

Helping the Consumers and Producers of Standards, Repositories and Policies to Enable FAIR Data Postprint

Authors: McQuilton, Peter, Batista, Dominique, Beyan, Oya, Granell, Ramon, Coles, Simon, Izzo, Massimiliano, Lister, Allyson L., Pergl, Robert, Rocca-Serra, Philippe, Schaap, Ben, Shanahan, Hugh, Thurston, Milo, Sansone, Susanna-Assunta, McQuilton, Peter

Date: 2022-11-16T00:00:00+00:00

Abstract

Thousands of community-developed (meta)data guidelines, models, ontologies, schemas and formats have been created and implemented by several thousand data repositories and knowledge-bases, across all disciplines. These resources are necessary to meet government, funder and publisher expectations of greater transparency and access to and preservation of data related to research publications. This obligates researchers to ensure their data is FAIR, share their data using the appropriate standards, store their data in sustainable and community-adopted repositories, and to conform to funder and publisher data policies. FAIR data sharing also plays a key role in enabling researchers to evaluate, re-analyse and reproduce each other's work. We can map the landscape of relationships between community-adopted standards and repositories, and the journal publisher and funder data policies that recommend their use. In this paper, we show how the work of the GO-FAIR FAIR Standards, Repositories and Policies (StRePo) Implementation Network serves as a central integration and cross-fertilisation point for the reuse of FAIR standards, repositories and data policies in general. Pivotal to this effort, the FAIRsharing, an endorsed flagship resource of the Research Data Alliance that maps the landscape of relationships between community-adopted standards and repositories, and the journal publisher and funder data policies that recommend their use. Lastly, we highlight a number of activities around FAIR tools, services and educational efforts to raise awareness and encourage participation.

Full Text

Preamble

Helping the Consumers and Producers of Standards, Repositories and Policies to Enable FAIR Data

Peter McQuilton¹⁺, Dominique Batista¹, Oya Beyan², Ramon Granell¹, Simon Coles³, Massimiliano Izzo¹, Allyson L. Lister¹, Robert Pergl⁴, Philippe Rocca-Serra¹, Ben Schaap⁵, Hugh Shanahan⁶, Milo Thurston¹ & Susanna-Assunta Sansone¹

¹Oxford e-Research Centre, Department of Engineering Sciences, University of Oxford, Oxford OX1 3PJ, UK

²Fraunhofer Institute for Applied Information Technology (FIT), 53754 Sankt Augustin, Germany

³School of Chemistry, Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ, UK

⁴Czech Technical University in Prague, Faculty of Information Technology (FIT CTU), 160 00 Prague 6, Czech Republic

⁵Wageningen University, Wageningen 6700 HB, The Netherlands

⁶Royal Holloway, University of London, London Surrey TW20 0EX, UK

Keywords: Convergence; Data repositories; Data policies; Data standards; FAIR data; FAIR enabling community standards

Citation: P. McQuilton, D. Batista, O. Beyan, R. Granell, S. Coles, M. Izzo, ...& S.-A. Sansone. Helping the consumers and producers of standards, repositories and policies to enable FAIR data. *Data Intelligence* 2(2020), 151-157. doi: 10.1162/dint_a_{00037}

Abstract

Thousands of community-developed (meta)data guidelines, models, ontologies, schemas and formats have been created and implemented by several thousand data repositories and knowledge-bases across all disciplines. These resources are necessary to meet government, funder and publisher expectations of greater transparency and access to and preservation of data related to research publications. This obligates researchers to ensure their data is FAIR, share their data using the appropriate standards, store their data in sustainable and community-adopted repositories, and conform to funder and publisher data policies. FAIR data sharing also plays a key role in enabling researchers to evaluate, re-analyse and reproduce each other's work.

We can map the landscape of relationships between community-adopted standards and repositories, and the journal publisher and funder data policies that recommend their use. In this paper, we show how the work of the GO-FAIR

FAIR Standards, Repositories and Policies (StRePo) Implementation Network serves as a central integration and cross-fertilisation point for the reuse of FAIR standards, repositories and data policies in general. Pivotal to this effort is FAIRsharing, an endorsed flagship resource of the Research Data Alliance that maps the landscape of relationships between community-adopted standards and repositories, and the journal publisher and funder data policies that recommend their use. Lastly, we highlight a number of activities around FAIR tools, services and educational efforts to raise awareness and encourage participation.

1. Making the Research Data Cycle FAIR

Science is constantly in flux, and (meta)data standards, databases and repositories are dynamic in nature, with a “life cycle” that encompasses creation, development, and maintenance. Once a standard is matured and databases and repositories that implement that standard are available, these resources need to be promoted to the relevant stakeholder community (such as database maintainers or funder data policy creators). The community, in turn, needs to recommend their implementation (e.g., in data policies of journals, publishers, funders and other organizations) or use (e.g., to define a data management plan), in order to facilitate a high-quality, FAIR research cycle.

To foster a culture such that the use of standards, databases and repositories for FAIRer data is pervasive, we need to reduce the mystique surrounding the FAIR principles [1] and provide practical guidelines for researchers, data policy makers, data managers and data stewards, and the developers/curators of these resources themselves.

A number of stakeholders—representing academia, industry, funding agencies, standards organizations, infrastructure providers and scholarly publishers—have come together as part of the FAIRsharing [2,3] community to help data consumers discover, select and use standards, repositories and policies with confidence, and producers to make their resource more discoverable, widely adopted and cited. FAIRsharing is now an endorsed flagship resource of the Research Data Alliance (RDA) [4], and works with the community via the joint Force11 and RDA FAIRsharing working group, as well as the GO-FAIR Standards, Repositories and Policies (StRePo) Implementation Network (IN), bringing together representatives of other GO-FAIR INs [5].

In this manuscript, as the representatives of the GO-FAIR StRePo IN, we present guidelines that highlight the role each stakeholder group should take to increase the visibility and adoption of standards, databases and repositories, and how FAIRsharing can enable this. We will also summarise other technical and educational activities, including FAIRsharing, that contribute to turning FAIR into reality.

2. Researchers and Data Stewards in Academia, Industry and Government

Researchers need to identify, use and ultimately cite the standards and databases (both knowledge-bases and repositories) that exist for their data and discipline and follow the recommendations and mandatory usage as stipulated in journal publisher and funder data policies. It is important therefore, both for the efficiency and subsequent quality of the data, that researchers keep this in mind as they create a Data Management Plan (DMP) for a grant proposal or funded project. This behaviour will enable researchers to provide and store publicly the relevant data and metadata associated with their research publications to help maximise the use and reuse of their data now and in the future. FAIRsharing can assist with identifying the most appropriate repositories, alongside the associated relevant standards they implement.

3. Developers and Curators of Standards, Databases and Repositories

Developers and curators of standards and databases need to increase the visibility and “discoverability” of their resource in order to increase not only the use but, equally as important, the number of citations they receive, which is still critical, unfortunately, for infrastructure funding. Developers and curators can use resources such as FAIRsharing to explore what resources exist in their area of interest (and if those resources can be used or extended for their work), as well as to enhance the discoverability and exposure of their own resource. The representatives of a database or repository are uniquely placed to describe their resource, and to declare the standards implemented by that resource. The same is true for standard developers. A metadata record for each resource can be created on FAIRsharing and then claimed by the developer of that resource. This allows the maintainer of the record to describe their resource via FAIRsharing metadata, which is marked up with schema.org [6] to ensure machine readability and discoverability via major search engines, and also receives a unique, persistent identifier (DOI) to allow the record to be cited.

In addition, the improved visibility—achieved through the linking of repositories and the standards that they implement—increases the likelihood that these resources are recommended by publisher and funder data policies. Apart from accessing FAIRsharing directly, third-party services may connect through its REST API interoperability to further make the records more findable and accessible to their users. Such an integration, for example, is currently being developed in the Data Stewardship Wizard [7].

4. Journal Publishers or Organizations with Data Policies

Data policy creators need to provide clear guidance to their communities on how they can select an appropriate repository for FAIR data and identify the relevant standards. FAIRsharing enables the maintenance of an interrelated list of citable standards and databases linked to the data policies that recommend them [8].

These recommendations can then be used by authors to refine and select the appropriate standards and repositories to use in their work while journals/publishers can revise their selections over time, enabling the recommendation of additional resources with more confidence. Funders and publishers that do not currently have such data statements are encouraged to develop them to ensure all data relating to an article or project are as FAIR as possible. Existing statements and recommendations from other journals and publishers can help to provide a valuable starting point for such a process. Funders and publishers should encourage authors to cite the standards, databases and repositories they use or develop via the “how to cite this record” statement and DOI, found on each FAIRsharing record [9].

5. Research Data Facilitators, Librarians and Trainers

Trainers, educators, librarians and the organizations and services they work with need to provide greater data management support for FAIR data. They can use FAIRsharing to provide a foundation on which to create or enrich educational lectures, training and teaching materials, and to inform data management planning tools. These stakeholders play a pivotal role in preparing the new generation of scientists and in delivering courses and tools that address the need to guide or empower researchers to organize data and make it FAIR.

6. Learned Societies, Unions and Associations

Learned societies, international scientific unions, associations, and alliances of these organisations should raise awareness around standards, databases, repositories and data policies, in particular on their availability, scope and value for FAIR and reproducible research. FAIRsharing is also working with global organizations like RDA, Force11, CODATA and GO-FAIR to mobilize their community members to take action, to promote the use and adoption of key resources, and to initiate new or participate in existing initiatives to define and implement policies and projects. Further, those data resources endorsed by their community can be promoted and described through the creation of a FAIRsharing collection for their domain.

7. Funders and Data Policy Makers

Funders can use FAIRsharing to help select the appropriate resources to recommend in their data policy and highlight those resources that awardees should consider when writing their data management plan. If we are to make FAIR data a reality, funders should recognize standards, databases and repositories as digital objects in their own right, which have their own associated research, development and educational activities. New funding frameworks need to be created to provide catalytic support for the technical and social activities around standards, both in specific domains and within and across disciplines. This support will enhance their implementation in databases and repositories and, ultimately, the interoperability and reusability of data.

8. Beyond the Registry: Tools, Services and Education

The FAIR principles are intended as a guide to enable data as well as other digital resources to become more Findable, Accessible, Interoperable and Reusable for humans and machines. Turning that guide into a workable ecosystem of services that allow the flow and exchange of FAIR data between FAIR resources across disciplines is the real challenge. Working with communities under GO FAIR, as the StrRePo IN, FAIRsharing has become a key element to build the matrix of resources that enable FAIRness [10], curating and linking metadata on each repository, knowledge-base, standard and data policy. This metadata can be used to assess the FAIRness of a resource and also to make distributed data analytics possible by improving the AI readiness of the data [11]. FAIRsharing, however, is not just a registry of (meta)data standards, repositories and policies. FAIRsharing is involved in a number of FAIR-enabling activities, all listed under its community page [12], exemplars of which are highlighted in the following sections.

9. Change: Policies and Incentives

Key to the penetration of FAIR data practices into the community is the work by journal and journal publisher data policy editors to surface the repositories and standards that conform to a high level of FAIRness. The collaboration between FAIRsharing, Datacite, and a number of journal and publisher representatives focuses on creating a shared list of core criteria for repository selection to help harmonize journals and publishers' data deposition guidelines for authors. The criteria are also informative for repository developers and maintainers, and useful as a reference for certification and other evaluation initiatives, because they highlight the features that journals and publishers believe are important for the identification and selection of appropriate data repositories. A collaboration between FAIRsharing and the Centre of Open Science (COS) is aimed at

standardising the classification of these journal/funder policies in the hope of encouraging more transparent research.

10. Build: Tools and Technologies

Good data management begins early in the research data lifecycle. Nowadays many funders mandate the creation of a Data Management Plan (DMP) to ensure that the FAIRness of data is not an afterthought. FAIRsharing collaborates with the Data Stewardship Wizard to drive the selection of domain-appropriate standards, repositories and data policies, part of a smart questionnaire that effortlessly guides users through the extensive requirements to be met to achieve good and FAIR data management.

How to measure the level of FAIRness, however, is still an open discussion. A number of efforts [13] are focused on developing maturity models as well as manual, semi-automated and automated evaluations; FAIRsharing provides content to many of these tools [14,15]. To measure the level of compliance of a data set (or other digital object) against the relevant metadata standards, FAIRsharing works to maximize the “computability” of these standards and their use in the evaluation process. This work also forms a contribution to the GO-FAIR “FAIR Funder pilot programme” [16], which recognizes the importance of enabling high quality metadata and careful planning for FAIR data stewardship.

Despite its success as a global brand, many elements that make FAIR actionable are still in their infancy. Building FAIR-related competencies and delivering training are one of these essential yet incomplete elements. In collaboration with other GO-FAIR INs, alongside CODATA, ELIXIR and FAIRsFAIR, FAIRsharing has founded the Terms4FAIRskills initiative [17] to create a formalised terminology that describes the competencies, skills and knowledge associated with making and keeping data FAIR. This work will assist with the creation and assessment of stewardship curricula, facilitating the annotation, discovery and evaluation of FAIR-enabling materials (e.g., training materials) and resources, and enable the formalisation of job descriptions and CVs with recognised, structured competencies.

11. Data Is Important: Help Us Make It FAIR

We are in the midst of a data explosion. More and more data are being generated, but are they available? Are they stored in an appropriate format? Have they been deposited in a sustainable and widely-known repository? Do they comply to the appropriate (meta)data guidelines, models, schema, formats and use unique and persistent identifiers? The challenge of this data-driven age, to stem data loss and to encourage good sustainable data, can only be met through the creation and use of FAIR-enabling standards, repositories and policies and

related advocacy and training. This will not only allow for greater accountability and data reuse and validation, but will also save a substantial amount of money and resources for organisations.

FAIRsharing contributes to the growing ecosystem of FAIR-enabling tools, services and educational frameworks. With the FAIR brand now established, there are a wealth of initiatives tackling the technological and social infrastructure that is turning FAIR into reality. This FAIR ecosystem, however, is still at an early stage. There will be new extinction and speciation events before a steady-state ecosystem develops. Even in this embryonic state, we can be proud of what we have achieved so far. Many communities are adopting FAIR principles and are creating FAIR data policies backed up by FAIR metadata. What matters now is widening the net, increasing participation in discussion. Join the ecosystem. Through working together and being an active participant we can build a FAIRer future.

Acknowledgements

Some of the discussion points in this article and the call for action were developed as part of the joint RDA and Force11 working group and the GO-FAIR StRePo IN. We therefore gratefully acknowledge the support provided by the RDA, Force11 and GO-FAIR communities and structures. FAIRsharing is funded by grants awarded to S.-A.S. that include elements of this work; specifically, grants from the UK BBSRC and Research Councils (BB/L024101/1, BB/L005069/1), European Union (H2020-EU.3.1, 634107, H2020-EU.1.4.1.3, 654241, H2020-EU.1.4.1.1, 676559), IMI (116060) and NIH (U54 AI117925, the new 1U24AI117966-01, 1OT3OD025459-01, 1OT3OD025467-01, 1OT3OD025462-01) and FAIRsharing award from the Wellcome Trust (212930/Z/18/Z), as well as a related award (208381/A/17/Z). S.-A.S. is also funded by the Oxford e-Research Centre, Department of Engineering Science of the University of Oxford.

References

- [1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, ...& B. Mons. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(2016), Article No. 160018. doi: 10.1038/sdata.2016.18.
- [2] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A.L. Lister & M. Thurston. FAIRsharing as a community approach to standards, repositories and policies. *Nature biotechnology* 37(358) (2019), 358–367. doi: 10.1038/s41587-019-0080-8.

- [3] FAIRsharing. Available at: <https://fairsharing.org/>.
- [4] The FAIRsharing registry and recommendations: interlinking standards, databases and data policies. doi: 10.15497/RDA00030.
- [5] Including the Metabolomics IN, Chemistry IN, Food-System IN, Personal Health Train IN, as well as the GO-Train pillar. Available at: <https://www.go-fair.org/implementation-networks/overview>.
- [6] SchemaORG. Available at: <https://www.schema.org>.
- [7] Data stewardship wizard. Available at: <https://ds-wizard.org>.
- [8] Examples of the resources recommended by the Wellcome Trust Open Research journal's data policy. Available at: <https://fairsharing.org/recommendation/WellcomeOpenResearch>.
- [9] An example (UniProKB), with both a DOI for the record and a publication citation for the resource. doi: 10.25504/FAIRsharing.s1ne3g.
- [10] H.P. Sustkova, K.M. Hettne, P. Wittenburg, A. Jacobsen, T. Kuhn, R. Pergl, ...& E. Schultes. FAIR convergence matrix: Optimizing the reuse of existing FAIR-related resources. *Data Intelligence* 2(2020), 158-170. doi: 10.1162/dint_a_{00038}.
- [11] O. Beyan, A. Choudhury, J. van Soest, O. Kohlbacher, L. Zimmermann, H. Stenzhorn, Md. R. Karim, M. Dumontier, S. Decker, L.O. Bonino da Silva Santos & A. Dekker. Distributed analytics on sensitive medical data: The Personal Health Train. *Data Intelligence* 2(2020), 96-107. doi: 10.1162/dint_a_{00032}.
- [12] FAIRsharing. Available at: <https://fairsharing.org/communities>.
- [13] FAIRassist, the nascent educational component of FAIRsharing. Available at: <https://fairassist.org>.
- [14] D.J.B. Clarke, L. Wang, A. Jones, M.L. Wojciechowicz, D. Torre, K.M. Jagodnik ...& A. Ma'ayan. FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources. 2019. bioRxiv 657676. doi: 10.1101/657676.
- [15] M.D. Wilkinson, M. Dumontier, S.-A. Sansone, L.O. B. da S. Santos, M. Prieto, D. Batista ...& E. Schultes. Evaluating FAIR maturity through a scalable, automated, community-governed framework. 2019. bioRxiv 649202. doi: 10.1101/649202.
- [16] P. Wittenburg, H.P. Sustkova, A. Montesanti, S.M. Bloemers, S.H. de Waard, M.A. Musen ...& E.A. Schultes. The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR data. 2019. Available at: <https://arxiv.org/abs/1902.11162>.
- [17] Terms for FAIR skills. Available at: <https://terms4fairskills.github.io>.

⁺ Corresponding Author: Peter McQuilton (E-mail: peter.mcquilton@oerc.ox.ac.uk, ORCID: 0000-0003-2687-1982).

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.