

FAIR Data in the Scholarly Communications Lifecycle

Chris Erdmann

Engagement, Support, Training Expert
RENCI | NHLBI BioData Catalyst | GO FAIR US
@libcce | erdmannc@renci.org

FAIR image/logo https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

Data infused research lifecycle... and software!

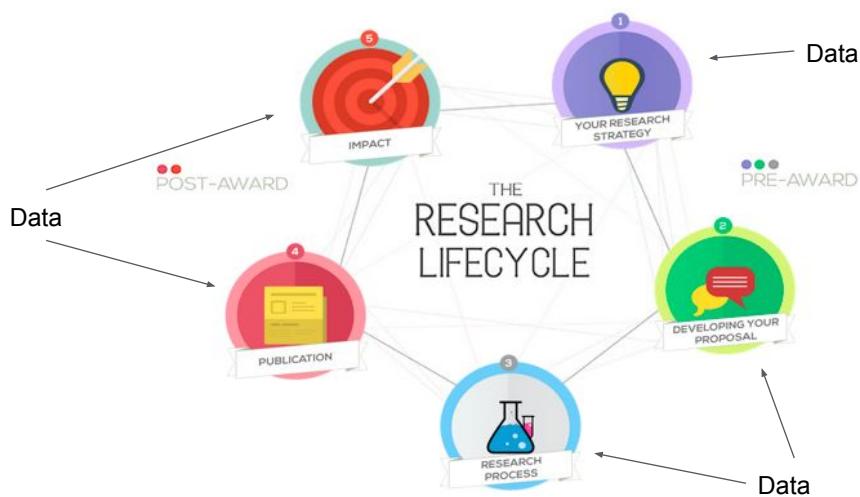


Diagram: [Bournemouth University UK](#)

How to support FAIR?

FAIR is a shared responsibility

FAIR needs an ecosystem of services

FAIR needs people

Open (research) needs FAIR



What is motivating you to learn more about FAIR?

For this session, we will be using a collaborative Etherpad to collect responses:

<https://pad.carpentries.org/FAIRlifecycle>

Please take a minute to bring up the Etherpad in your browser.

Note: When responding to questions in the Etherpad, if someone has a similar response, you can add a “+” to it. Also, add your first names to the responses, just in case we might want to follow up on the call to discuss further or for clarification.

FAIR Motivations

- Research Practices / Culture
- Research Impact
- Institutional Policies / Support
- Infrastructure / Publishing
- Funders
- Legal
- Just Interested



National Institutes
of Health



U.S. DEPARTMENT OF
ENERGY

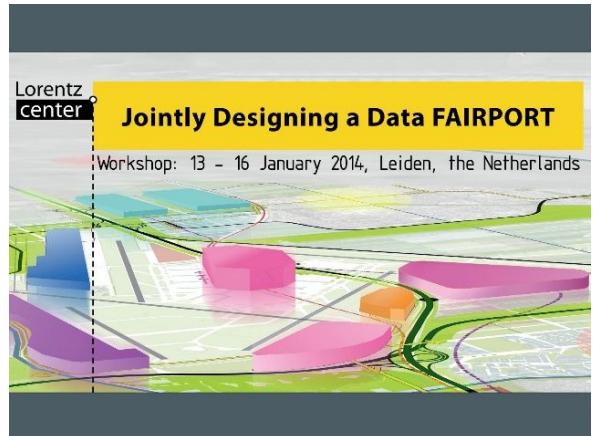
Outline

- Background
- FAIR principles in more depth
- Understanding what FAIR means in your context (assessment)
- FAIR examples
- Types of output (research software)
- Communities to follow
- Q&A

Origins of FAIR

2014 Leiden, Lorentz Center Workshop

- Not to be Prescriptive
- Community Driven
- Avoid the Standards Path
- Avoid Narrative w/ Trapped Data
- Machine Actionable
- Guiding Principles



Data FAIRport (Lorentz Workshop)

<https://www.datafairport.org/>

MENU ▾

SCIENTIFIC DATA



Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons✉

Scientific Data 3, Article number: 160018 (2016) | Cite this article

128k Accesses | 1650 Citations | 1568 Altmetric | Metrics

An Addendum to this article was published on 19 March 2019

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).

<https://doi.org/10.1038/sdata.2016.18>

15 FAIR 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

15 principles total



Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

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 they describe

F4. (Meta)data are registered or indexed in a searchable resource

Go FAIR Principles

<https://www.go-fair.org/fair-principles/>

What is FAIR?

The FAIR universe is populated by concepts and acronyms that you may not be familiar with. You will learn about all of them, as you make your way through this website. On this page, you can read about three fundamental concepts: the FAIR principles, FAIR data, and FAIRification



- Humans and machines as digesters of data
- FAIR Principles apply to both metadata and data
- As Open as possible, as closed as necessary
- The Principles are not rules or standards
- Documentation, file formats, RDM training...
- FAIR from the disciplines

How to FAIR

<https://www.howtofair.dk/>

How are you responding to the principles?

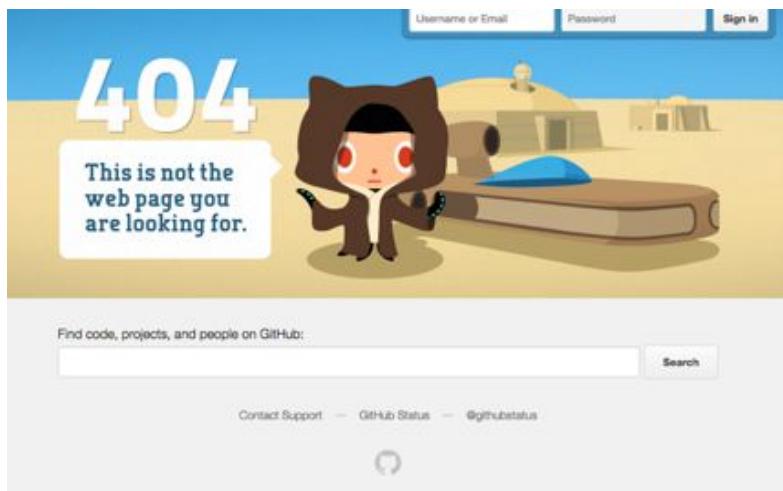
Briefly describe how you are responding (or would respond) to the subset of principles (below) in the Etherpad:

<https://pad.carpentries.org/FAIRlifecycle>

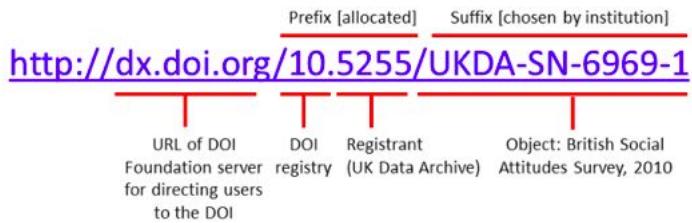
- F1: (Meta) data are assigned globally unique and persistent identifiers
 - R1.3: (Meta) data meet domain-relevant community standards
- You can also find these listed in the Etherpad...

F1: (Meta) data are assigned globally unique and persistent identifiers

What's the problem?

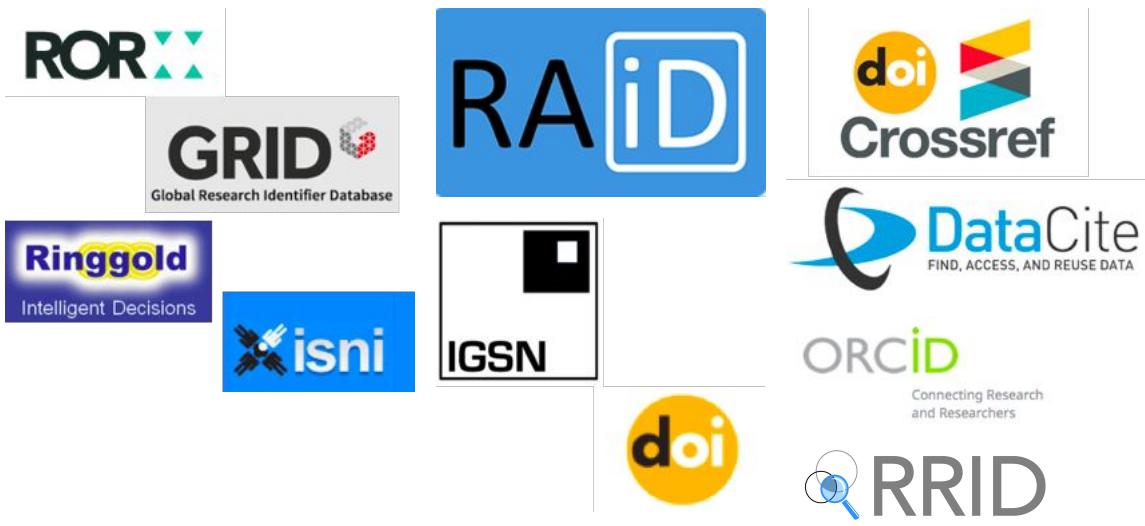


What does a persistent identifier (PID) look like?



Example: a DataCite Digital Object Identifier (DOI)

There are many PIDs and PID providers to choose from



Also RRID <https://www.rrids.org/> and Identifiers.org <http://identifiers.org/>

RESEARCH-ARTICLE

An Open Ecosystem for Pervasive Use of Persistent Identifiers



PEARC '20: Practice and Experience in Advanced Research Computing July 2020
Pages 99–105 <https://doi.org/10.1145/3311790.3396660>

MINIDs

Pidforum.org - for the harder PID questions

the PID Forum

[Sign Up](#) [Log in](#)



[all categories](#)

Categories

[Latest](#)

[Top](#)

Category

Topics

General

Topics that don't need a category, or don't fit into any other existing category.

Latest



[Welcome to the PID Forum!](#)

0
Jun '19

PID Best Practices

A category to bring together information (papers, guidelines etc) and ideas on PID best practices for different communities and disciplines.

11

Translations of the Power of PID educational video

1
3d

PID Resources in Languages Other than English

PID News & Blogs

Share interesting PID news & blogs here

27



[Global PID survey](#)

0
3d

PID News & Blogs

PID Graph

Persistent identifiers and associated metadata describe resources such as datasets, software, publications, people, research organizations, funders, and grants. An important part of this metadata is the description of connections between these resources. Together these resources and their connection...

27

PID Policy consultation closing 20th August

0
5d

PID Policy for EOSC

PID Services

Post anything related to PID Services here.

3



[Ambassador Competition Now Open](#)

0
10d

FREYA Ambassadors Chat Room

Recommendations for Services in a FAIR Data Ecosystem -> Collaboration between FREYA, RDA Europe, FAIRsFAIR, OpenAIRE, and EOSC-hub

0
11d

Jisc PID Agency Project

The Tickell report recommended that Jisc lead sector agreement and adoption of a basket of 'unique identifiers'

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774956/Open-access-to-research-publications-2018.pdf

Developing a persistent identifier roadmap for open access to UK research.

<http://repository.jisc.ac.uk/7840/>

Open access to research publications - 2018

Independent advice

Professor Adam Tickell
Vice-Chancellor, University of Sussex
Chair of the Universities UK Open Access Coordination Group

June 2018

The proposed UK PID consortium would both help enable the UK's 'open research infrastructure' and also support the use of open PID infrastructures as needed by the community.

Joint Information Systems Committee (JISC)

Five priority PIDs

Entities	Communities	Repository community	Infrastructure community	Publishers	Librarians	Research Funders	Research Managers
People							
				Focus group for PIDs related to people eg ORCID			
Institutions							
				Focus group for PIDs related to institutions eg ISNI, ROR			
Grants							
				Focus group for PIDs related to funded grants eg Crossref Grant ID			
Projects							
				Focus group for PIDs related to projects eg RAiD			
Outputs							
				Focus group for PIDs related to outputs eg DOI			

R1.3: (Meta)data meet domain-relevant community standards

What is metadata?

Standardised

Structured

Machine and human readable

<https://www.wikidata.org/wiki/Q180160>

Source: <https://flickr.com/photos/33255628@N00/8071729256>

**METADATA IS A
LOVE NOTE
TO THE FUTURE**

Metadata is a subset of documentation: Documentation is a catch-all term which can include some very high-level, human-readable, loosely structured information about the dataset – for example a description of the laboratory method used to generate a set of results or a sketch book accompanying a work of sculpture. The term metadata has come to define a subset of documentation information that uses standardised terms and is presented in a structured way. There are different categories of metadata, including descriptive, preservation, technical and administrative metadata.

Minimum metadata required: This could correspond to the DataCite mandatory minimum metadata set although it is very likely that other elements may be required, whether by the repository or by the research funder. This metadata set is designed for citation and disambiguation, in other words ensuring that the dataset a researcher is reusing is identical to one cited in a journal article or online. This set of metadata does little to make the resource visible to researchers speculatively searching for relevant data to reuse. Most repositories only require a few mandatory fields to reduce the burden on depositors.

Providing good documentation: Researchers should provide good documentation at the study level such as a description of the research methodology employed to create your data. There should also be sufficient documentation at the file, item and variable level to enable a third party to understand it. This could be ensuring that Excel spreadsheets have sensible row and column descriptions or that a document is included with the dataset which properly explains any abbreviations used.

For supporting FAIR, data must be findable by both humans and machines. The best way to do this is to encourage researchers to deposit their research data with a FAIR-aligned repository that is recognised in their field. This way, their data will be searchable and discoverable online and the repository will employ domain-specific metadata standards.

Domain specific standards for description

RDA Metadata Standards Directory

<https://rd-alliance.github.io/metadata-directory/standards/>

FAIRsharing <https://fairsharing.org/standards/>



Social and Behavioral Sciences

Standards

[Add](#)

[DDI \(Data Documentation Initiative\)](#) [Edit](#)

A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel:

- DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.
- DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.

Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.

Project policies for (meta)data

Multidisciplinary drifting Observatory for the Study of the Arctic Climate (MOSAiC) data policy

Metadata shall make data findable and provide additional contextual information about measurement details, methods, relevance, lineage, quality, usage and access restrictions of the data. It shall allow coupling users, software, and computing resources to the data. Hence, metadata must be machine-readable and interpretable as well as human-understandable. Furthermore, metadata for each data set should follow the FAIR data principles in terms of fitness for purpose and fitness for re-use. The metadata should be agreed on, listed, and explained within the MSOPs.

Includes recommendations for metadata and vocabularies, listing examples for oceanography, climatology, and modelling; biology; provenance.

Sources: <https://mosaic-expedition.org/science/mosaic-data/>; <https://spaces.awi.de/display/EFPW/MOSAiC+Data+Policy>
[Direct link to data policy \(pdf\)](#)

Metadata Standards / Templates / Communities



schema.org



The CodeMeta Project



Play with metadata in a sandbox



Recent uploads

January 1, 2016 Thesis Open Access

[Kritische Soziale Arbeit und ihr Gegenstand: Eine kritische Auseinandersetzung](#)

[View](#)

Stalder, Bruno; Vifian, Karin

In vorliegender Bachelorarbeit „Kritische Soziale Arbeit und ihr Gegenstand – Eine kritische Auseinandersetzung“ beschreiben die Autorin Karin Vifian und der Autor Bruno Stalder die Perspektive kritischer Sozialer Arbeit in Theorie und Praxis. Die kritische Soziale Arbeit, deren wichtigste...

Uploaded on August 30, 2016

Zenodo Sandbox: <https://sandbox.zenodo.org/>

Zenodo Sandbox

<https://sandbox.zenodo.org/>

Understanding what FAIR means in your context



FAIR questions

FINDABLE

1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository? 
 Yes
 No
2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others? 
 Yes
 No
3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans? 
 Yes
 No

ACCESSIBLE

4. Are you aware that access to your dataset may need to be controlled and that metadata should include licence information under which the data can be reused? 
 Yes
 No
5. Are you aware that metadata should remain available over time, even if the data is no longer accessible? 
 Yes
 No

FAIR Aware <https://fairaware.dans.knaw.nl/>

More on Assessment

This lesson is still being designed and assembled (Pre-Alpha version)



Library
Carpentry

Home

Code of Conduct

Setup

Episodes ▾

Extras ▾

License

Improve this page

Search...



Library Carpentry: FAIR Data and Software



Assessment

?

Overview

Teaching: 10 min
Exercises: 50 min

Questions

- How can I assess the FAIRness of myself, organisation, service, community... ?
- Which FAIR assessment tools exist to understand how FAIR you are?

Objectives

- Assess the current FAIRness level of myself, organisation, service, community...
- Know about available tools for assessing FAIRness.
- Understand the next steps you can take to being FAIRer.

Library Carpentry: FAIR Data and Software

<https://librarycarpentry.org/lc-fair-research/07-assessment/index.html>

FAIR Zenodo Principles

The screenshot shows the Zenodo website's navigation bar. The logo 'zenodo' is on the left. To its right, the word 'About' is highlighted in a larger font. A horizontal line separates this from the rest of the menu. On the right side of the line, there are several links: 'About', 'Blog', 'Help', and 'Developers'. Below this line, there are more links: 'About', 'Privacy Policy', 'Terms of Use', 'General Policies', 'Infrastructure', 'Principles' (which is also highlighted in a larger font), 'Roadmap', and 'Contact'.

Principles

Best Effort Principles

Zenodo does not sign SLAs (service-level agreements). This is not a weakness, it is by design and marks a philosophy that we believe is most appropriate for Science. Instead, Zenodo is run by leading practitioners according to best practices.

What Science needs is inherent reliability, or more accurately demonstrated reliability based on open best practices. Furthermore the users should be able to influence these best practices. In the long-term, a service which is trusted is much more valuable than one for which assurances must be bought.

Service failure can never be undone. Enforcing an SLA means being prepared to litigate against the contract, which means compensation, frequently assessed on the basis of loss of revenue... but none of these concepts have any place or relevance in the free exchange of research results!

Living by these principles, Zenodo strives to make available architecture, implementation, practices and statistics. Please see for example the infrastructure page. We are also aiming to have these certified.

<https://about.zenodo.org/principles/>

Repository Checklist

- General / discipline / object
- Suits user needs
- Persistent and unique identifier
- Landing page / metadata to find, cite
- Access, download usage metrics
- Availability in the long term
- Responsive to community
- Certificates / standards
- Clear terms, conditions, licensing



See OpenAire's Guides for Researchers:

[How to select a repository?](#)

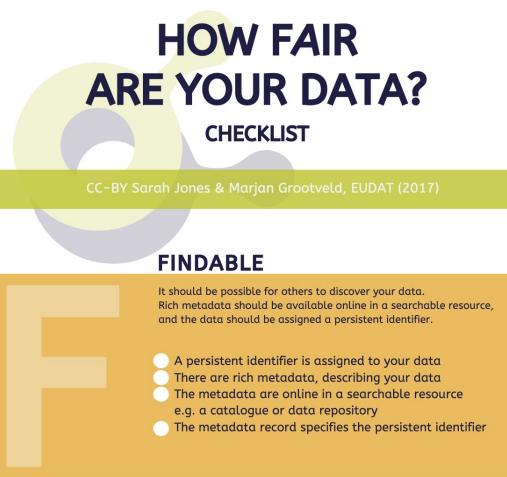
OpenAire's Guides for Researchers: How to select a repository?

<https://www.openaire.eu/opendatapilot-repository-guide>

<http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/>

<http://www.copdess.org/enabling-fair-data-project/author-guidelines/>

FAIR Checklists



The Machine Learning Reproducibility Checklist (v2.0, April 7, 2020)

For all **models** and **algorithms** presented, check if you include:

- A clear description of the mathematical setting, algorithm, and/or model.
- A clear explanation of any assumptions.
- An analysis of the complexity (time, space, sample size) of any algorithm.

For any **theoretical claim**, check if you include:

- A clear statement of the claim.
- A complete proof of the claim.

For all **datasets** used, check if you include:

- The relevant statistics, such as number of examples.
- The details of train / validation / test splits.
- An explanation of any data that were excluded, and all pre-processing steps.
- A link to a downloadable version of the dataset or simulation environment.
- For new data collected, a complete description of the data collection process, such as instructions to annotators and methods for quality control.

For all shared **code** related to this work, check if you include:

- Specification of dependencies.
- Training code.
- Evaluation code.
- (Pre-)trained model(s).
- README file includes table of results accompanied by precise command to run to produce those results.

For all reported **experimental results**, check if you include:

- The range of hyper-parameters considered, method to select the best hyper-parameter configuration, and specification of all hyper-parameters used to generate results.
- The exact number of training and evaluation runs.
- A clear definition of the specific measure or statistics used to report results.
- A description of results with central tendency (e.g. mean) & variation (e.g. error bars).
- The average runtime for each result, or estimated energy cost.
- A description of the computing infrastructure used.

Reproduced from: www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist-v2.0.pdf

Jones, Sarah, & Grootveld, Marjan. (2017, November). How FAIR are your data?. Zenodo.

<http://doi.org/10.5281/zenodo.3405141>

The Machine Learning Reproducibility Checklist

<https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf>

FAQs / Guidelines

COPDESS

Coalition for Publishing Data in the Earth and Space Sciences

The Coalition for Publishing Data in the Earth and Space Sciences ▾

Enabling FAIR Data Project ▾



ENABLING FAIR DATA – FAQS

[HOME](#) / [ENABLING FAIR DATA PROJECT](#) / ENABLING FAIR DATA – FAQS

[http://www.copdess.org/enabling-fair-data-project%E2%80%BA
/enabling-fair-data-faqs/](http://www.copdess.org/enabling-fair-data-project%E2%80%BA/enabling-fair-data-faqs/)

Core Trust Seal



Home About ▾ Certification ▾ Certified Repositories ▾ Apply ▾ Contact ▾



TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Towards Sustainable Data Infrastructures

See the Requirements for 2020–2022

Core Trust Seal <https://www.coretrustseal.org/>

TRUSTworthy Digital Repositories

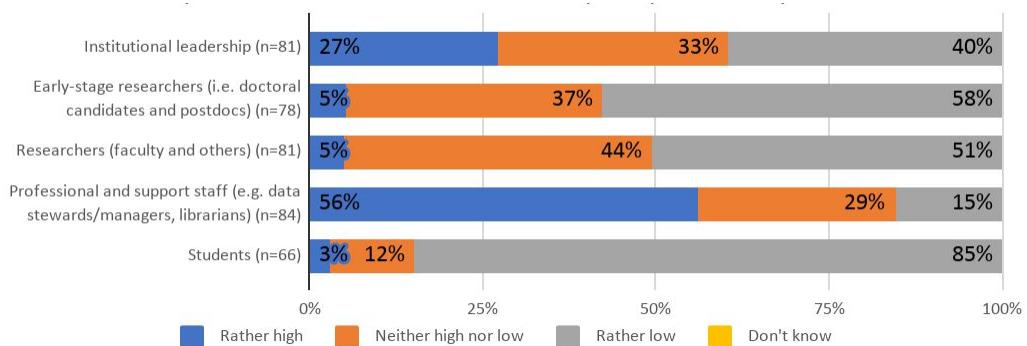
“...to make data FAIR whilst preserving them over time requires trustworthy digital repositories (TDRs) with sustainable governance and organizational frameworks, reliable infrastructure, and comprehensive policies supporting community-agreed practices”

Principle	Guidance for Repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Lin, D., Crabtree, J., Dillo, I. et al. **The TRUST Principles for digital repositories**. *Sci Data* 7, 144 (2020).
<https://doi.org/10.1038/s41597-020-0486-7>

Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020).
<https://doi.org/10.1038/s41597-020-0486-7>

How would you rate the awareness of the FAIR principles within your institution?



FAIR in European Higher Education (FAIRsFAIR) <https://doi.org/10.5281/zenodo.3629682>

#FAIRexamples

Additional Examples of FAIR in Disciplines

Top 10 FAIR Data & Software Things

about github repository download/cite license #top10fair

The Top 10 FAIR Data & Software Things are brief guides (stand alone, self paced training materials), called "Things", that can be used by the research community to understand how they can make their research (data and software) more FAIR ([Findable, Accessible, Interoperable and Reusable](#)). Each discipline/topic has its own specific list:

Nanotechnology

Astronomy

Linked Open Data

Imaging

Top 10 FAIR Data & Software Things

<https://librarycarpentry.org/Top-10-FAIR/>

Who is the FAIRest one of all?

Mirror, Mirror on the Wall: TKFDM announces “FAIRest Dataset” Award

Kiesel, Johannes, Potthast, Martin, Hagen, Matthias, Kneist, Florian, & Stein, Benno. (2017).

Webis-Web-Archive-17 (Version 1.0.0) [Data set].

Zenodo. <http://doi.org/10.5281/zenodo.1002204>

Rath, Michael, Mäder, Patrick. (2019). The SEOSS Dataset - Requirements, Bug Reports, Code History,

and Trace Links for Entire Projects. Harvard

Dataverse, V1. <https://doi.org/10.7910/DVN/PDDZ4Q>

THÜRINGER
FDM-TAGE
2020



Wettbewerb
FAIRest Dataset

Mirror, Mirror on the Wall: TKFDM announces “FAIRest Dataset” Award

<https://forschungsdaten-thueringen.de/nachricht/start-des-fairest-dataset-wettbewerb.html>

Turning FAIR into reality



Lowndes, J., Best, B.,
Scarborough, C. et al. **Our path
to better science in less time
using open data science tools.**
Nat Ecol Evol 1, 0160 (2017).
<https://doi.org/10.1038/s41559-017-0160>

Software and data skills training



THE
CARPENTRIES

<https://carpentries.org/>

Share your #FAIRexamples in the Etherpad

<https://pad.carpentries.org/FAIRlifecycle>

Types of Research Output... Research Software.



Image source:

<https://blog.wellcomeopenresearch.org/2018/07/12/supporting-our-researchers-in-shaping-data-underpinning-publications/>



Digital Object Identifiers (DOI) are the backbone of the academic reference and metrics system. If you're a researcher writing software, this guide will show you how to make the work you share on GitHub citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool [Zenodo](#).

ProTip: This tutorial is aimed at researchers who want to cite GitHub repositories in academic literature. Provided you've already set up a GitHub repository, this tutorial can be completed without installing any special software. If you haven't yet created a project on GitHub, start first by [uploading your work to a repository](#).

Make Your Code Citable

<https://guides.github.com/activities/citable-code/>

Software Heritage Library



Software Heritage

Mission ▾ Archive ▾ Community ▾ Support ▾ About ▾

```
13 */
14 static int do_sched_cfs_period_timer(struct cfs_bandwidth *cfs_b, u64 expires);
15 {
16     u64 runtime, runtime_expires;
17     int throttled;
18
19     /* no need to continue the timer with no bandwidth remaining */
20     if (cfs_b->quota == RUNTIME_INF)
21         goto out_deactivate;
22
23     throttled = !list_empty(&cfs_b->throttled_sds_rq);
24     cfs_b->nr_pericda += overrun;
25
26     /*
27      * idle depends on !throttled (for the case of a single period)
28      * we're going inactive then everything else can be disabled
29      */
30     if (cfs_b->idle && !throttled)
31         goto out_deactivate;
32
33     __refill_cfs_bandwidth(runtime(cfs_b));
34
35     if (!throttled) {
36         /* mark as potentially idle for the upcoming timer */
37         cfs_b->idle = 1;
38         return 0;
39     }
40
41     /* account preceding periods in which throttling occurred */
42     cfs_b->nr_throttled += overrun;
43
44     runtime_expires = cfs_b->runtime_expires;
45
46 }
```

Software [is our] Heritage

Software Heritage

<https://www.softwareheritage.org/>

Software DOI, now what? Related identifiers

The screenshot shows the DataCite website with a navigation bar at the top featuring the DataCite logo, Feedback, and Home links. Below the navigation is a breadcrumb trail: Guides > DataCite Metadata Schema 4.3. The main content area is titled "DataCite Metadata Schema 4.3". On the left, there are two sections: "GETTING STARTED" and "DEVELOPER DOCUMENTATION", each with a list of links. The "GETTING STARTED" section includes "Getting Started", "Contact DataCite", "Transferring DOIs and Prefixes", "Testing Guide", and "DataCite or Crossref". The "DEVELOPER DOCUMENTATION" section includes "DataCite REST API Guide", "DataCite MDS API Guide", "DataCite GraphQL API Guide", "DataCite EZ API Guide", "DataCite OAI-PMH Guide", and "DataCite Content Negotiation". To the right of the documentation sections is a callout box with a blue icon and the text "PDF is the version of record". It explains that the official home for the DataCite Metadata Schema is at <https://schema.datacite.org>, and that PDFs of the schema documentation found there are the versions of record. It notes that these pages on the support site are provided for convenience, but the PDF is the official version. A citation for the PDF version is also provided.

DataCite Metadata Schema 4.3

GETTING STARTED

- Getting Started
- Contact DataCite >
- Transferring DOIs and Prefixes
- Testing Guide >
- DataCite or Crossref

DEVELOPER DOCUMENTATION

- DataCite REST API Guide >
- DataCite MDS API Guide
- DataCite GraphQL API Guide
- DataCite EZ API Guide
- DataCite OAI-PMH Guide
- DataCite Content Negotiation

PDF is the version of record

The official home for the DataCite Metadata Schema is at <https://schema.datacite.org>, and the PDFs of the schema documentation found there are the versions of record. These pages on our support site are provided as a convenience. We try to duplicate the PDF as much as possible, but the PDF is the official version.

Citation for the PDF version: DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. DataCite e.V.

Towards FAIR principles for research software

Cite

Issue title: FAIR Data, Systems and Analysis

Guest editors: Paul Groth and Michel Dumontier

Article type: Position Paper

Authors: Lamprecht, Anna-Lena^{a,*} | Garcia, Leyla^b | Kuzak, Mateusz^{c,d} | Martinez, Carlos^e | Arcila, Ricardo^f | Martin Del Pico, Eva^g | Dominguez Del Angel, Victoria^h | van de Sandt, Stephanieⁱ | Ison, Joniⁱ | Martinez, Paula Andrea^k | McQuilton, Peter^l | Valencia, Alfonso^{m,n} | Harrow, Jennifer^o | Psomopoulos, Fotis^p | Gelpi, Josep LL.^{q,r} | Chue Hong, Neil^{s,t} | Goble, Carole^u | Capella-Gutierrez, Salvador^{v,w}

Affiliations: [a] Utrecht University, The Netherlands. E-mail: a.Llamprecht@uu.nl | [b] ZBMED Information Centre for Life Sciences, Germany. E-mail: l.garcia@zbmed.de | [c] Netherlands eScience Center, The Netherlands | [d] Dutch Techcentre for Life Sciences, The Netherlands. E-mail: m.kuzak@esciencecenter.nl | [e] Netherlands eScience Center, The Netherlands. E-mail: c.martinez@esciencecenter.nl | [f] EMBL-EBI, UK. E-mail: arcila@ebi.ac.uk | [g] Barcelona Supercomputing Center (BSC), Spain. E-mail: eva.martin@bsc.es | [h] L'Institut Français de Bioinformatique (IFB), France. E-mail: victoria.dominguez@france-bioinformatique.fr | [i] CERN, Switzerland. E-mail: stephanie.van.de.sandt@cern.ch | [j] National Life Science Supercomputing Center, Technical University of Denmark, Denmark. E-mail: jison@cbs.dtu.dk | [k] National Imaging Facility, Australia. E-mail: p.martinez@uq.edu.au | [l] Oxford e-Research Centre, UK. E-mail: peter.mcquilton@oerc.ox.ac.uk | [m] Barcelona Supercomputing Center (BSC), Spain | [n] Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain. E-mail: alfonso.valencia@bsc.es | [o] ELIXIR Hub, UK. E-mail: harrow@ebi.ac.uk | [p] Institute of Applied Biosciences, CERTH, Greece. E-mail: fpsom@certh.gr | [q] Barcelona Supercomputing Center (BSC), Spain | [r] University of Barcelona, Spain. E-mail: gelpi@ub.edu | [s] Software Sustainability Institute, UK | [t] EPCC, University of Edinburgh, UK. E-mail: n.chuehong@software.ac.uk | [u] University of Manchester, UK. E-mail: carole.goble@manchester.ac.uk | [v] Barcelona Supercomputing Center (BSC), Spain. E-mail: salvador.capella@bsc.es

Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 – 59.

<https://content.iospress.com/articles/data-science/ds190026>

A Fresh look...

<https://arxiv.org/abs/2101.10883>

Software Quality Checklists/Management Plans

- [CLARIAH](#) (Pages 38-42)
- [SSI](#)
- [EURISE](#)
- [Deutsches Zentrum für Luft- und Raumfahrt](#) (Class 1)
- [ELIXIR](#) (SMP)

Software Quality Checklist

General

- Does the software have a descriptive name?
- Is there a short high-level description of the software?
- Is the purpose of the software clear?
- Is the targeted audience of the software clear?
- Does it (and its dependencies) use OSI approved licenses?
- Is the software under version control?
- Is there a website for the software?
- Does the software have a release mechanism?
- Is the software available in packaged format or only sources?
- Are maintainer and development status clear, including contact information?
- Are the requirements listed and up to date?
- Is the interface responsive and accessible?
- Is copyright and authorship clear?
- Is there a contribution guide?

EURISE Network Software Quality Checklist

<https://github.com/eurise-network/technical-reference/blob/v0.1/quality/software-checklist.rst>

Top 10 FAIR Data & Software Things & Recommendations

TOP 10 FAIR DATA & SOFTWARE THINGS

Research Software

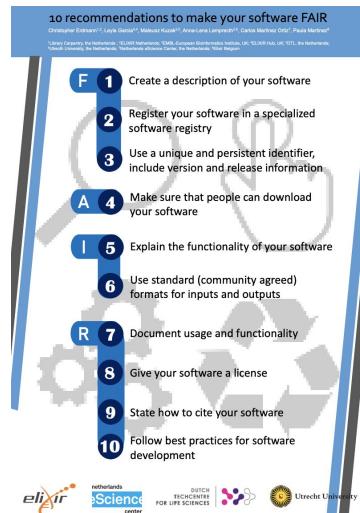
Sprinters

Anna-Lena Lamprecht, Carlos Martinez Ortiz, Chris Erdmann, Leyla Garcia, Mateusz Kuzak, Paula Andrea Martinez

Description:

The [FAIR data principles](#) are widely known and applied today. What the FAIR principles mean for (scientific) software is an ongoing discussion. However, there are some things on which there is already agreement that they will make software (more) FAIR. In this document, we go for some 'low hanging fruit' and describe 10 easy FAIR software things that you can do. To limit the scope, "software" here refers to scripts and packages in languages like R and Python, but not to other kinds of software frequently used in research, such as web-services, web platforms like myexperiment.org or big clinical software suites like OpenClinica.

A [poster summarizing these 10 FAIR software things](#) is also available.



Top 10 FAIR Data & Software Things: Research Software

<https://librarycarpentry.org/Top-10-FAIR//2018/12/01/research-software/>

10 recommendations to make your software FAIR

<https://f1000research.com/posters/8-1541>

FIVE RECOMMENDATIONS FOR FAIR SOFTWARE

LET'S GO! →



WHAT'S FAIR

ABOUT US

Five Recommendations for FAIR Software

<https://fair-software.eu/>

Netherlands eScience Center : Research Software Directory

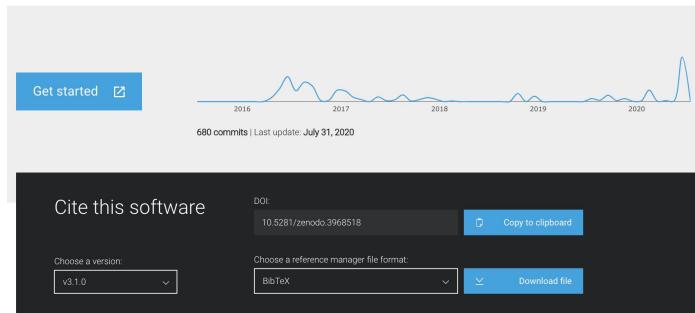
Search the directory

Metrics About

mcfly

7
mentions
6
contributors

Helps you find a suitable neural network configuration for deep learning on time series.



Netherlands eScience Center : Research Software Directory

<https://www.research-software.nl/>

Papers with Code



Search for papers, code and tasks



Browse State-of-the-Art

Methods

Trends

About



Log In/Register

Trending Research

Trending

Latest

Greatest

✉ Subscribe



The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies

28 Apr 2020 • salesforce/ai-economist

In experiments conducted on MTurk, an AI tax policy provides an equality-productivity trade-off that is similar to that provided by the Saez framework along with higher inverse-income weighted social welfare.



2.13 stars / hour

Paper

Code

Papers with Code

<https://paperswithcode.com/>

Global Initiatives: Who is doing what and how to get involved



- <https://www.rd-alliance.org/fair>
- <https://www.stm-researchdata.org>
- <https://www.force11.org>
- <https://codata.org/initiatives/working-groups/fair-data-training/>
- <https://gofair.us/>
- <https://www.fairsfair.eu/>



International FAIR Convergence Symposium 2020

[About](#)[Call for Sessions](#)[Call for Posters and Lightning Talks](#)[Sponsorship Opportunities](#)[Sponsors](#)[Registration](#)[Login](#)[Sign Up](#)

Welcome to International FAIR Convergence Symposium 2020

Change of Symposium format and dates!

The International FAIR Convergence Symposium will now take place as a fully virtual event in the week 30 November - 4 December. Please save the dates!

In order to facilitate the new approach, and in response to a number of enquiries, we have also extended the deadline for submissions of session proposals to 30 September and of posters and lightning talks to 31 October.

To ensure an engaging and dynamic event, with clear outcomes, the organisers now envision the Symposium as a series of online sessions, dispersed over one week to accommodate a range of timezones as best as possible.

The International FAIR Convergence Symposium is organised by CODATA and GO FAIR in the week of November 30 - December 4. This event will provide a unique forum for advancing international and cross-domain convergence around FAIR. The two internationally recognized organisations will convene a global community with an interest in combining data across domains for a host of research issues - including major global challenges, such as those relating to the SDGs or the COVID-19 pandemic. Outcomes will directly inform the CODATA Decadal Programme 'Making data work for

Conference Schedule

[+ Session proposals and abstracts](#)

To submit session proposals and conference abstracts please use the button above. Conference logistics and accommodation information will be available on the [Conference Website](#).

	Status	Open	Close
Session proposals	OPEN FOR SUBMISSION	2020-04-30	2020-09-30
Conference poster	OPEN FOR SUBMISSION	2020-04-30	2020-10-31

<https://conference.codata.org/FAIRconvergence2020/>



About Get Involved News Events FAIR Principles

Advancing FAIR in the US

Our Aim is to connect FAIR stakeholders and foster a community where FAIR approaches can be shared, discussed, and advanced collaboratively.

[Read More](#)

<https://gofair.us/>

Asking others, turning to your colleagues



Malte Lau Petersen @LauMalte · Feb 25

Thoughts on hosting Open Access data? I'm looking for the right way to make ~10GB available through an R package. Technically should be easy with osf, gitlab, ec2 or similar, but what is the *right* way to do it?
#rstats #openscience #opendata

4

7

10



Pieter Huybrechts @HuyPieter · Feb 25

You are looking for a **data repository**, have a look here:
nature.com/sdata/policies...

Zenodo is a great choice. Is your data FAIR?



FAIR Principles - GO FAIR

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data. The authors ...
go-fair.org

Malte Lau Petersen @LauMalte · Feb 25

<https://twitter.com/LauMalte/status/1232325843797540870?s=20>

FAIR Data in the Scholarly Communications Lifecycle

Natasha Simons, Chris Erdmann, Daniel Bangert, & Fiona Murphy. (2020, August). FAIR Data in the Scholarly Communications Lifecycle. Presented at the FORCE11 Scholarly Communications Institute (FSCI), Online: Zenodo.

<http://doi.org/10.5281/zenodo.3987052>



Q&A Discussion



Thank you!

