DSA and FAIR: The best guarantee for the open and long term accessibility of research data.

Authors: Ingrid Dillo, Rob Hooft

Conference paper

download presentation (/media/presentation/initial_import/58.dilloi.datafitness.pptx)

Summary

Data fitness is multifaceted and covers various aspects related to a dataset such as its level of documentation, curation, peer review, and citability. Both the Data Seal of Approval as well as the FAIR data guiding principles are important means to establish trust with stakeholders (data depositors, data re-users, funders, etc.) in the fitness of data [1].

Data Seal of Approval

The Data Seal of Approval (DSA) is a community adopted and core level approach to certify repositories against a set of requirements for applying and verifying quality aspects concerning the creation, storage, use and reuse of digital data. The requirements are aimed at ensuring that the certified repository is recognized in the community as a trustworthy source of data. The objectives of the Data Seal of Approval are to safeguard data, ensure high quality and guide reliable management of data for the future without requiring implementation of new standards, regulations or heavy investments.

The idea to develop a core seal of approval for digital archives originated in the Netherlands ten years ago. From that moment the Data Seal of Approval organically grew, slowly but surely, from a national into an international certification standard. Nowadays certification standards are available at different levels, from a core level to extended and formal levels. Even at the core level, certification offers many benefits to a repository and its stakeholders.

Core certification involves a minimally intensive process whereby digital repositories supply evidence that they are sustainable and trustworthy. A repository first conducts an internal self-assessment, which is then reviewed by community peers. Such assessments help data communities—producers, repositories, and consumers—to improve the quality and transparency of their processes, and to increase awareness of and compliance with established standards. This community approach guarantees an inclusive atmosphere in which the candidate repository and the reviewers closely interact.

In addition to external benefits, such as building stakeholder confidence, enhancing the reputation of the repository, and demonstrating that the repository is following good practices, basic certification provides a number of internal benefits to a repository. Specifically, basic certification offers a benchmark for comparison and helps to determine the strengths and weaknesses of a repository.

Over the last two years the DSA community has been working with the ICSU/World Data System, also using a core certification procedure for its regular and network membership, within the framework of the Research Data Alliance [2]. This collaborative work has resulted in a catalogue of common requirements for core certification and a common procedure. Both organizations will start working with these common requirements this year.

FAIR Data

Over the last few years the acronym FAIR has entered the world of research data management. The FAIR movement summarizes the needs for data management and preservation into making data **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable both for humans and for computers. This requires machine readability of the metadata, a license and ideally also the data elements themselves. The FAIR data guiding principles [3] can help people to make decisions about the data management processes that need to be adopted during and after projects. FAIR data can come in a variety of formats, but the principles ensure that transforming data from one format to the other without complicated and error prone procedures is possible. Efforts are already underway to build a series of software products that facilitate making

QUICK LINKS
BACK TO SESSION
BACK TO PROGRAMME
BACK TO SESSION
Data fitness: What are the processes and components to assess and communicate re-usability of data? (/2016/sessions/53/)
BACK TO PROGRAMME
Back to Programme (/2016/programme/)

available as FAIR linked data already, and plans are under way between ELIXIR and NIH to do this for all reference data sets in Life Science research. This effort has the potential of significantly reducing the investment by people on what has been called *data munching*.

DSA and FAIR: complementarity and synergy

DSA and the FAIR data principles are fully aligned in the general aim to render data re-usable by more than those who originally generated them. A very important aspect of DSA that is only implicit in the “Accessibility” of the FAIR data principles is the check on long term sustainability of the data repository and the data in it. The explicit inclusion of machine readability is the main added value of the FAIR data principles, as well as the explicit focus on interoperability across datasets.

Another element in this complementary is the fact that the DSA is given at the level of the repository instance and the FAIR data principles are applicable at the level of individual data sets in the repository.

In practice this all means that data repositories in which only human readable data are stored and where these data are not (yet) interoperable for machines can perfectly qualify for the DSA. DSA, by virtue of the organization of the repositories, can also play a role in improving the Findability of the stored data. In turn, the Interoperability guaranteed by FAIR data will make it easier for repositories to clone each other’s data sets, thereby fulfilling one of the requirements for the sustainability guarantees in DSA.

Both FAIR and DSA have their own intrinsic value and should be further developed. Some data will never be FAIR (for instance images) but clearly, the metadata, image annotations and data set descriptors, wherever possible render the underlying data much better 're-usable in practice' when they are FAIR.

Thus, a very attractive 'endorsement' level for a repository (ranging from a single implementation of Dataverse or a light FAIRpoint to a fully fledged Digital Curation Centre) would be a fully FAIR DSA holder.

QUICK LINKS

BACK TO SESSION
BACK TO PROGRAMME

BACK TO SESSION

Data fitness: What are the processes and components to assess and communicate re-usability of data? (/2016/sessions/53/)

BACK TO PROGRAMME

Back to Programme (/2016/programme/)

Competing Interests

The authors declare that they have no competing interests.

Notes

1 http://datasealofapproval.org/en/ (http://datasealofapproval.org/en/), https://www.force11.org/group/fairgroup/fairprinciples (https://www.force11.org/group/fairgroup/fairprinciples)

2 https://rd-alliance.org/group/repository-audit-and-certification-dsa–wds-partnership-wg/outcomes/dsa-wds-partership (https://rd-alliance.org/group/repository-audit-and-certification-dsa–wds-partnership-wg/outcomes/dsa-wds-partership)

3 **Wilkinson, M.D., Dumontier, M. [...], Mons, B.** 2016, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, 160018 doi:10.1038/sdata.2016.18 http://www.nature.com/articles/sdata201618 (http://www.nature.com/articles/sdata201618)