

FAIRsharing: standards, databases, repositories and policies

October 2018

Joint RDA-Force11 FAIRsharing Working Group

Co-chairs: Susanna-Assunta Sansone (0000-0001-5306-5690) Peter McQuilton (0000-0003-2687-1982) Simon Hodson (0000-0003-3179-7270) Rebecca Lawrence (0000-0003-4817-8206)	WG page: https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html WG email: rda-fairsharing-wg@rda-groups.org Adopters: https://fairsharing.org/communities
---	---

This document describes the outputs of a joint RDA-Force11 FAIRsharing WG¹.

1. The **FAIRsharing registry** of interlinked records on **standards** (for identifying, reporting, citing data and metadata), **databases** (repositories and knowledge-bases) and data **policies** (from journals, publishers, funders and other organizations), ranging from the generic and multi-disciplinary, to those from specific domains.
2. The **FAIRsharing recommendations**, to guide the users and producers of standards, databases and repositories on how to best select and describe these resources; and to guide funders and publishers on how to recommend them in data policies.

These outputs are tightly bound together as the registry enacts the recommendations, through the provision of well-described and interlinked records, which are curated and maintained with the support and input of the producers of these resources.

FAIRsharing is an informative and educational service that increases guidance to consumers of standards, databases, repositories, and data policies, accelerating the discovery, selection and use of these resources. It also aims to increase the satisfaction of resource producers through increased visibility, reuse, adoption and citation and their resource. Targeted to researchers and other stakeholders involved in producing, managing, serving, curating, preserving, publishing or regulating data, FAIRsharing supports and enables the implementation of the FAIR principles².

FAIRsharing has been built iteratively since its launch in 2011, and has matured under the guidance of the joint RDA-Force11 Working Group (WG), the FAIRsharing international Advisory Board and input from the community. FAIRsharing has become a sustainable service, part of an ecosystem embedded in international research infrastructure programmes, with an ever-expanding community of users, collaborators and adopters from various stakeholder communities.

¹ Previously known as BioSharing WG, the name change reflects the broader constituency and potential user base of its outputs, which go beyond the biosciences community. This change was approved by the RDA Council following interaction with RDA Technical Advisory Board and the WG Co-chairs on the 31st of July 2018.

² Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 15(3):160018 <https://doi.org/10.1038/sdata.2016.18> (2016).

Scope and Background

In this data-driven age, governments, funders and publishers expect greater transparency and reuse of research data, as well as greater access to and preservation of the data that supports research findings. This requires greater researcher responsibility for the produced data, which should result in greater confidence in, and the reuse of, existing data.

Community-driven standards, such as those for the identification, citation and reporting of data and metadata, underpin reproducible and reusable research, aid scholarly publishing, and drive both the discovery and evolution of scientific practice. The number of these standardization efforts, driven by large organizations or at the grass roots level, has been on the rise since the early 2000s. Thousands of community-developed standards are available (across all disciplines), many of which have been created and/or implemented by several thousand data repositories. Nevertheless, their uptake by the research community has been slow and uneven. This is mainly because investigators lack incentives to follow and adopt standards.

Problems with uptake of a standard by the community is exacerbated if the standard is not implemented by databases, repositories and other research tools, or endorsed by infrastructures. Furthermore, the fragmentation of community efforts results in the development of arbitrarily different, incompatible standards. In turn, this leads to standards becoming rapidly obsolete in fast-evolving research areas.

As with any other digital object, standards, databases and repositories are dynamic in nature, with a 'life cycle' that encompasses formulation, development and maintenance; their status in this cycle may vary depending on the level of activity of the developing group or community. When a standard is mature and standard-compliant databases and repositories become available these resources need to be promoted to the relevant stakeholder community, who in turn need to recommend their implementation (e.g., in data policies of journals, publishers, funders and other organizations) or use them (e.g. to define a data management plan) to facilitate a high-quality research cycle.

To foster a culture change into one where the use of standards, databases and repositories for FAIRer data is pervasive, we need to reduce the knowledge gap across the stakeholder around these resources, to better promote their existence, value, and use.

The FAIRsharing Registry and Related Services

Working with the community, the FAIRsharing team³ carefully curates information on standards, monitoring the evolution and implementation of standards in databases and repositories, and their recommendation by journal and funder data policies. Quantity is not the end goal here. Rather, the richness and accuracy of each record, their discoverability and interlinks are the priority. The content within FAIRsharing is licensed via the Creative Commons Attribution ShareAlike license 4.0 (CC BY-SA 4.0)⁴; the SA clause enhances the open heritage and aims to create a larger open commons, ensuring the users downstream share back.

³ Based at and (operated by) at the University of Oxford, UK.

⁴ FAIRsharing content license: <https://fairsharing.org/license>

Here are some of the key features and services that FAIRsharing offers.

Interlinked content. FAIRsharing covers the following type of standards. Minimum reporting guidelines (also known as guiding principles or checklists) outline the necessary and sufficient information vital for contextualizing and understanding a digital object. Terminology artifacts or semantics, ranging from dictionaries to ontologies, provide definitions and unambiguous identification for concepts and objects. Models and formats define the structure and relationship of information for a conceptual model or schema, and include transmission formats to facilitate the exchange of data between different systems. Identifier schema are formal systems for resources and other digital objects that allow their unique and unambiguous identification. These standards range from generic and multi-disciplinary, to standards that are tailored for specific disciplines. Databases in FAIRsharing encompass repositories and knowledge resources for datasets as well as other digital objects. Policies in FAIRsharing focus on formal guidelines by funders, journals and publishers, but are not limited to these. By interlinking these records FAIRsharing shows the number of databases in which a standard is implemented, the type and number of community standards a repository uses, and how many and which journal and funder data policies recommend the use of a repository or standard.

Growth and staggered rollout. As of 28th of September 2018, FAIRsharing has over 2497 records: 1252 standards, 1132 data repositories, 113 data policies (of which 80 are from journals and publishers and 23 from funders). Although FAIRsharing does not yet implement versioning, records are not deleted but their evolution is tracked and their status tagged. The life, agricultural, environmental, biomedical and health sciences has been the first focus areas, but the coverage has already started to expand to cover other disciplines working with and for new communities, following their request; new areas include engineering, natural, humanities and social sciences.

Up-to-date, credited description. Producers of standards, databases, and repositories and data policies are able to register and/or claim the record(s) for the resource(s) they maintain or have developed, ensuring that the description is accurate. Maintainers of each record can be linked with their Open Research and Contributor Identifier (ORCID)⁵, gaining personal recognition; a privacy policy describes how FAIRsharing collects and uses your personal data during your use of the site⁶. In communication with the maintainer(s), FAIRsharing assigns indicators to show the status in the resource's life cycle:

- 'Ready' for use;
- 'In Development';
- 'Uncertain' (when any attempt to reach out to the developing community has failed); and
- 'Deprecated' (when the community no longer mandates its use or has been superseded, together with an explanation where available).

A successful curation and communications process has already been in place for a couple of years, but it is iteratively refined as the content grows, and based on users feedback. Currently, each record is reviewed and curated by a FAIRsharing curator at least once a year, with priority to those frequently accessed, or that have more interlinks with others and with higher indicators of maturity. If a user updates a record, those updates will be assessed against our curation guidelines to ensure they are an accurate representation of the resource. Likewise, if a record is updated by a FAIRsharing curator, an email notification is sent to the record claimant, so providing a data cross-check to reduce the introduction of inaccuracies.

⁵ ORCID: <https://orcid.org>

⁶ Operated by the University of Oxford, FAIRsharing privacy policy is in accordance with the General Data Protection Regulation and associated data protection legislation: <https://fairsharing.org/privacy>

Citable, discoverable resources. After quality control checks by our curators, FAIRsharing mints digital object identifiers (DOIs) for each resource's metadata record⁷, which provides a persistent and unique identifier to enable the referencing of these resources. Updates to each record are captured in a publicly visible log, allowing all users to see the refinement of each record. This citation offers a unique, at-a-glance view of all descriptors and indicators associated with a resource, as well as any evidence of adoption or endorsement by a data policy or organization. Further machine readability of each record is ensured through markup annotation of the web content with schema.org⁸ vocabulary, which is employed by Google and other major search engines.

Resource finder, visualization tools. Records are manually categorized according to subject and domain by FAIRsharing curators, via two open application ontologies^{9,10}. This facilitates more accurate browsing, discovery and selection, via a number of search options (simple, advanced interface and a step-by-step wizard). A further feature which has proved popular with our adopters, is the ability to collate or group resources into Collections, where they are related to a particular group or initiative, or Recommendations, where the grouping is based on a data policy from a journal, society or funder. Both Recommendations and Collections are customizable with the adopter's name and logo, and maintained by a representative of their organization or project. Collections and Recommendations can be viewed on FAIRsharing as a table, grid or as an interactive network graph. In the graph, each node is clickable and provides some minimal information about the resource, along with a link to the main FAIRsharing record for that resource. The network graph can also be filtered by domain or species and can be expanded to include a further level of related databases and standards. The network graph, alongside new visually accessible statistics show the interlinking relationships between standards, databases, repositories and data policies (where present).

FAIRer resources for FAIRer data. FAIRsharing ensures these resources are Findable (e.g., by providing DOIs), Accessible (e.g., identifying their level of openness and licence type), encouraged to be Interoperable (e.g., highlighting which repositories implement the same standards to structure and exchange data), and Reusable (e.g., knowing the coverage of a standard and its level of endorsement by a number of repositories should encourage its use or extension in neighbouring domains, rather than reinvention). Collaborative work to improve FAIRsharing is happening on many fronts¹¹, e.g. to drive selection and decision-making via enrichment of indicators based on community-endorsed and discipline-specific criteria, such as FAIR metrics and FAIRness level.

⁷ FAIRsharing itself is a citable record: <https://doi.org/10.25504/FAIRsharing.2abjs5>

⁸ Schema.org: <https://schema.org>

⁹ The Domain Resource Application Ontology (DRAO), an application ontology describing cross-discipline research domains used within FAIRsharing records by curators and the user community
<https://github.com/FAIRsharing/domain-ontology>

¹⁰ The Subject Resource Application Ontology (SRAO), an application ontology describing subject areas / academic disciplines <https://github.com/FAIRsharing/subject-ontology>

¹¹ A 'live', updated list is maintained at <https://fairsharing.org/communities/activities>

The FAIRsharing Recommendations

Several stakeholders play catalytic roles to foster a culture change within the research community into one where the use of standards, databases and repositories for FAIRer data is pervasive; this following figure summarizes FAIRsharing guidance to each stakeholder group.

Researchers in academia, industry, government

Use FAIRsharing to identify and cite the standards, databases or repositories that exist for your discipline when creating a data management plan, releasing data or submitting a manuscript to a journal.



Developers and curators of resources

Make your standard, database or repository discoverable, by adding or claiming it in FAIRsharing; increase exposure and credit outside of your immediate community and promote adoption.

Journal publishers or organizations with data policy

Create in FAIRsharing your interrelated list of citable databases, repositories and relevant standards to recommend to your authors, users or their community; maintain and revise your recommendation over time.



Research data facilitators, librarians, trainers

Use FAIRsharing to provide a foundation on which to create or enrich educational lectures, training and teaching material; and to plug into data management planning tools and other FAIR-supporting resources.

Learned societies, unions and associations

Collaborate with FAIRsharing to raise awareness of standards, databases, repositories and data policies; mobilize your community to take action to promote registration, use and citation of key resources.



Funders and data policy makers

Recommend FAIRsharing to your awardees or community, to inform the development of their data management plan; and select the appropriate resources to recommend in your data policy.

Researchers in academia, industry and government

Researchers can use FAIRsharing as a lookup resource to identify, use and cite the standards, databases (both knowledge-bases and repositories) that exist for their data and discipline. This can be especially useful for example, when creating a data management plan for a grant proposal or funded project. Similarly, when submitting a manuscript to a journal, FAIRsharing can assist with identifying the most appropriate databases and repositories, alongside the associated relevant standards they implement. This will enable authors to provide the relevant data and metadata and to store it in the most appropriate resource. This will help maximise the use and reuse of their data.

Developers and curators of standards, databases and repositories

Standard developers and database curators can use FAIRsharing to explore what resources exist in their areas of interest (and if those resources can be used or extended), as well as to enhance the discoverability and exposure of their own resource. This resource might then receive credit outside of its immediate community and ultimately promote adoption¹². A representative of a community standardization initiative is best placed to describe the status of a standard(s) and to track its evolution. This can be done by creating an individual record¹³ or by grouping several records together in a collection¹⁴. To achieve FAIR data, linked data models need to be provided

¹² To learn how to add your resource to FAIRsharing, or to claim it: <https://fairsharing.org/new>

¹³ For example, the DDI standard for social, behavioral, economic, and health data: <https://doi.org/10.25504/FAIRsharing.1t5ws6>

¹⁴ For example, as done by the COMBINE, an initiative coordinating the development of the various community standards and formats for computational models: <https://fairsharing.org/collection/ComputationalModelingCOMBINE>.

that allow the publishing and connection of structured data on the web. Representatives of a database or repository are uniquely placed to describe their resource, and to declare the standards implemented¹⁵. The more adopted a resource is, the greater its visibility. For example, if your standard is implemented by a repository, these two records will be interlinked; thus, if someone is interested in that repository they will see that your standard is used by that resource. If your resource is recommended in a data policy from a journal, funder or other organization, it will be given a 'recommended' ribbon, which is present on the record itself and clearly visible when the resource appears in search results.

Journal publishers or organizations with data policies

For journal publishers or organizations with a data policy, FAIRsharing enables the maintenance of an interrelated list (Recommendation) of citable standards and databases, grouping those that the policy recommends to users or their community¹⁶. Journals/publishers can also revise their selections over time, enabling the recommendation of additional resources with more confidence. Journals 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. Finally, journals and publishers should also encourage authors to cite the standards, databases and repositories they use or develop via the 'how to cite this record' statement, found on each FAIRsharing record, which includes a DOI.

Research data facilitators, librarians and trainers

Trainers, educators as well as librarians and those organizations and services involved in supporting research data can use FAIRsharing to provide a foundation on which to create or enrich educational lectures, training and teaching material, and to plug it into data management planning tools. These stakeholders play a pivotal role to prepare the new generation of scientists and deliver courses and tools that address the need to guide or empower researchers to organize data and to make it FAIR.

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. They should also mobilize their community members to take action^{17,18,19}, 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; and create collection for their domain²⁰.

¹⁵ For example, the ICPSR archive of behavioral and social science research data that uses the DDI standard: <https://doi.org/10.25504/FAIRsharing.y0df7m>

¹⁶ Examples of Recommendations created by eight main publishers and journals are at: <https://fairsharing.org/recommendations>, including some generalist and many domain-specific databases and repositories.

¹⁷ CODATA, First ICSU-CODATA Workshop on Data Standards. Zenodo <http://doi.org/10.5281/zenodo.1193642> (2018).

¹⁸ JISC, Report on the Findable Accessible Interoperable and Reusable Data Principles. Zenodo <http://doi.org/10.5281/zenodo.1245568> (2018).

¹⁹ Science Europe, Presenting a Framework for Discipline-specific Research Data Management (D/2018/13.324/1). https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf (2018).

²⁰ For example, as done by the International Virtual Observatory Alliance, an organisation that debates and agrees the technical standards for astrophysics and astronomy: <https://fairsharing.org/collection/IVOA>

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 plan²¹. 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 activities²². 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, to enhance their implementation in databases and repositories, and ultimately the interoperability and reusability of data.

The FAIRsharing Community and Next Steps

Operating since 2011, and born from an early community-driven portal²³, FAIRsharing has become a sustainable service, hosted at the University of Oxford, and run with funds from a portfolio of infrastructure grants²⁴ that will ensure ongoing management and curation of this invaluable resource.

This is a major undertaking, but it is a journey we are not doing alone. An example is the recent article co-authored by 68 diverse stakeholders, representing academia, industry, funding agencies, standards organizations, infrastructure providers and scholarly publishers that have come together as a community, representing the core adopters, advisory board members, and/or key collaborators of the FAIRsharing resource²⁵. The group introduced the FAIRsharing mission and community network, along with a cohesive community approach to the growth in standards, databases, repositories and policies. This article also provides evidence that FAIRsharing is at the epicentre of international FAIR-enabling resources such as FAIRmetrics²⁶, and major community efforts including CODATA²⁷, ELIXIR²⁸, European Open Science Cloud (EOSC) Pilot²⁹, Global and Open FAIR (GO-FAIR)³⁰, International Organization for Standardization (ISO)³¹, and the US National Institutes of Health (NIH) Data Commons³².

²¹ European Research Council, Scientific Council: Open Research Data and Data Management Plans Information for ERC grantees. https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf (2018).

²² Sansone SA and Rocca-Serra P Review: Interoperability Standards. Wellcome Trust. Figshare. <https://doi.org/10.6084/m9.figshare.4055496.v1> (2016).

²³ Taylor CF, Field D, Sansone SA, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol.* 26(8):889-96 <https://doi.org/10.1038/nbt.1411> (2008).

²⁴ Grants are to its Principal Investigator, Professor Sansone SA, who is a permanent faculty member of the Engineering Science Department of the University of Oxford, as well as an Associate Director of the Department's Oxford e-Research Centre.

²⁵ Sansone SA, McQuilton P, Rocca-Serra P et al. FAIRsharing, a cohesive community approach to the growth in standards, repositories and policies. Pre-print bioRxiv 245183; doi: <https://doi.org/10.1101/245183> (2018).

²⁶ Wilkinson, Sansone, Schultes et al. A design framework and exemplar metrics for FAIRness. *Sci Data* 5:180118. <https://doi.org/10.1038/sdata.2018.118> (2018).

²⁷ CODATA: <http://www.codata.org>

²⁸ ELIXIR: <https://www.elixir-europe.org>

²⁹ EOSC Pilot: <https://eosc-pilot.eu>

³⁰ GO-FAIR: <http://go-fair.org>

³¹ ISO: <https://www.iso.org>

³² NIH Data Commons: <https://commonfund.nih.gov/commons>

Next steps include working on a broad communication plan to ensure the uptake of the recommendations - presented in this document - amongst each of the core stakeholder communities mentioned above.

How to join

Anyone can be a user of FAIRsharing. If you are a producer or maintainer of standards, databases and policies, you can add a new record or claim an existing record in FAIRsharing³³. Adopters, are representatives of institutions, libraries, journal publishers, infrastructure programmes, societies and other organizations or projects (that themselves serve and guide individual researchers or other stakeholders on research data management matters). Adopters use FAIRsharing specifically to do at least one of the following:

1. Educate their user community on the variety of existing standards, databases and policies and actively encourage them to submit/claim records, where relevant, and cite them;
2. Create Recommendations by registering their data policy, and then linking it to standards and/or databases recommended in the policy; and/or
3. Create a Collection by pulling together a list of standards and/or databases around a given domain of interest relevant to them; and lastly
4. Have a FAIRsharing logo on their websites with a link from their website to the registry homepage.

This document, as well as the content within FAIRsharing is licensed via the Creative Commons Attribution ShareAlike license 4.0 (CC BY-SA 4.0).

³³ To learn how to add your resource to FAIRsharing, or to claim it: <https://fairsharing.org/new>