A Simple Approach to Assessing the FAIRness of Data in Trusted Digital Repositories

Peter Doorn and Eleftheria Tsoupra

Data Archiving and Networked Services (DANS), Haag,
The Netherlands
`{peter.doorn, eleftheria.tsoupra}@dans.knaw.nl`

The FAIR principles (short for Findable, Accessible, Interoperable, Reusable) have in a short time become a household name in the research data world. In this paper, we describe a simple approach to evaluate datasets in Trusted Digital Repositories (TDR), using an assessment tool based on a brief questionnaire. It is argued that repositories complying with the Data Seal of Approval (or the new Core Trust Seal, the new common requirements of DSA and World Data System) already provide a basic level of FAIRness. In fact, the principles underlying the DSA, which were formulated in 2006, are quite similar to the FAIR principles formulated in 2014.

Essentially, the DSA gives quality criteria for digital repositories, whereas the FAIR principles target individual data sets (or data objects – this is an issue of granularity, but that is not fundamental for the discussion of this paper). Additionally, the FAIR Guiding Principles explicitly express it an aim that they are valid not only for humans, but also for machines, a demand never formulated explicitly by the DSA. In spite of these differences in orientation, there is a remarkable resemblance of the (four) FAIR Guiding Principles to the (five) principles underlying the DSA.

There is a growing demand for quality criteria for research datasets. In this paper we will argue that the DSA and FAIR principles get as close as possible to giving quality criteria for research data. They do not do this by trying to make value judgements about the content of data sets, but rather by qualifying the fitness for data reuse in an impartial and measurable way. By bringing the ideas of the DSA and FAIR together, we will be able to offer an operationalization that can be implemented in any Trustworthy Digital Repository.

In the simple DANS operationalization, it will be possible to score each data set upon ingest in a TDR, where a data archivist will be able to do the assessment, helped by pre- set scores on criteria that can be established automatically. After reuse, researchers will be asked to give user reviews based on the same criteria. In this way, every data set can receive a FAIR profile, of which we will provide some examples.

Work on creating FAIR metrics is ongoing in various context. This paper will compare our proposed approach with work going on elsewhere, especially by the FAIR metrics group of the GO FAIR initiative, see: https://www.dtls.nl/fair-data/fair-metrics-group/ in which we also participate. The ultimate aim is to arrive at a core set of FAIR metrics and methods that is applicable across disciplines and continents, although it may appear that different domains may require additional criteria and approaches.