FAIR Data in Trustworthy Data Repositories

Ingrid Dillo
Data Archiving and Networked
Services (DANS)
The Netherlands
ingrid.dillo@dans.knaw.nl

KEYWORDS

Certification, data sharing, FAIR, CoreTrustSeal, open data, repository, Data Seal of Approval, World Data System

1 OPEN DATA AND DATA SHARING

Research funding in recent years often comes with the condition to make some of the resulting data openly available. Most researchers appreciate the benefits of sharing research data, but on an individual basis they may be reluctant to share their own data. Why are some researchers hesitant to share? And what are the most important motivations of researchers who do share their data?[1]

2 THE CONCEPT OF TRUST AND REPOSITORY CERTIFICATION

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data. Open data and data sharing can only become a success if we put the concept of trust central stage. The certification of digital repositories is an important means to provide this trust to the different stakeholders involved.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that thedata they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

Within the tiered framework of certification standards that has developed over the last decade core level certification has been embraced by a large number of repositories around the globe. Within the framework of the Research Data Alliance the ICSU World Data System (WDS) and the Data Seal of Approval (DSA) have developed a unified catalogue of requirements.[2] The group built on inherent complementarity between the criteria previously established by the two organizations to harmonize unified and universal requirements reflecting the core characteristics of trustworthy data repositories. The first new CoreTrustSeals (CTS) have already been acquired.

3 FAIR DATA IN TRUSTWORTHY DIGITALREPOSITORIES

The condition to make data resulting from publicly funded research openly available has the effect that more and more data are rapidly becoming available. Therefore, there also is a growing demand for quality criteria for research datasets. The CTS requirements and the 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 datasets, but rather by qualifying the fitness for data reuse in an impartial and measurable way.

In 2014 the FAIR Guiding Principles (Findable, Accessible, Interoperable and Reusable) were formulated.[3] The well-chosen FAIR acronym is highly attractive. In a relatively short term, the FAIR data principles have been adopted by many stakeholder groups, including research funders.

The FAIR principles are remarkably similar to the underlying principles of DSA (2005): the data can be found on the Internet, are accessible (clear rights and licenses), in a usable format, reliable and are identified in a unique and persistent way so that they can be referred to. Essentially, CTS presents quality criteria for digital repositories, whereas FAIR targets individual datasets.

Bringing the ideas of the CTS and FAIR together offers an operationalization that can be implemented in any certified Trustworthy Digital Repository.

[1] Van den Eynden, V. and Bishop, L. (2014). Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report, available from knowledge-echange.info/Default.aspx?ID=733

[2] https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa–wdspartnership-wg.html https://www.datasealofapproval.org/en/news-andevents/news/2017/3/30/new-standards-and-certification-entity-horizon/
[3] https://www.force11.org/group/fairgroup/fairprinciples