Towards Transparent Presentation of FAIR-enabling Data Repository Functions & Characteristics

Robert Huber[1], Alejandra Gonzalez Beltran[2], Charlotte Neidiger[3], Robert Ulrich[3], and Hervé L'Hours[4]

[1]Universität Bremen, MARUM, PANGAEA, Bremen, Germany (rhuber@uni-bremen.de)
[2]Science and Technology Facilities Council, UK Research and Innovation
[3]Karlsruhe Institute of Technology
[4]UK Data Service, UK Data Archive, University of Essex

Identifying, finding and gaining a sufficient overview of the functions and characteristics of data repositories and their catalogues is essential for users of data repositories and catalogues in the environmental and geosciences, as well as in other domains. However, achieving this is not trivial within a reasonable amount of time and effort, especially for less experienced users. This lack of transparent, human- and machine-friendly exposure of essential data repository information impacts many possible stakeholders that need up to date and reliable information about data repositories to serve a broad range of users. These include, for example, search engines and registries such as GEOSS, re3data or FAIRsharing. Researchers need to be able to find FAIR enabling trustworthy repositories to deposit, curate and preserve their own digital objects, as well as to reliably find FAIR data already gathered by others in order to reuse it. Assessment bodies such as CoreTrustSeal need transparent access to data repositories' functions and characteristics in order to facilitate their certification process. An overview of the data and metadata standards, exchange services and interfaces offered by repositories is essential to data scientists in order to effectively integrate these into their workflows.

In this study we present how seemingly self-evident information about how the identity, purpose ('this is a data repository'), mandate and areas of responsibility of data repositories is exposed to humans and machines via websites and/or catalogues. Our findings are that such information is difficult to find and in many cases, machine-readable metadata is not clear, not relevant or missing altogether. We also show that despite all the efforts and successes in developing discipline specific standards over the last decades, these are insufficiently linked to from more domain agnostic standards. This absence of domain specific information in PID systems and search engines makes it to large extent invisible in the FAIR ecosystem. In particular, relevant metadata representations or links to discipline specific, standardised services, such as the Open Geospatial Consortium (OGC) suite of services, are rarely exposed.

In this paper, we seek to present the simple and effective methods being elaborated within the FAIR-IMPACT project to improve this situation by using existing and emerging methods and

standards. To this end, we will show effective ways that repositories can expose services information and standards via typed-link-based sign-posting as currently summarised in the FAIRiCAT approach. We will evaluate the options for implementation across domain-specific metadata as well as domain-independent formats such as DCAT or schema.org and show how they can be used in combination with FAIRiCAT in practice. We will also present methods for exposing the FAIR status of digital objects and the FAIR-enabling and trustworthiness status of data repositories to improve cooperation and information exchange between data repositories, registries, assessment providers and certification authorities.