



How FAIRsharing can help FAIRify your standards, databases and data policies

Peter McQuilton, PhD

ORCID: 0000-0003-2687-1982 | Twitter: @Drosophilic



Data Readiness Group

<https://datareadiness.eng.ox.ac.uk>



DEPARTMENT OF
ENGINEERING
SCIENCE



Outline

- Introduction to FAIR
- FAIRsharing – helping to FAIRify standards, databases and data policies
- Connections – building a FAIR ecosystem

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



What is it and what
can it do for you?

F
indable



A
ccessible



I
nteroperable



R
eusable



- **Findable**

- Discoverable on the web
- Uses globally unique, resolvable and persistent identifiers (e.g. DOI)

- **Accessible**

- Clearly defined access and security protocols (e.g. for sensitive data, like patient samples)

- **Interoperable**

- Machine-actionable
- Community-adopted standards (e.g. formats, guidelines)
- Linked with other resources, shares data

- **Reusable**

- Clear licensing, data provenance, uses community standards and stored appropriately

Lots of data aren't FAIR

Nature Genetics **41**, 149 - 155 (2009)
Published online: 28 January 2008 | doi:10.1038/ng.295



Repeatability of published microarray gene expression analyses

See associated Correspondence: [Baggerly, *Nature* 467, 401 \(September 2010\)](#)

John P A Ioannidis^{1,2,3}, David B Allison⁴, Catherine A Ball⁵, Issa Coulibaly⁴, Xiangqin Cui⁴, Aedín C Culhane^{6,7}, Mario Falchi^{8,9}, Cesare Furlanello¹⁰, Laurence Game¹¹, Giuseppe Jurman¹⁰, Jon Mangion¹¹, Tapan Mehta⁴, Michael Nitzberg⁵, Grier P Page^{4,12}, Enrico Petretto^{11,13} & Vera van Noort¹⁴

Given the complexity of microarray-based gene expression studies, guidelines encourage transparent design and public data availability. Several journals require public data deposition and several public databases exist. However, not all data are publicly available, and even when available, it is unknown whether the published results are reproducible by independent scientists. Here we evaluated the replication of data analyses in 18 articles on microarray-based gene expression profiling published in *Nature Genetics* in 2005–2006. One table or figure from each article was independently evaluated by two teams of analysts. We reproduced two analyses in principle and six partially or with some discrepancies; ten could not be reproduced. The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis. Repeatability of published microarray



Lots of data aren't FAIR

Nature Genetics **41**, 149 - 155 (2009)
Published online: 28 January 2008 | doi:10.1038/ng.295



Repeatability of published microarray gene expression analyses

See associated Correspondence: [Baggerly, *Nature* 467, 401 \(September 2010\)](#)

John P A Ioannidis^{1,2,3}, David B Allison⁴, Catherine A Ball⁵, Issa Coulibaly⁴, Xiangqin Cui⁴, Aedín C Culhane^{6,7}, Mario Falchi^{8,9}, Cesare Furlanello¹⁰, Laurence Game¹¹, Giuseppe Jurman¹⁰, Jon Mangion¹¹, Tapan Mehta⁴, Michael Nitzberg⁵, Grier P Page^{4,12}, Enrico Petretto^{11,13} & Vera van Noort¹⁴

Given the complexity of microarray-based gene expression studies, guidelines encourage transparent design and public data availability. Several journals require public data deposition and

The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis. Repeatability of published microarray studies is apparently limited. More strict publication rules enforcing public data availability and explicit description of data processing and analysis should be considered.

Not FAIR: Low 'findability' and interoperability



- Not always well cited, stored
 - Software, codes, workflows are hard(er) to get hold of
- Poorly described for third party reuse
 - Different level of detail and annotation
- Curation, reporting and annotation activities are perceived as time consuming
 - Sometimes rushed and minimally done if professional curation is not available

FAIR principles – built on metadata

Principles put emphasis on enhancing the ability of ***machines*** to automatically find and use the data, in addition to supporting its reuse by ***individuals***

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Licensing

- License your digital object
- Without a license, an object cannot be reused
- Which license?
 - Most permissive you can
 - License doesn't mean open, just provides a framework for use
- Publish your data – retain ownership but allow others to reuse, with attribution and credit

Put your data somewhere FAIR

- A community repository
- Trusted and vetted by the community
 - Funded/sustainable
 - Has a clear data management and sustainability plan
 - Uses community-adopted standards
- With the appropriate license
- Use standards and repositories that have been endorsed by funders, journal publishers, other organisations (e.g. ELIXIR)



ELIXIR position paper on FAIR Data Management in the life sciences

Findable, Accessible, and Interoperable life-science data are reused

Biological sciences have a long tradition of open research data where vast data sets are made available for community reuse -- sometimes even prior to publication. ELIXIR, the European research infrastructure for life-science data, is committed to coordinating, integrating and sustaining deposition databases and supports European life scientists in making their data Findable, Accessible, Interoperable and Reusable (FAIR).¹ The coordinated action of national Nodes ensures harmonised data handling and management and provides the mechanism for FAIR data in collaborative European life-science projects.

Blomberg N and ELIXIR Consortium.

ELIXIR position paper on FAIR data management in the life sciences

[version 1; not peer reviewed].

F1000Research 2017, **6**(ELIXIR):1857 (document)

 [10.7490/f1000research.1114985.1](https://doi.org/10.7490/f1000research.1114985.1)

Ways to help make your repository, standard or data policy FAIR

- **Findable** - use PID schemas, use schema.org mark-up, add metadata to FAIRsharing
- **Accessible** - Define level of openness – access protocol and license clearly in a policy findable from the homepage
- **Interoperable** – Use community standards for reporting, models, formats and terminologies
- **Reusable** - Licensing, provenance of data, follow reporting standards – clear policy linked from homepage



FAIRify your standards,
databases and data policies

Since 2011

FAIRsharing.org

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

Find

Recommendations

Standards and/or databases recommended by journal or funder data policies.

Discover

Collections

Standards and/or databases grouped by domain, species or organization.

Learn

Educational

About standards, their use in databases and policies, and how we can help you.

 Search FAIRsharing

Search

Standards Databases Policies Collections/Recommendations

Advanced Search



Fine grained control over your search.

Search Wizard



Let us guide you to your results.



1352 Standards

Terminology Artifact	752
Model/Format	394
Reporting Guideline	163
Identifier Schema	14
FAIR metrics	29



1325 Databases

Natural Sciences	1222
Engineering Science	221
Humanities	45
Social Sciences	49

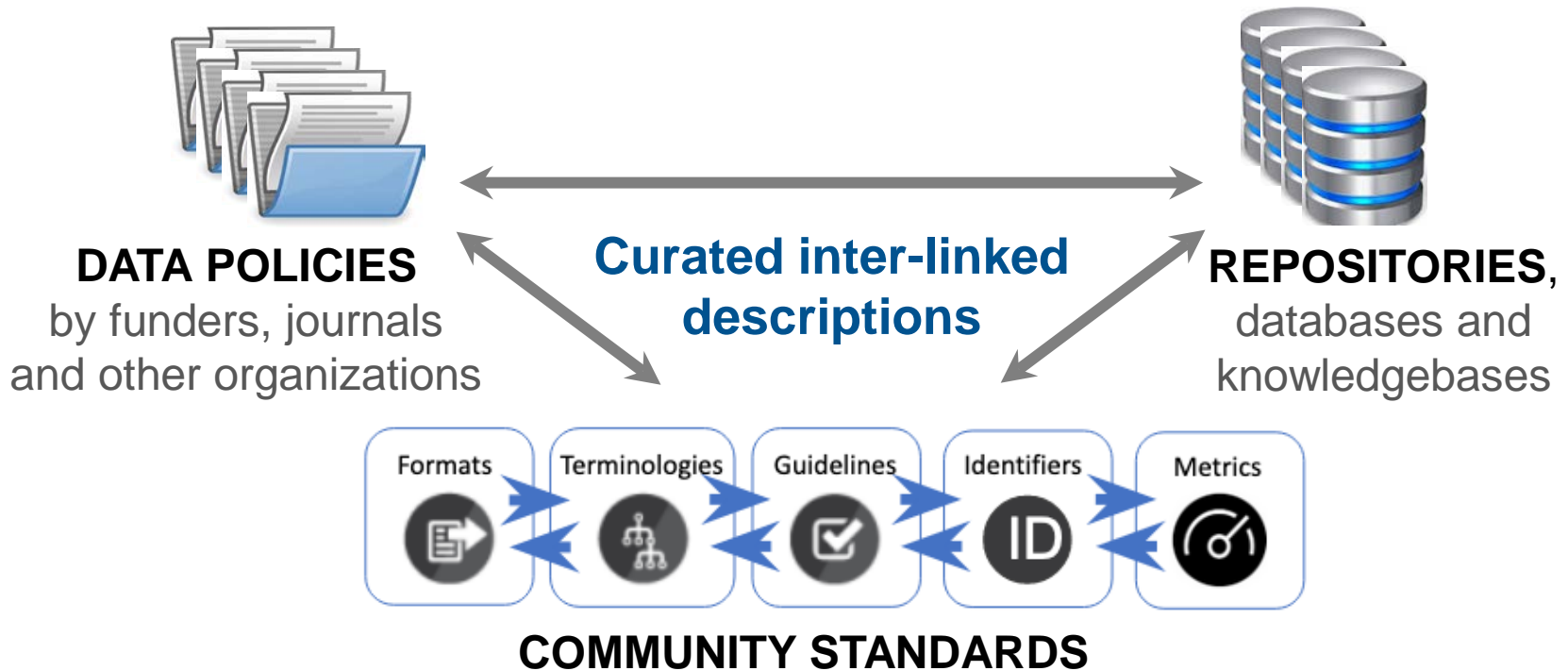


130 Policies

Funder	23
Journal	85
Society	9

FAIRsharing.org

informative and educational resource



We guide *consumers* to discover, select and use these resources with confidence

We help *producers* to make their resources more visible, more widely adopted and cited

Providing rich descriptive metadata for resources

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards Databases Policies Collections Add/Claim Content Stats PeteMcQ

Open Science Framework

Abbreviation: OSF

General Information

The OSF supports the entire research lifecycle: planning, execution, reporting, archiving, and discovery. Features include: automated versioning, logging of all actions, collaboration support, free and unlimited file storage, registrations, and connections to other tools/services (ie. Dropbox, figshare, Amazon S3, Dataverse, GitHub). It is 100% free, open source, and intended for use in all domain areas.

Homepage <http://osf.io>

Countries that developed this resource [Worldwide](#)

Created in 2011

Taxonomic range

All

Knowledge Domains

Cancer Data Storage Study Design

Subjects

Life Science Psychology

User-defined Tags

General Purpose

RECOMMENDED

Providing rich descriptive metadata for resources

FAIRsharing.org standards, databases, policies

Search all of FAIRsharing

Standards Databases Policies Collections Add/Claim Content Stats PeteMcQ

R Open Science
Abbreviation: OSF

General Information

The OSF supports the en logging of all actions, coll figshare, Amazon S3, Dat

Homepage <http://osf.io>

Countries that develop

Created in 2011

Taxonomic range

All

Knowledge Domains

Cancer Data Storage

Subjects

Life Science Psychology

User-defined Tags

General Purpose

Support

General

support email	Technical support
support email	General enquiries
forum	https://groups.google.com/forum/?hl=en#! ...
online document ation	https://osf.io/getting-started/
twitter	@OSFramework

Additional Information

Contact	Andrew Sallans	ORCID
---------	--------------------------------	-----------------------

RECOMMENDED

Providing rich descriptive metadata for resources

FAIRsharing.org
standards, databases, policies

Search all of FAIRsh...

PeteMcQ

Open Science

Abbreviation: OSF

General Information

The OSF supports the en...
logging of all actions, coll...
figshare, Amazon S3, Dat...

Homepage <http://osf.io>

Countries that developed

Created in 2011

Taxonomic range

All

Knowledge Domains

Cancer Data Storage

Subjects

Life Science Psychology

User-defined Tags

General Purpose

Support

General

- support email
- support email
- forum
- online document
- ation
- twitter

Additional Information

Contact

Andrew Sarans

Tools

Code Repository	https://github.com/centerforopen-science
-----------------	---

Access / Retrieve Data

Conditions of Use

Data Versioning

Complete Versioning Of All Files Stored In System, Available Permanently	No link recorded
--	------------------

Data Access

Search	https://osf.io/search/
--------	---

Data Release

Continuous Release	No link recorded
--------------------	------------------

RECOMMENDED

Highlighting relationships with standards, databases and data policies

The screenshot displays the FAIRsharing.org website interface. The top left features the FAIRsharing.org logo with the tagline 'standards, databases, policies'. The top right shows a user profile for 'PeteMcQ'. The main content area is divided into two sections: 'Related Databases' and 'Implementing Policies'. The 'Related Databases' section lists 'figshare', 'Mendeley Data', and 'DataverseNL'. The 'Implementing Policies' section lists various guidelines and standards, including 'African Academy of Sciences Open Research Data Guidelines for Authors', 'MNI Open Research', 'Wellcome Open Research Data Guidelines', 'Taylor and Francis Data Policy', 'F1000Research Data Policy', 'GigaScience - Minimum Standards of Reporting Checklist', 'HRB Open Research', 'eLife Recommended Repositories and Standards', 'Scientific Data Data Policy', 'Public Library of Science Recommended Data Repositories', 'UCL Child Open Health Research', 'Open Research Data and Data Management Plans - Information for ERC grantees by the ERC Scientific Council', and 'Gates Open Research Data Guidelines'. On the left side, there is a sidebar with 'Open Science' (Abbreviation: OSF) and 'General Information' (The OSF supports the logging of all actions, coll figshare, Amazon S3, Dat). Below this, there are sections for 'Homepage http://osf.io', 'Countries that developed', 'Created in 2011', 'Taxonomic range' (All), 'Knowledge Domains' (Cancer, Data Storage), 'Subjects' (Life Science, Psychology), and 'User-defined Tags' (General Purpose). A red 'RECOMMENDED' banner is visible on the right side of the page.

FAIRsharing.org
standards, databases, policies

PeteMcQ

Related Databases

- figshare
- Mendeley Data
- DataverseNL

Implementing Policies

- African Academy of Sciences Open Research Data Guidelines for Authors
- MNI Open Research
- Wellcome Open Research Data Guidelines
- Taylor and Francis Data Policy
- F1000Research Data Policy
- GigaScience - Minimum Standards of Reporting Checklist
- HRB Open Research
- eLife Recommended Repositories and Standards
- Scientific Data Data Policy
- Public Library of Science Recommended Data Repositories
- UCL Child Open Health Research
- Open Research Data and Data Management Plans - Information for ERC grantees by the ERC Scientific Council
- Gates Open Research Data Guidelines

Open Science
Abbreviation: OSF

General Information
The OSF supports the logging of all actions, coll figshare, Amazon S3, Dat

Homepage <http://osf.io>

Countries that developed

Created in 2011

Taxonomic range

All

Knowledge Domains

Cancer Data Storage

Subjects

Life Science Psychology

User-defined Tags

General Purpose

RECOMMENDED

Highlighting relationships with standards, databases and data policies

FAIRsharing.org standards, databases, policies

PeteMcQ

Related Databases

In the following recommendations:

HRB Health Research Board

UCL Child Health Open Research

neuro

PLOS

AAAS

Genetics & Genomics Next

Taylor and Francis

Gates Open Research

eLife Recommended Repositories and Standards

W

SCIENTIFIC DATA

BioMed Central The Open Access Publisher

F1000 Research

GIGA SCIENCE

How to cite this record FAIRsharing.org: OSF; Open Science Framework DOI: <https://doi.org/10.25504/FAIRsharing.g4z879>; Last edited: Jan. 8, 2019, 1:21 p.m.; Last accessed: Oct 21 2019 8:26 p.m.

This record is maintained by [asallans](#) ORCID and [sarabowman](#) ORCID

Record updated: July 27, 2018, 8:31 a.m. by [The FAIRsharing Team](#).

Life Science Psychology

User-defined Tags

General Purpose

UCL Child Open Health Research
Open Research Data and Data Management Plans - Information for ERC
grantees by the ERC Scientific Council
Gates Open Research Data Guidelines

SPRINGER NATURE

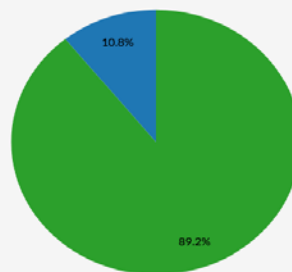
SCIENTIFIC DATA

This record is maintained by: [ScientificData](#)

This recommendation is taken from the following policy:

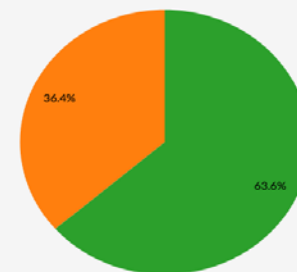
[Scientific Data Data Policy.](#)

General Statistics



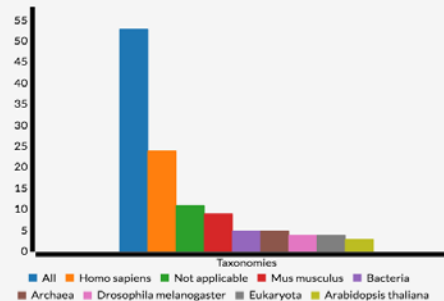
Standards Policies Databases

Standard Types



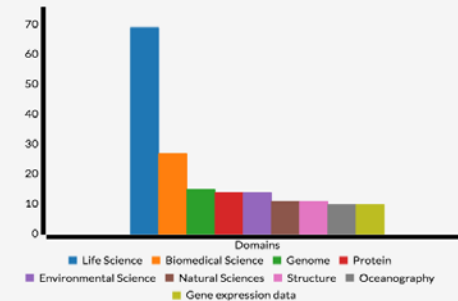
Model/Format Reporting Guideline Terminology Artifact Other

Taxonomies (top ten)



All Homo sapiens Not applicable Mus musculus Bacteria Archaea Drosophila melanogaster Eukaryota Arabidopsis thaliana

Domains (top ten)



Life Science Biomedical Science Genome Protein Environmental Science Natural Sciences Structure Oceanography Gene expression data

Showing records 1 - 50 of 102.

« 1 2 3 »

R
RECOMMENDED

Ag Data Commons

DATABASE

- Standards 7
- Publications 0
- In Collections 1
- Recommended 1

9 Taxa types, including:

All
 Animalia
 Archaea
 Bacteria

Chromista
 Fungi
 Plantae
 Protozoa

43 Data types, including:

Agricultural Products
 Agriculture

Agroecology
 Agroecosystem
 Agronomy

R
RECOMMENDED

ADS

Archaeology Data Service

DATABASE

- Standards 0
- Publications 0
- In Collections 0
- Recommended 1

No taxa defined.

4 Data types, including:

Classical Archeology
 History
 Humanities

Natural Sciences

R
RECOMMENDED

ArrayExpress

DATABASE

- Standards 7
- Publications 2
- In Collections 3
- Recommended 13

1 Taxa types, including:

All

9 Data types, including:

DNA Microarray
 Gene Expression Data

Genotyping
 Life Science
 Methylation

View as Table
View as Grid

Sort by

Best Match

Recommended Records

Recommended

Associated Publication?

No Publication

Has Publication

Claimed?

No Maintainer

Has Maintainer

Record Status

Uncertain

Deprecated

In development

Ready

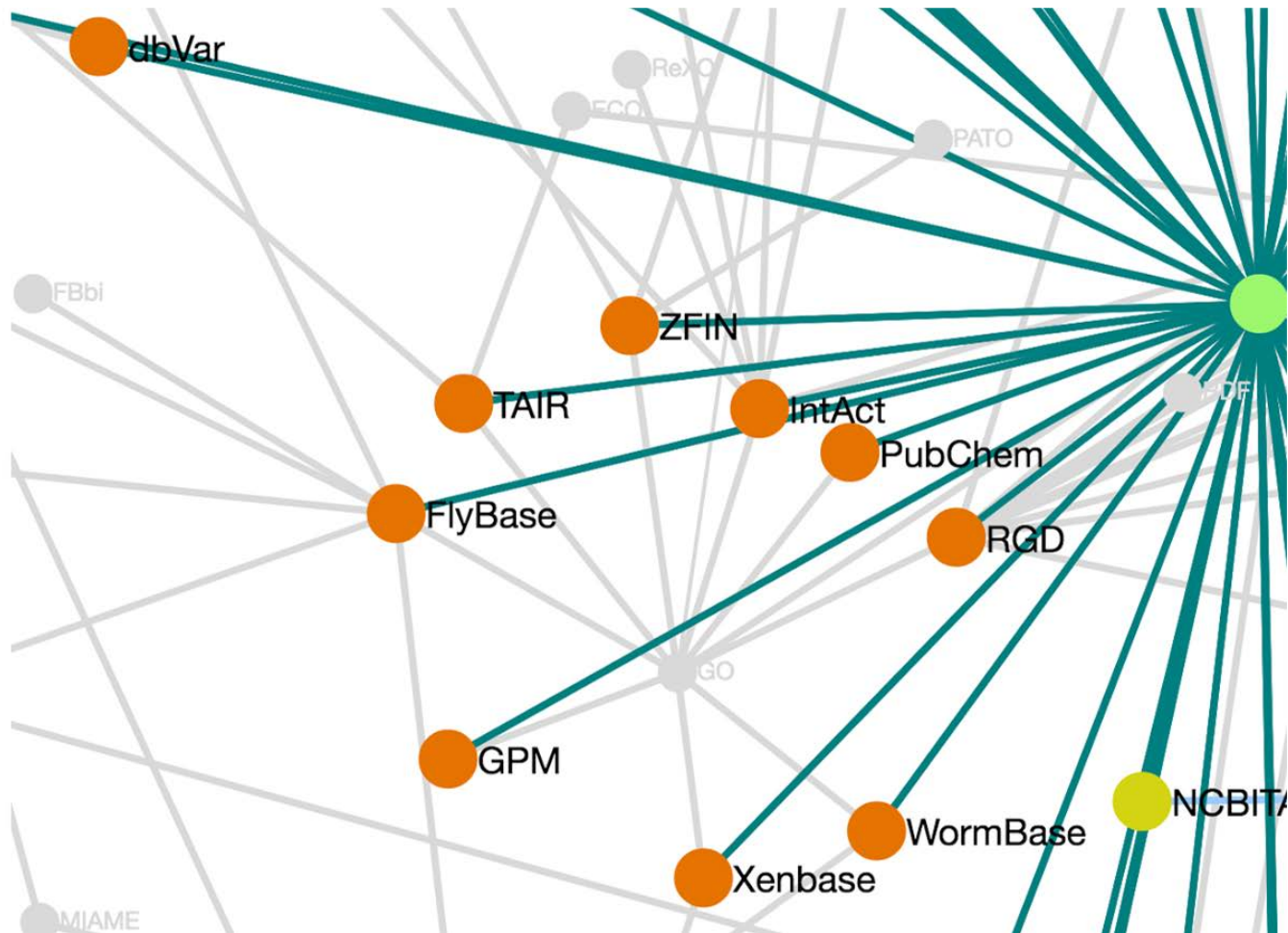
Standard Type

Terminology Artifact
7

Reporting Guideline
4

Registry

Database
91



*“The interactive browser will **allow us to discover** which **databases and standards** are not currently included in our author guidelines, **enabling us to regularly monitor and refine our policies** as appropriate, in support of our mission to help our authors enhance the reproducibility of their work.”*

H. Murray. Publishing Editor, **F1000Research**

Mapping the landscape of badges and certification

ELIXIR Core Data Resources and Deposition Databases

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research.

This record is maintained by: [PeteMcQ](#) [ORCID](#), [RachelDrysdale](#)

Record added: March 9, 2017, 1:28 p.m..

Record updated: Feb. 20, 2019, 7:55 p.m. by [The FAIRsharing Team](#).

[Homepage](#)

[Reference](#)

Taxonomic range

[All](#)

Knowledge Domains BETA

[Biocuration](#)

[Literature Curation](#)

Subjects BETA

[Bioinformatics](#)

[Biomedical Science](#)

[Data Integration And Warehousing](#)

[Data Subm](#)

User-defined Tags

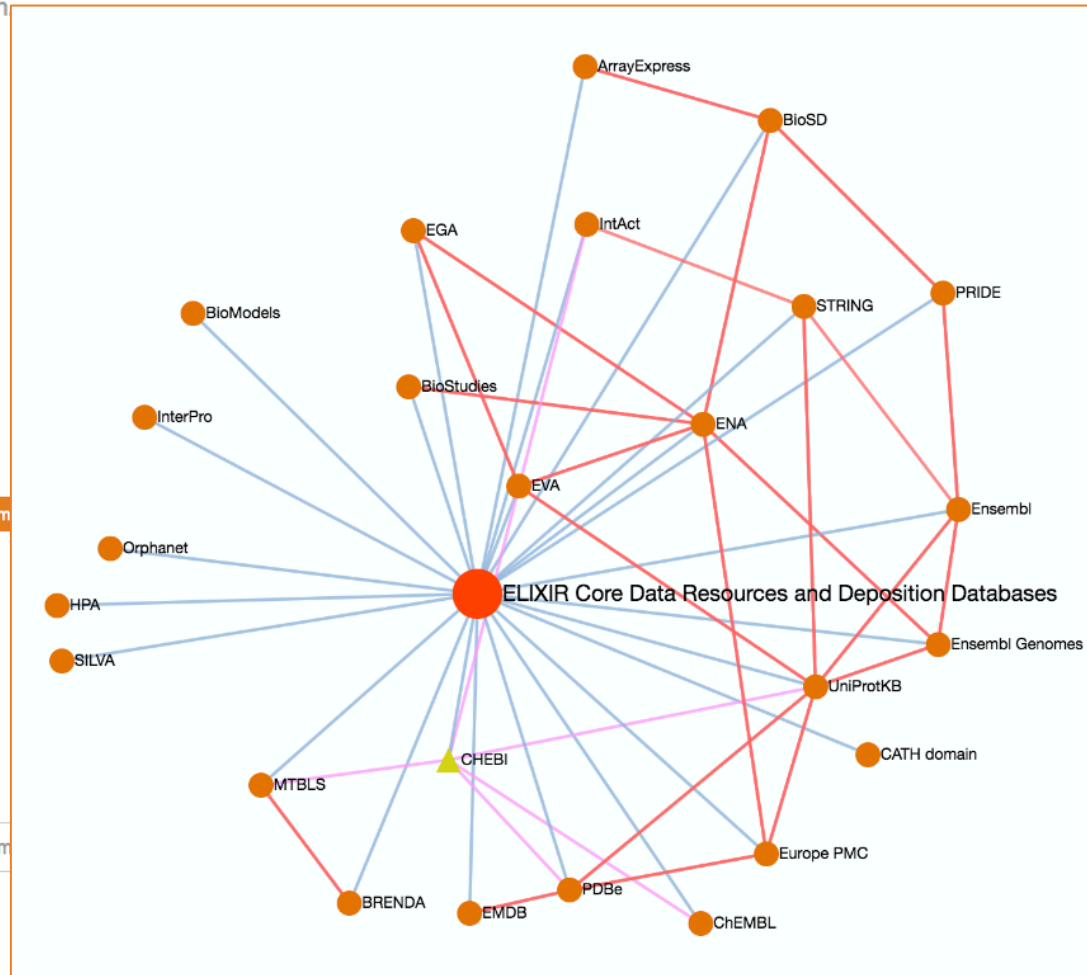
[Interoperability](#)

[View as Graph](#)

[Show edit history](#)

[Compare with collection/recommendation \(Beta\)](#)

Please select a Collection or Recomm



Mapping the landscape – collections of resources



International Virtual Observatory Alliance (IVOA)

The Virtual Observatory (VO) is the vision that astronomical datasets and other resources should work as a seamless whole. Many projects and data centres worldwide are working towards this goal. The International Virtual Observatory Alliance (IVOA) is an organisation that debates and agrees the technical standards that are needed to make the VO possible. It also acts as a focus for VO aspirations, a framework for discussing and sharing VO ideas and technology, and body for promoting and publicising the VO. This collection lists the documents & standards agreed by IVOA.

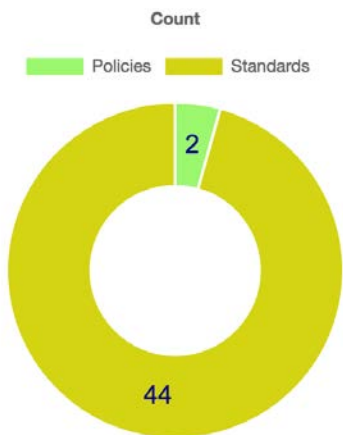
This record is maintained by: [carviset](#)

Record added: March 15, 2018, 12:56 p.m..

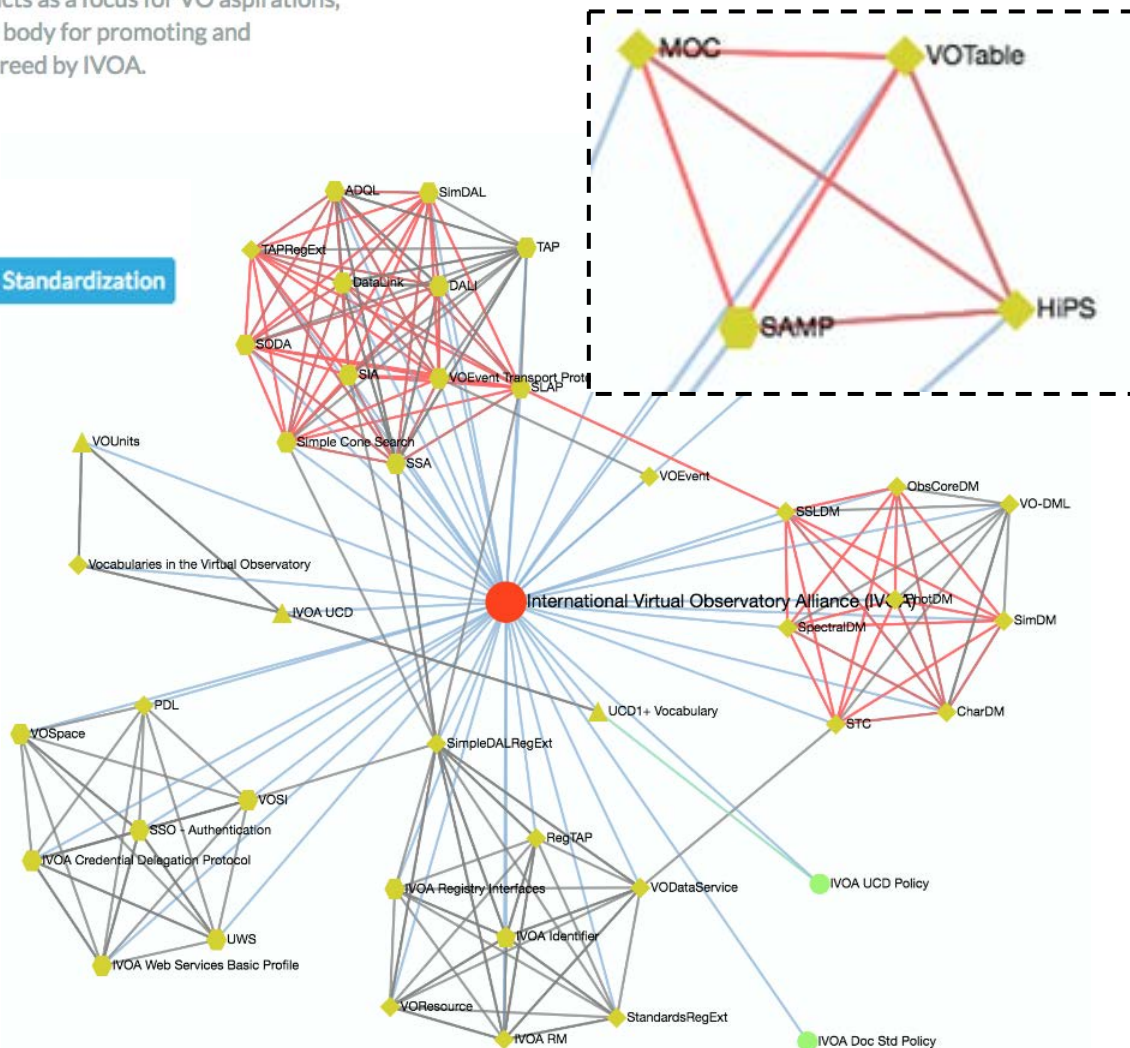
Record updated: Oct. 11. 2018. 9:57 a.m. by [carviset](#).

Scope and data types

- Astrophysics And Astronomy
- Data Standards
- Metadata Standardization



- DATABASE
- POLICY
- COLLECTION
- TERMINOLOGY ARTIFACT
- MODEL/FORMAT
- IDENTIFIER SCHEMA
- REPORTING GUIDELINE
- RECOMMENDS
- COLLECTS
- RELATED TO



FAIRsharing redesign – what's coming next?

- Redesign the data model
 - Split databases into repositories and knowledgebases
- Adding more fields to each record
- Adding more network graph tools
- Adding better search and manipulation tools

Your ideas are welcome!

FAIRsharing enables the FAIR principles

Ensures that standards, databases, repositories, policies are:

- **Findable**, e.g., by providing DOIs and marking up records in schema.org, allowing users to register, claim, maintain, interlink, classify, search and discover them
- **Accessible**, e.g., identifying their level of openness and/or license type
- **Interoperable**, e.g., highlighting which repositories implement the same standards to structure and exchange data
- **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 neighboring domains, rather than reinvention



FAIRsharing.org

 Adopters

 Collaborators

Researchers in
academia, industry,
government

Journal publishers
or organizations
with data policy

Learned societies,
unions and
associations

Developers and
curators of
resources

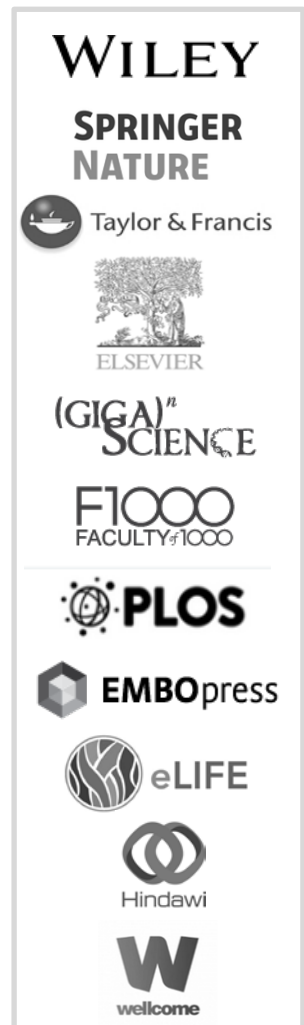
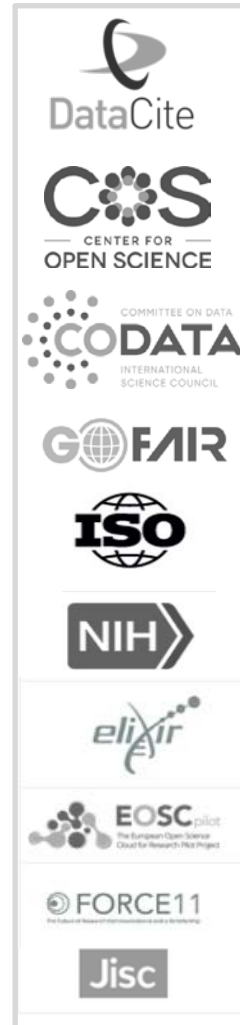
Research data
facilitators,
librarians, trainers

Funders and
data policy
makers

A flagship output of the:



Recommended by



Connections— building a
FAIR ecosystem

GO FAIR Initiative

Home > GO FAIR Initiative

- > **GO FAIR Initiative**
 - > **Vision and Strategy**
 - > **GO CHANGE**
 - > **GO TRAIN**
 - > **GO BUILD**
 - > **Governance**
 - > **Steering Committee**
 - > **Stakeholder Forum**
 - > **Executive Board**
 - > **GO FAIR Offices**
 - > **Map**
 - > **Contact**

GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the **FAIR data principles**, making data Findable, Accessible, Interoperable and Reusable. It offers an open and inclusive ecosystem for individuals, institutions and organisations working together through **Implementation Networks** (INs). The INs are active in three activity pillars: **GO CHANGE**, **GO TRAIN** and **GO BUILD**.



FAIR StRePo

[Home](#) › [Implementation Networks](#) › [Current Implementation Networks](#) › FAIR StRePo

*FAIR StRePo: making **Standards**, **Repositories**, and **Policies** FAIR*

Main tasks

- Map the landscape of standards, repositories and data policies across and related to the GO FAIR INs.
- Curate metadata associated with these digital objects and describe the relationships between them.
- Work with **OPEDAS IN** to provide support to the FAIR metrics to assess the FAIRness of digital objects.
- Work as part of **GO TRAIN** to provide assistance in the generation of a FAIR curriculum.
- Work with the **Data Stewardship Wizard** team to provide support for data management.
- Work as regular collaborators in the ongoing **Metadata for Machine (M4M) Workshop** series, to help self-identified communities to define and use domain-relevant community metadata standards in machine readable form for FAIRification and validation purposes.
- Collaborate with other efforts to standardize the selection and adoption of standards and repositories in data policies.

data policies in general, especially, and it works together with other GO FAIR INs to create digital objects (datasets, code, workflow, articles etc.) and data policies associated with them, creating informative and educational material, guidance, tools and services to serve producers and users of these resources. This IN touches and interoperates with a number of other GO FAIR INs. To achieve these objectives, the IN will leverage and expand on the already existing community-driven work of **FAIRsharing**. FAIRsharing is already involved in some GO FAIR activities that catalyse the **FAIRification** of data and services.

parency and
that supports
ducers to make
ty and standard-

itative of FAIRsharing
representative of FAIRsharing
representative of the GOFAIR Metabolomics IN
representative of the GOFAIR Chemistry IN
representative of the GOFAIR Food-System IN
representative of the GOFAIR FAIR Curriculum IN
representative of the FAIR Metrics.org and GOFAIR OPEDAS IN
representative of the GOFAIR GO-Train
representative of the GOFAIR GO-Train
representative of the Data Stewardship Wizard
representative of the Personal Health Train IN



Connections



Personal Health Train



TRAIN

FAIRassist.org



OPEDas

<http://w3id.org/AmIFAIR>

[10.1038/s41597-019-0184-5](https://doi.org/10.1038/s41597-019-0184-5)

Article | Open Access | Published: 20 September 2019

Evaluating FAIR maturity through a scalable, automated, community-governed framework

Mark D. Wilkinson, Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Mercè Crosas & Erik Schultes

Scientific Data 6, Article number: 174 (2019) | Cite this article

1781 Accesses | 1 Citations | 55 Altmetric | Metrics



[10.1162/dint_a_00038](https://doi.org/10.1162/dint_a_00038)

FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

Hana Píngl, Jekelova¹, Kristina Maria Hitzner², Peter Wübbeling³, Anriikka Miettinen⁴, Tobias Kuhn⁵, Robert Píngl⁶, Jan Škufel⁷, Peter An-Quilhan⁸, Barbara Mauerer⁹, Susanna-Assunta Sansone¹⁰, Markus Becker¹¹, Maurice Hering¹², Lucy L. Lawless¹³, Peter Phipps¹⁴, Erik Schultes¹⁵

¹CSI IBM International Support and Coordination Office, Karlsruhe

²Centre for Digital Scholarship, London University, London, UK

³Open Access Publishing and Data Services, Royal Holloway, Surrey

⁴Toronto University, Health Care Center, Toronto, Canada

⁵Department of Linguistic Science, Vrije Universiteit Amsterdam, 1018 XH Amsterdam, The Netherlands

⁶Cardiff Technical University in Higher Education, Cardiff, CF11 3ET, 100 Woodhouse, Cardiff, Wales

⁷Open Access Research Centre, Department of Informatics, University of Oxford, Oxford, OX1 3PS, United Kingdom

⁸Open Access Publishing and Data Services, Royal Holloway, Surrey

⁹IBM Research, Armonk, NY, USA

¹⁰Department of Linguistic Science, Vrije Universiteit Amsterdam, 1018 XH Amsterdam, The Netherlands

¹¹Open Access Publishing and Data Services, Royal Holloway, Surrey

¹²Open Access Publishing and Data Services, Royal Holloway, Surrey

¹³Open Access Publishing and Data Services, Royal Holloway, Surrey

¹⁴Open Access Publishing and Data Services, Royal Holloway, Surrey

¹⁵Open Access Publishing and Data Services, Royal Holloway, Surrey

Keywords: FAIR, interoperability, Challenges and Challenges, Convergence, FAIR, Convergence
© Author(s) 2019. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction in any medium, provided the original work is properly cited.



Matrix

FAIRsharing.org



StRePo



[10.1162/dint_a_00037](https://doi.org/10.1162/dint_a_00037)

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

Peter McQuilton¹, Dominique Batista², Oya Beyan³, Ramon Granell⁴, Susanna-Assunta Sansone⁵, Massimo Izzo⁶, Allyson L. Lister⁷, Robert Píngl⁸, Philippe Rocca-Serra⁹, Eric Schultes¹⁰, Peter Shunahara¹¹, Milo Thomson¹² & Susanna-Assunta Sansone¹³

¹Chaired Research Centre, Department of Engineering Sciences, University of Oxford, Oxford, OX1 3PS, UK

²Quantum Institute for Applied Information Technology, 10750 Appleton, Germany

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

⁴Cardiff Technical University in Higher Education, Cardiff, CF11 3ET, 100 Woodhouse, Cardiff, Wales

⁵Open Access Publishing and Data Services, Royal Holloway, Surrey

⁶Open Access Publishing and Data Services, Royal Holloway, Surrey

⁷Open Access Publishing and Data Services, Royal Holloway, Surrey

Keywords: Convergence, Data repositories, Standards, Data standards, FAIR data, FAIR enabling community standards

Citation: P. McQuilton, D. Batista, O. Beyan, R. Granell, S. Sansone, M. Izzo, A. L. Lister, R. Píngl, P. Rocca-Serra, E. Schultes, P. Shunahara, M. Thomson & S. 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



Metabolomics



Chemistry



Food-System



FAIR Curriculum
terms4FAIRskills

FAIR StRePo Projects

-  Matrix – mapping the landscape
- **terms4FAIRskills**
 - A terminology for data stewardship and FAIR curricula
 - <https://terms4fairskills.github.io/>
- **FAIRassist.org**
 - Discover resources that measure and improve FAIRness
 - <https://www.fairassist.org>

Join us!

- FAIRsharing.org
 - Help us map the IN Matrix
 - Register your repositories and standards in FAIRsharing
 - Create a Collection for your IN
- terms4FAIRskills
 - We are looking for more terminology annotators
 - Contact terms4FAIRskills@codata.org
- FAIRassist.org
 - Tell us what's missing
 - Register your resource/questionnaire



Collaboration: FAIRsharing.org and FAIR evaluation tools


The use of community **standards** for **(meta)data** and **identifiers** are among the FAIRness indicators

FAIRsharing content powers 2 (semi)automatic evaluation tools:

Article | [Open Access](#) | Published: 20 September 2019


Evaluating FAIR maturity through a scalable, automated, community-governed framework

Mark D. Wilkinson , Michel Dumontier, Susanna-Assunta Sansone , Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Mercè Crosas & Erik Schultes 

Scientific Data 6, Article number: 174 (2019) | [Download Citation](#) 

1393 Accesses | 54 Altmeteric | [Metrics](#) 

SPRINGER NATURE


 <https://doi.org/10.1038/s41597-019-0184-5>

FAIRassist.org

FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources

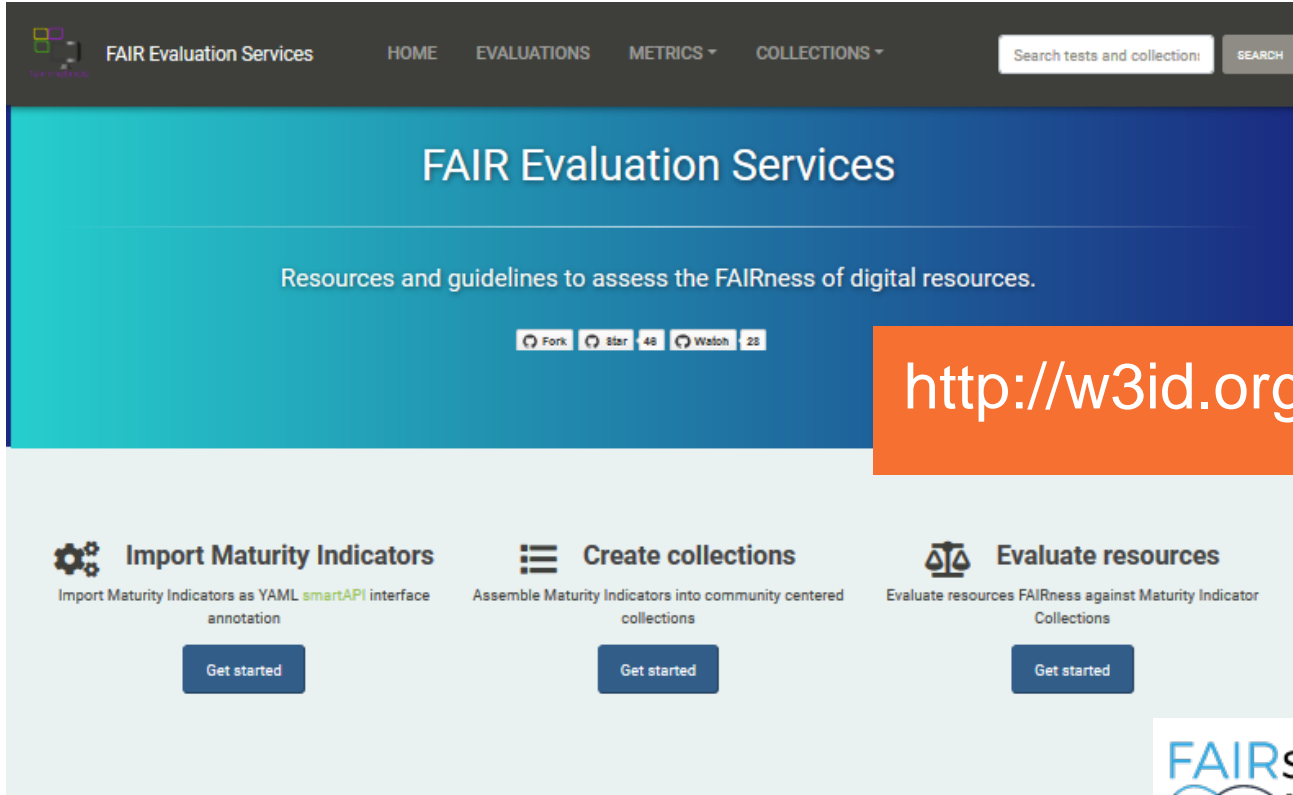
 Daniel J. B. Clarke, Lily Wang, Alex Jones,  Megan L. Wojciechowicz, Denis Torre,  Kathleen M. Jagodnik,  Sherry L. Jenkins,  Peter McQuilton, Zachary Flamholz, Moshe C. Silverstein,  Brian M. Schilder,  Kimberly Robasky, Claris Castillo, Ray Idaszak, Stanley C. Ahalt,  Jason Williams,  Stephan Schurer,  Daniel J. Cooper,  Ricardo de Miranda Azevedo,  Juergen A. Klenk, Melissa A. Haendel,  Jared Nedzel, Paul Avillach,  Mary E. Shimoyama, Rayna M. Harris, Meredith Gamble, Rudy Poten, Amanda L. Charbonneau,  Jennie Larkin,  C. Titus Brown,  Vivien R. Bonazzi,  Michel J. Dumontier,  Susanna-Assunta Sansone,  Avi Ma'ayan

doi: <https://doi.org/10.1101/657676>

 <https://doi.org/10.1101/657676>

The FAIR Evaluator

Designed as a bottom-up community effort, building on 'generation 1' FAIR metrics (human entry) to create 'generation 2' FAIR maturity indicators that use FAIRsharing metadata



The screenshot shows the homepage of the FAIR Evaluation Services website. The header includes the logo, navigation links (HOME, EVALUATIONS, METRICS, COLLECTIONS), and a search bar. The main content area features a large blue banner with the title 'FAIR Evaluation Services' and a subtitle 'Resources and guidelines to assess the FAIRness of digital resources.' Below the banner are social media icons for Fork, Star, and Watch. The footer contains three main sections: 'Import Maturity Indicators', 'Create collections', and 'Evaluate resources', each with a 'Get started' button. The FAIRsharing.org logo is visible in the bottom right corner.

<http://w3id.org/AmIFAIR>

Community driven

- Communities can decide which Maturity Indicators are relevant to them (working with FAIR data maturity model)
- These are registered in the Evaluator as a “Collection”, with some documentation about what MIs are included, and to what communities the Collection would be relevant
 - the purpose being re-use of Indicator Collections between similar communities/agencies
- Anyone can execute an evaluation on any GUID

FAIRsharing Metadata – powering the DSW questionnaires

The image displays two screenshots of web-based questionnaires. The left screenshot shows the 'FAIR Matrix Wizard' interface for 'BBMRI-ERIC (FAIR-Matrix-V1, 1.0.6)'. It features a sidebar with navigation options: Organization, Users, Knowledge Models, Questionnaires, and KM Editor. The main content area is titled 'III. Repository Questions' and includes a list of chapters (I to XI) and a 'More' section with a 'Summary Report' link. The current question is '1 Which repositories do you use for data?'. The right screenshot shows the 'DS Wizard' interface for 'Fairsharing Integration Demo (Fairsharing Integration Demo, 1.0.0)'. It features a sidebar with navigation options: Organization, Users, Knowledge Models, Questionnaires, and KM Editor. The main content area is titled 'I. Chapter 1' and includes a 'Current Phase' dropdown menu set to 'Before Submitting the Propo:'. The current question is '1 Select something from Fairsharing'. The dropdown menu is open, showing a list of options: 'The Protein Database', 'FAIRsharing https://fairsharing.org/bsg-d000448', 'protein', 'The Protein Database', 'LINCS Extended Metadata Standard: Proteins', 'Protein Data Bank Japan', 'Protein Classification Benchmark Collection', 'mini Protein Data Bank Format', and 'LINCS Pilot Phase 1 Metadata Standards: Protein Reagents'. A black arrow points from the 'FAIRsharing https://fairsharing.org/bsg-d000448' option in the top screenshot to the 'protein' option in the bottom screenshot.

Semantically-enabled drop-down menus and auto-complete functions (using data from FAIRsharing)

Gaps and hurdles

- Many of us, as well as many stakeholders (incl. publishers and funders) have been doing and supporting FAIR things before FAIR was a thing
- We need to reconcile views and needs, not just on paper
- Make this ecosystem participatory; easily said than done



Alastair Dunning
Susanna-Assunta Sansone
Marta Teperék

**THE LAYERED
CAKE OF FAIR
COORDINATION:
HOW MANY IS
TOO MANY?**

Lastly, it is essential that coordination efforts do not become echo chambers. In this layer cake of FAIR coordination, the higher you go, the further away you are from the researchers. So something needs to change. Whatever option is taken, it's clear that solving the current profusion of FAIR coordination projects simply by adding another layer of coordination might not be the best solution. While an extra layer may seem to offer a sweet way of bringing various ingredients together, the result can be a gooey mess.



FAIRsharing

WG

i Group details

Status: Completed

Chair (s): Susanna-Assunta Sansone, Rebecca Lawrence, Peter McQuilton, Simon Hodson

Secretariat Liaison: Lynn Yarmey

TAB Liaison: Paul Uhler

Case Statement: [Download](#)



WGs Maintaining deliverables (maintenance group)

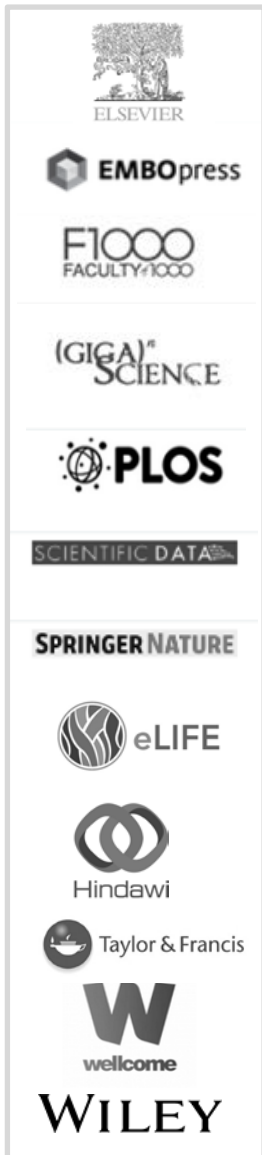
This WG has delivered its RDA-endorsed Flagship outputs:

FAIRsharing
registry

FAIRsharing
recommendations



Working with journal editors and publishers



nature
biotechnology

Correspondence | **OPEN** | Published: 02 April 2019 **Open Access CC-BY**

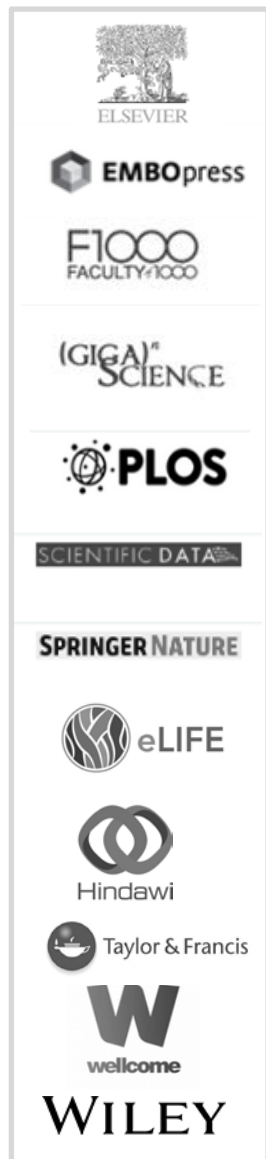
FAIRsharing as a community approach to
standards, repositories and policies

<https://doi.org/10.1038/s41587-019-0080-8>

69 authors (adopters, collaborators, users)
representing different stakeholder groups

Analysed the **data policies** by
journals/publishers, and the **standards** and
repositories they recommend

What have we learned and what are we doing now?



Discrepancy in recommendations across the data policies

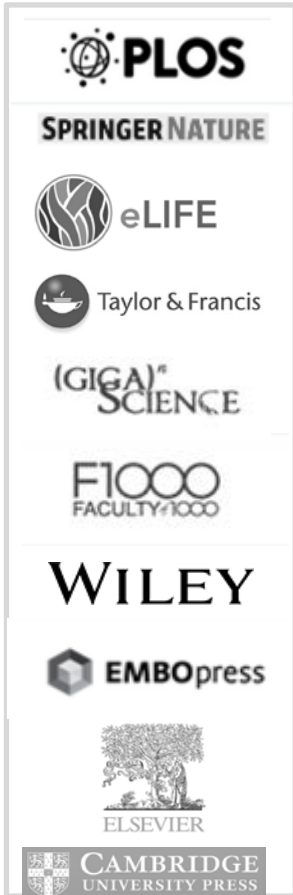
- some repositories are named, but very few standards are
- cautious approach due to the wealth of existing resources

Recommendations are often driven by

- the editor's familiarity with one or more standards, notably for journals or publishers focusing on specific disciplines
 - the engagement with learned societies and researchers actively supporting and using certain resources
- Consensus: FAIRsharing plays a key role in helping editors to discover and recommend appropriate resources, but repositories and standards could be more FAIR!

Data Repository Selection: Criteria That Matter

Started Jan 2018



We propose a **set of criteria** that journals and publishers believe are important for the **identification** and **selection** of **data repositories**, which can be recommended to researchers when they are preparing to publish the data underlying their findings



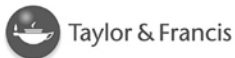
Data Repository Selection: Criteria That Matter
Pre-print:
 <https://osf.io/m2bce>

Objectives

Started Jan 2018



SPRINGER NATURE



(GIGA)[®]
SCIENCE

F1000
FACULTY1000

WILEY



CAMBRIDGE
UNIVERSITY PRESS

 DataCite

FAIRsharing.org

1. Guide **journals** and **publishers** in providing authors with consistent recommendations on data deposition
2. Reduce potential for confusion of **researchers** and **support staff**
3. Inform data **repository developers** and **managers** of the features believed to be important by journals and publishers
4. Apprise **certification** and other **evaluation initiatives**, serving as a reference and perspective from journals and publishers
5. Drive the curation in **FAIRsharing**, which will enable display, filter and search based on these criteria

Data Repository Selection: Criteria That Matter

Pre-print:

 <https://osf.io/m2bce>

Foreseen impact and next steps

Started Jan 2018



Our work will also **drive changes** by:

- Defining a common language across publishers;
- Helping publishers to maintain this information in a more automated way;
- Making the process for selection of recommended repositories more transparent to all stakeholders

The criteria are available and we are ready for your feedback – <https://tinyurl.com/RepoCriteriaFeedback>

Once agreed, we will add the criteria into FAIRsharing

 DataCite
FAIRsharing.org

Data Repository Selection: Criteria That Matter

Pre-print:

 <https://osf.io/m2bce>

FAIRsharing.org

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

Find

Recommendations

Standards and/or databases recommended by journal or funder data policies.

Discover

Collections

Standards and/or databases grouped by domain, species or organization.

Learn

Educational

About standards, their use in databases and policies, and how we can help you.



Peter
McQuilton

Project
Coordinator

with developers and curators

Advisory Board

See also the FAIRsharing [RDA](#) and [Force11](#) WG webpages.

Emma Ganley (PLOS) Co-chair
Varsha Khodiyar (NPG) Co-chair
Michael Ball (ESRC)
Theo Bloom (BMJ)
Jennifer Boyd (OUP)
Dave Carr (The Wellcome Trust and Wellcome Open Research)
Helena Cousijn (Datacite)
Scott Edmunds (GigaScience, BGI)
Dominic Fripp (JISC)
Chris Graf (Wiley)
Simon Hodson (CODATA), *Co-Chair of the RDA/Force11 WG*
Mike Huerta (Coordinator of Data & and Open Science Initiative, Associate Director for Programme Development at

the NIH National Library of Medicine)
Amye Kenall (BMC)
Rebecca Lawrence (F1000), *Co-Chair of the RDA/Force11 WG*
Thomas Lemberger (EMBO Press)
Jennifer Lin (CrossRef)
Luiz Olavo Bonino (GO-FAIR)
Gabriella Rustici (ELIXIR UK Node, University of Cambridge, UK)
Marina Soares E Silva (Elsevier)
Imma Subirats (Information Management Officer, FAO of the United Nations, Italy)
Marta Teperek (Data Stewardship Coordinator, TUDelft, The Netherlands)

Data Readiness Group

datareadiness.eng.ox.ac.uk

