potential areas for "negotiation" and hurdles to overcome. In order to illustrate this, we plan on explicitly addressing potential costs and benefits at each step in the process.

In order to reach these aims, we have defined the following action items.

#### **Actions:**

1. Survey data management plans to identify the functionality a FAIR package should have ([Data Stewardship Wizard](#); [DMPonline](#) by the Digital Curation Centre (DCC));
2. Identify other packages, ontologies, and infrastructures that we can build on or incorporate;
3. Define an exemplary "FAIRifying" end-to-end process, e.g., an executable data management plan (SORA project), e.g.:
  - o data selection (DOI/metadata/Zenodo),
  - o data analysis (Jupyter),
  - o data publication (DOI/metadata/Zenodo);
4. Identify and estimate the cost of being FAIR for each step in the process;
5. Highlight the incentives (benefits) for the researcher (package user);
6. Collect feedback on the FAIRifying Package (time allowing).