UNIVERSITY OF BERGEN LIBRARY

Sharing research data and software - FAIR & FAIR4RS

Collaborative scientific software development
Summer School, Day 9, 30.06.2022

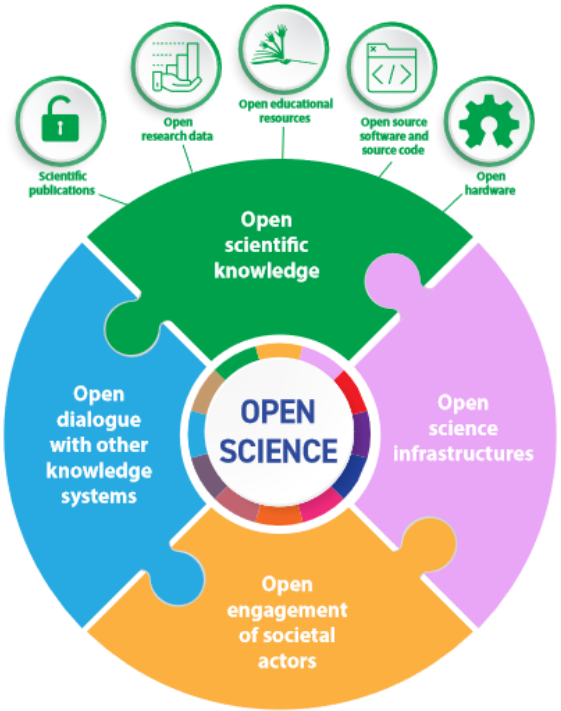
University of Bergen Library
Digital Lab
Jenny Ostrop

UNIVERSITY OF BERGEN

This presentation is available under:

https://doi.org/10.5281/zenodo.6778099

Opening the processes of scientific knowledge creation



Open scientific knowledge refers to open access to scientific publications, research data, metadata, open educational resources, ***software, and source code*** *and hardware […]*

UNESCO Recommendation on Open Science (2021)

https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

Open Science Benefits

- Reproducibility
- Efficiency
- New avenues of research

- Faster response to societal challenges
- Trust in science
- Citizen engagement

- Value creation based on open research results
- Open Innovation

Open Science on the agenda

Norwegian Ministry
of Education and Research

22 August 2017

National goals and guidelines
for open access to research articles

National strategy on
access to and sharing
of research data

Plan S
Making full & immediate
Open Access a reality

European
Commission

English

Home › Research and innovation › Strategy › Goals of research and innovation policy › Open Science

Open Science

An approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Expert groups, publications, news and events.

The Research Council Policy
for Open Science

In effect from 2020

Open Science requirements

- Mandatory: Open access to research articles & rights retention.
- Mandatory: FAIR & open research data by default. Data Management Plan.
- Recommended: Early & open sharing of research, Open peer review, **Output management beyond research data**, Including citizens

Forskningsrådet
The Research Council of Norway

- Full and immediate open access to all articles in accordance with Plan S.
- Data are to be shared in keeping with the FAIR principles. Data Management Plan required.

https://www.uib.no/en/ub/111372/open-access-research-data#open-access-to-research-data-is-a-requirement

Selfish reasons for Open Science Practices

What's in for me?

- Reaching wider audience (incl. general public, researchers in low-income countries)
- **Increased visibility (more references pointing to your work)**
- More citations (publications & data sets)
- **Collaborations, others can build on your work and credit you**
- Early feedback (on pre-registration, preprints, code etc.)

Further reading: Open Science Benefits

POINT OF VIEW

How open science helps researchers succeed

Abstract Open access, open data, open source and other open scholarship practices are growing in popularity and necessity. However, widespread adoption of these practices has not yet been achieved. One reason is that researchers are uncertain about how sharing their work will affect their careers. We review literature demonstrating that open research is associated with increases in citations, media attention, potential collaborators, job opportunities and funding opportunities. These findings are evidence that open research practices bring significant benefits to researchers relative to more traditional closed practices.

DOI: 10.7554/eLife.16800.001

ERIN C MCKIERNAN*, PHILIP E BOURNE, C TITUS BROWN, STUART BUCK, AMYE KENALL, JENNIFER LIN, DAMON MCDOUGALL, BRIAN A NOSEK, KARTHIK RAM, COURTNEY K SODERBERG, JEFFREY R SPIES, KAITLIN THANEY, ANDREW UPDEGROVE, KARA H WOO AND TAL YARKONI

PERSPECTIVE

Open science challenges, benefits and tips in early career and beyond

Christopher Allen[1]*, **David M. A. Mehler**[1,2]*

1 Cardiff University Brain Research Imaging Centre (CUBRIC), Wales, United Kingdom, 2 Department of Psychiatry, University of Muenster, Germany

CAREER FEATURE | 13 May 2019

Data sharing and how it can benefit your scientific career

Open science can lead to greater collaboration, increased confidence in findings and goodwill between researchers.

PERSPECTIVE

On the value of preprints: An early career researcher perspective

Sarvenaz Sarabipour[1]*, **Humberto J. Debat**[2], **Edward Emmott**[3], **Steven J. Burgess**[4], **Benjamin Schwessinger**[5], **Zach Hensel**[6]

1 Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, 2 Center of Agronomic Research, National Institute of Agricultural Technology (IPAVE-CIAP-INTA), Córdoba, Argentina, 3 Department of Bioengineering, Northeastern University, Boston, Massachusetts, United States of America, 4 Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 5 Research School of Biology, The Australian National University, Acton, Australian Capital Territory, Australia, 6 Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

A Beginner's Guide to Conducting Reproducible Research

Jesse M. Alston[1,2], and Jessica A. Rick[1,3]

[1]Program in Ecology, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071 USA

[2]Department of Zoology and Physiology, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071 USA

[3]Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071 USA

McKiernan et al. 2016: 10.7554/eLife.16800
Allen et al. 2019: https://doi.org/10.1371/journal.pbio.3000246
Sarabipour et al. 2019: https://doi.org/10.1371/journal.pbio.3000151
Popkin 2019: https://doi.org/10.1038/d41586-019-01506-x
Alston et al. 2021: https://doi.org/10.1002/bes2.1801

Agenda

- Open Science – a paradigm shift
- Sharing data: Open & FAIR research data
- Sharing software/code: FAIR for research software
- How to make your code citable

The research data life cycle

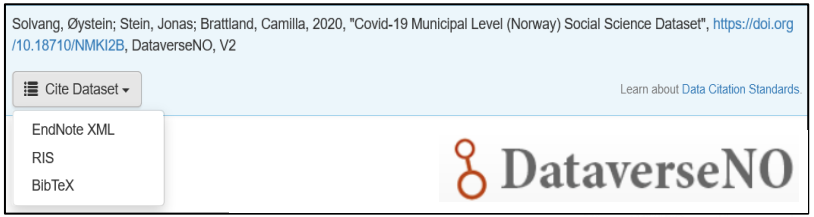
Research projects can:

1. Generate novel data
2. Reuse existing datasets (secondary data)

Data citation

- Principles: Attribution & Access
  - Joint Declaration of Data Citation Principles (JDDCP)
  - Creative Commons: TASL – Title, Author, Source, License

- Many archives contain information how a dataset should be cited



https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/NMKI2B

Data reuse requirements

1. Discovering suitable datasets

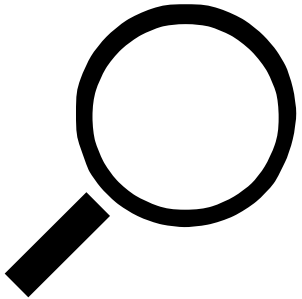
2. Retrieving the data
   - Scale? Manual, automated, or API-retrieval?

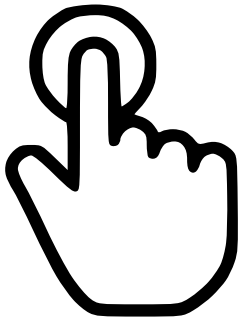
3. Understanding the data
   - Human-readable vs. machine-readable (metadata, data files)

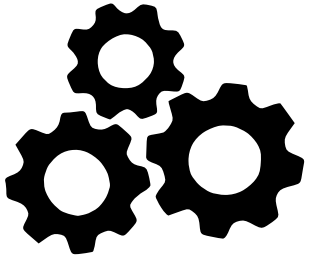
4. Permission to build upon the data

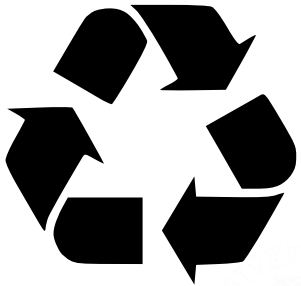
FAIR: prerequisites for reuse

Findeable **A**ccessible **I**nteroperable **R**euseable



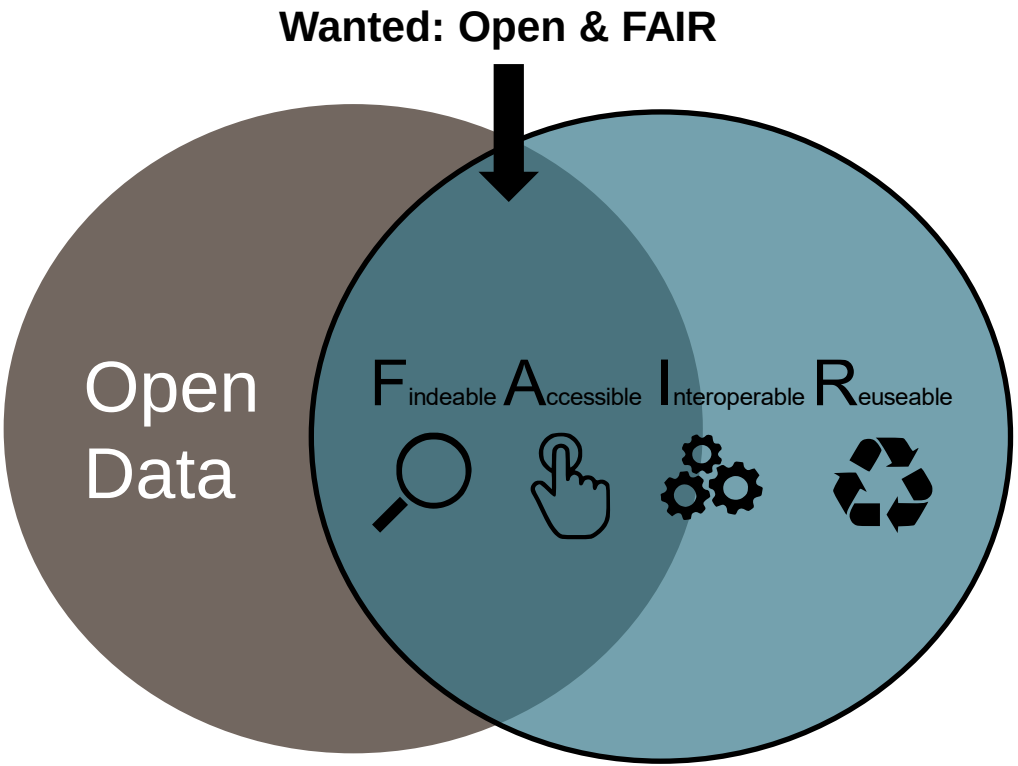






Wilkinson et al., 2016 Sci Data; https://www.go-fair.org/fair-principles/

Open data and FAIR data



“As open as possible – as closed as necessary”

FAIR principles: Findable

Findeable

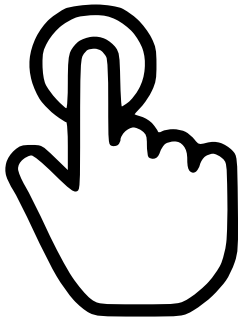
- Permanent, unique identifiers (PID) avoid ambiguity

https://identifiers.org/

- Rich metadata accompany dataset

FAIR principles: Accessible

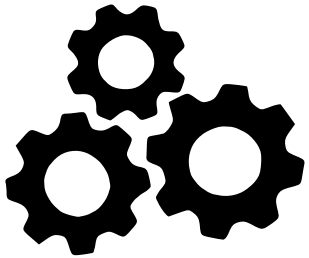
Accessible



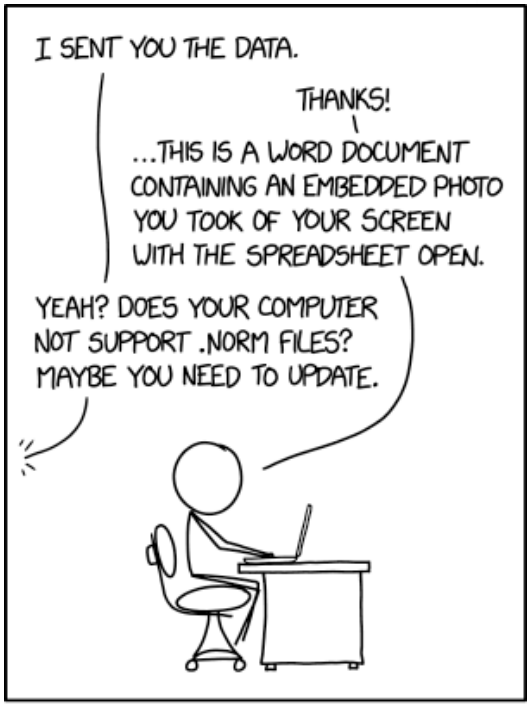
- Available in “approved” repository/archive
- Methods/tools for access is described and available
- Metadata remain available, even if dataset is deleted

FAIR principles: Interoperable

Interoperable

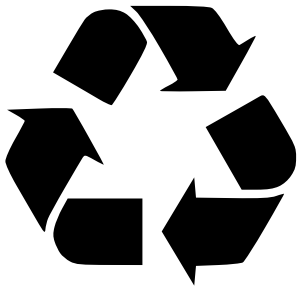


- Standardized metadata
- Standardized (open) file formats
- Controlled vocabularies
- Cross-referencing

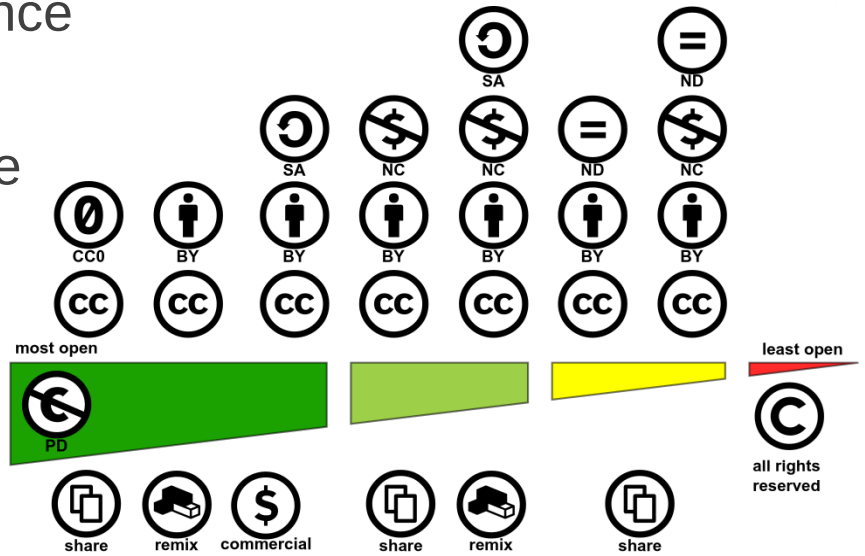


FAIR principles: Reusable

Reuseable



- Detailed provenance
- Licensed for reuse



creative commons

Agenda

- Open Science – a paradigm shift
- Sharing data: Open & FAIR research data
- Sharing software/code: FAIR for research software
- How to make your code citable

Research Software

Software that is used to generate, process or analyse results that you intend to appear in a publication (either in a journal, conference paper, monograph, book or thesis). Research software can be anything from a few lines of code written by yourself, to a professionally developed software package.

UK Research Software Survey 2014

Hettrick et al. 2014, UK Research Software Survey 2014: https://zenodo.org/record/14809

Software is not data

- Software is the result of a creative process that provides a tool for doing something, for example with data.
- Software is executable, while data is not.
- All software applications that are not written completely from scratch are of a composite nature that easily leads to complex dependencies.
- Lifetime of software is generally shorter than that of data, as versioning is applied more frequently and regularly leads to changes in behaviour and/or interfaces.

Lamprecht et al. 2019, Towards FAIR principles for research software: https://doi.org/10.3233/DS-190026

FAIR for research software

Data Science 3 (2020) 37–59
DOI 10.3233/DS-190026
IOS Press

Towards FAIR principles for research software

Anna-Lena Lamprecht [a,*], Leyla Garcia [b], Mateusz Kuzak [c,d], Carlos Martinez [e], Ricardo Arcila [f], Eva Martin Del Pico [g], Victoria Dominguez Del Angel [h], Stephanie van de Sandt [i], Jon Ison [j], Paula Andrea Martinez [k], Peter McQuilton [l], Alfonso Valencia [m,n], Jennifer Harrow [o], Fotis Psomopoulos [p], Josep Ll. Gelpi [q,r], Neil Chue Hong [s,t], Carole Goble [u] and Salvador Capella-Gutierrez [v,**]

Lamprecht et al. 2019, Towards FAIR principles for research software: https://doi.org/10.3233/DS-190026

FAIR4RS principles (2021)

FAIR Principles for Research Software (FAIR4RS Principles)

FAIR Guiding Principles (2016)	Towards FAIR Principles for research software (2020)	Taking a fresh look at FAIR for research software (2021)	FAIR4RS Principles (2021)
F. 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.	The main concern of findability for research software is to ensure software can be identified unambiguously when looking for it using common search strategies.	The first step in (re)using software is to find it. Metadata and software should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of software, so this is an essential component of the FAIRification process.	The software, and its associated metadata, should be easy to find for both humans and machines.

Chue Hong et al. 2021, FAIR principles for Research Software: https://doi.org/10.15497/RDA00065

FAIR4RS principles (2021)

FAIR Guiding Principles (2016)	Towards FAIR Principles for research software (2020)	Taking a fresh look at FAIR for research software (2021)	FAIR4RS Principles (2021)
F1. (Meta)data are assigned a globally unique and persistent identifier	F1. Software and its associated metadata have a global, unique and persistent identifier for each released version.	F1. Software is assigned a globally unique and persistent identifier	F1. Software is assigned a globally unique and persistent identifier.
			F1.1. Different components of the software must be assigned distinct identifiers representing different levels of granularity.
			F1.2. Different versions of the same software must be assigned distinct identifiers.

Chue Hong et al. 2021, FAIR principles for Research Software: https://doi.org/10.15497/RDA00065

FAIR4RS version 1.0 (2022)

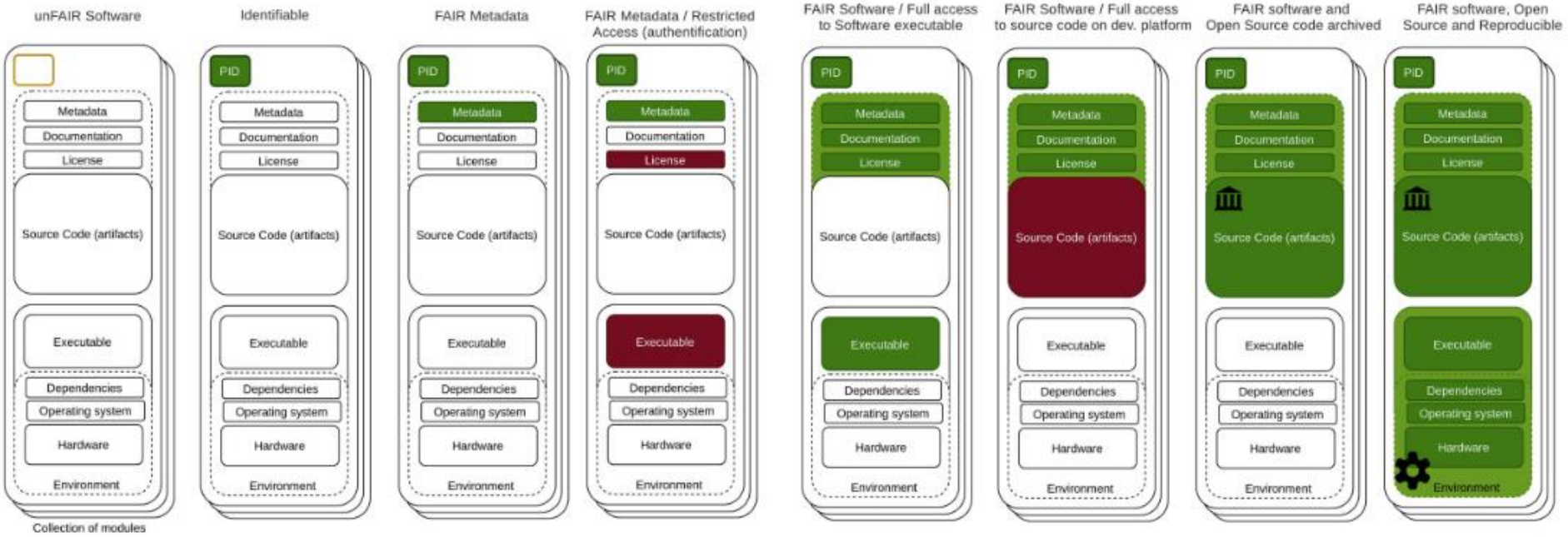
Date	Version Number	Description	Editor(s)
15/3/2022	1.0	First release of principles○	Neil Chue Hong
9/6/2021	0.3	Draft for formal RDA community review	Neil Chue Hong
7/6/2021	0.2.1	Amended abstract and text of F1, F1.1, F1.2, F4 and R1 for review by drafting group	Neil Chue Hong
1/6/2021	0.2	Second draft for review by FAIR4RS Steering Committee	Neil Chue Hong
17/5/2021	0.1	First draft for review by FAIR4RS WG	Neil Chue Hong, Michelle Barker

○: *The pre-1.0 drafts of the FAIR4RS Principles included sections describing the drafting process - these are now published separately.*

Chue Hong et al. 2022, FAIR principles for Research Software version 1.0: https://doi.org/10.15497/RDA00068

Software as FAIR research objects

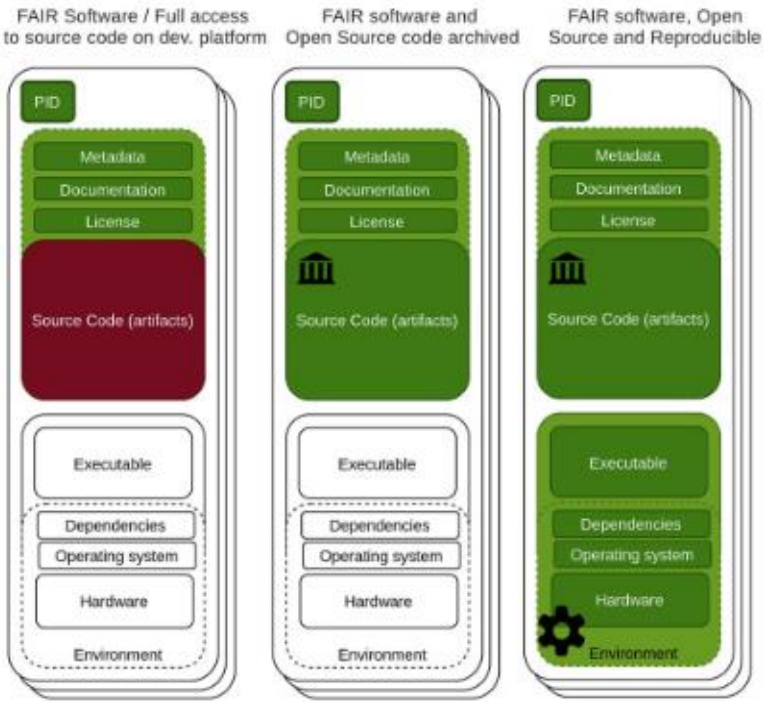
FAIRness



Modified from Katz et al. 2021, Taking a fresh look at FAIR research software: https://doi.org/10.1016/j.patter.2021.100222 (CC BY)

Required steps

e.g. GitHub zenodo



- Public repository with version control
- Software license: choosealicense.com
- Archive code & enable citation
  next on the agenda
- Recording of dependencies and environments
- Software quality measures

Modified from Katz et al. 2021 (CC BY); Five recommendations for FAIR software: https://fair-software.eu/

Agenda

- Open Science – a paradigm shift
- Sharing data: Open & FAIR research data
- Sharing software/code: FAIR for research software
- How to make your code citable

Make your code citable

- Your code will be archived and get a Digital Object Identifier (DOI)

- ➢ Makes it easy to use your code and give you **credit**
- ➢ Improves findability of your code
- ➢ Encourages others to build on your work
- ➢ Allows reproducibility and transparency

Code Refinery Software Citation: https://coderefinery.github.io/social-coding/software_citation/

Archives for research software

zenodo

Category **Archives**

Zenodo

Global multidisciplinary repository

Type
Public

Year of creation
2013

Legal status
Service in organisation

Geographical scope
International

Content scope
Global

Software projects handled
44,086 (more than 101K versions)

Source code archival
Own storage

Supported identifiers

Estimated resources
FTE:4

Software infrastructure
Licence: open source (MIT)
Data access: open API

Policy support *(optional)*
International endorsement

Website
zenodo.org

Software Heritage

Category **Archives**

Software Heritage

Universal software archive

Type
Private not for profit

Year of creation
2015

Legal status
Hosted organisation

Geographical scope
International

Content scope
Global

Software projects handled
140Million

Source code archival
Own storage plus mirror network

Supported identifiers

Estimated resources
Current: 14 FTE, 1,600,000€/year
Target: 50 FTE, 10,000,000€/year

Software infrastructure
Licence: open source
Data access: open API

Policy support *(optional)*
National Plan for Open Science (France)
Agreement with UNESCO

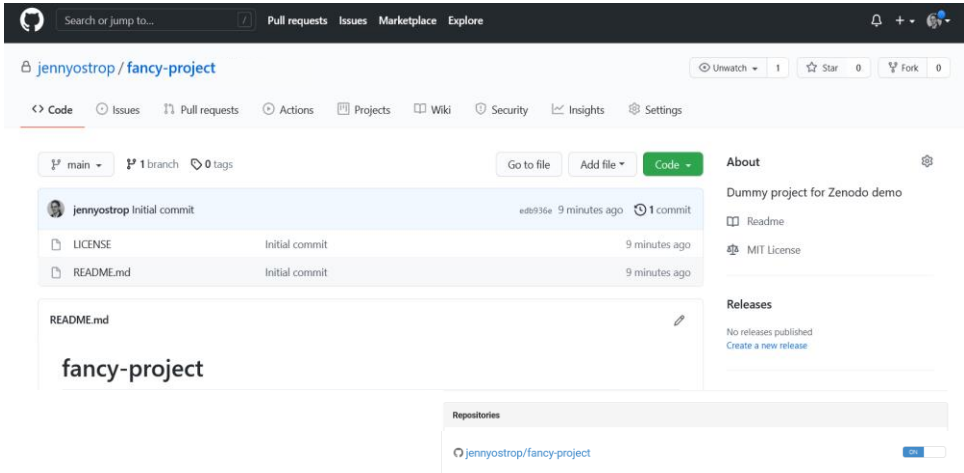
Website
softwareheritage.org

Demo Video https://www.youtube.com/watch?v=8nlSvYh7VpI

Reference Publications
(Di Cosmo, 2020) doi:10.1007/978-3-030-52200-1_36
(Abramatic et al., 2018b) doi:10.1145/3183558

GitHub release on Zenodo step-by-step

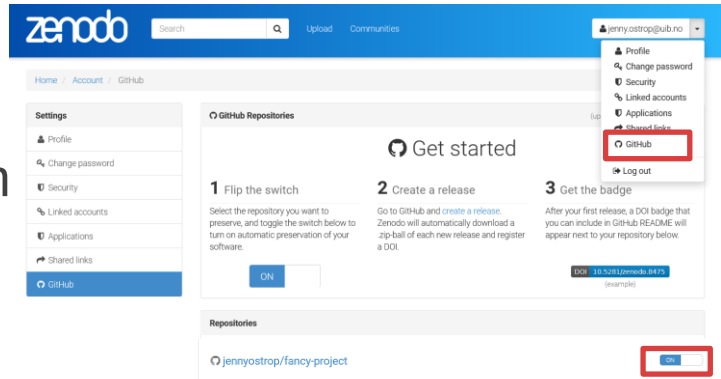
1. Create GitHub project (must be public)



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

GitHub release on Zenodo step-by-step

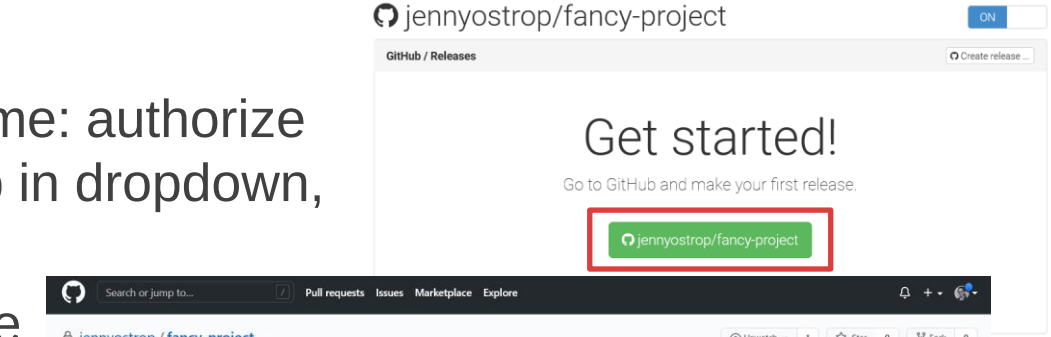
1. Create GitHub project
2. Login to Zenodo (first time: authorize GitHub), choose GitHub in dropdown, select project to release & toggle switch

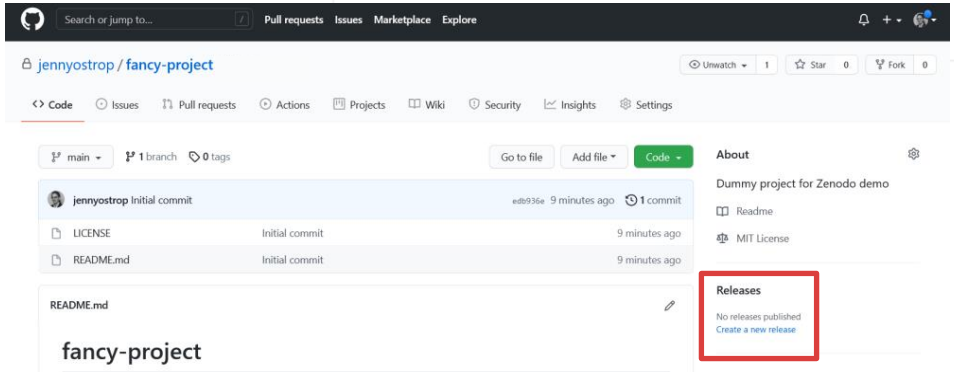


For testing:
https://sandbox.zenodo.org

GitHub release on Zenodo step-by-step

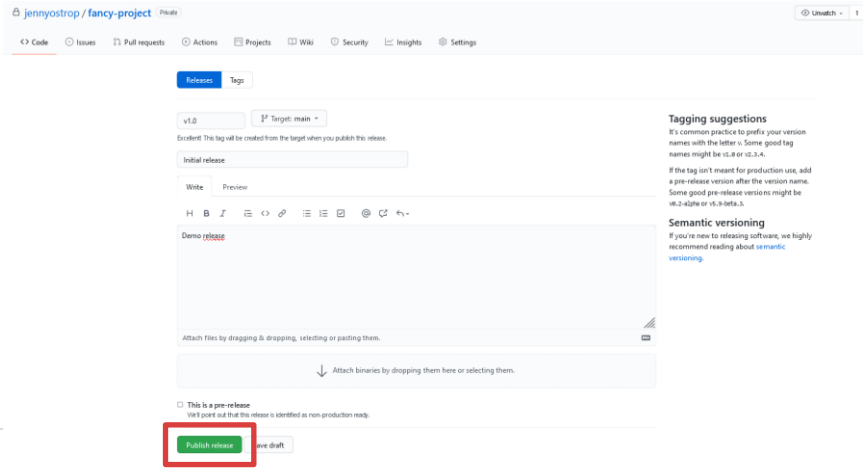
1. Create GitHub project
2. Login to Zenodo (first time: authorize GitHub), choose GitHub in dropdown, select project to release
3. In Github, create release





GitHub release on Zenodo step-by-step

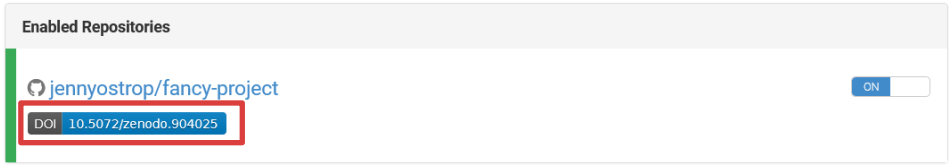
1. Create GitHub project
2. Login to Zenodo (first time: authorize GitHub), choose GitHub in dropdown, select project to release
3. In Github, create release
4. In Github, publish release



GitHub release on Zenodo step-by-step

1. Create GitHub project
2. Login to Zenodo (first time: authorize GitHub), choose GitHub in dropdown, select project to release
3. In Github, create release
4. In Github, publish release

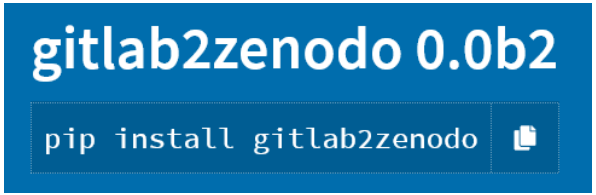
   >> release will get doi



GitLab: gitlab2zenodo

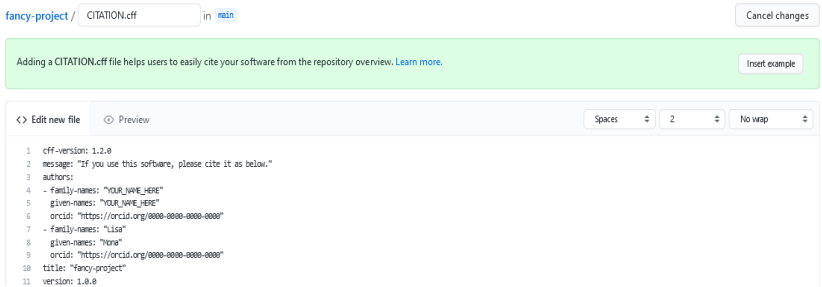
- UiB GitLab Community Edition: https://git.app.uib.no/users/sign_in

- gitlab2zenodo (beta) sends GitLab snapshots to Zenodo automatically: https://pypi.org/project/gitlab2zenodo/

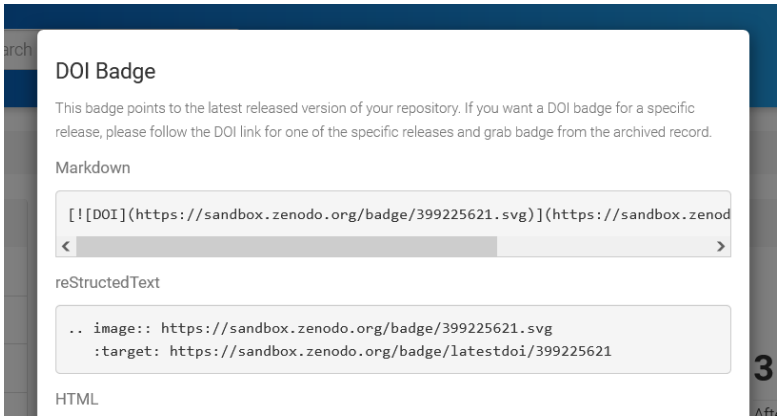


Citation information visibility

1. Add & fill out GitHub CITATION.cff template
   >> import in Zenodo (add pre-release!)
   >> supported by Zotero



2. Add doi-button to GitHub README (copy & paste from Zenodo)



https://docs.github.com/en/github/creating-cloning-and-archiving-repositories/creating-a-repository-on-github/about-citation-files

CITATION.cff

What is a `CITATION.cff` file?

`CITATION.cff` files are plain text files with human- and machine-readable citation information for software (and datasets). Code developers can include them in their repositories to let others know how to correctly cite their software.

This is an example of a simple `CITATION.cff` file:

```yaml
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/0000-0003-4925-7248
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2021-08-11
```


https://citation-file-format.github.io/

If you are citing software

- Minimal requirement: Creator(s), Title, Publication venue, Date, Identifier
- Recommended: Version, Type

Developer, A. A., Developer, B. B., & Developer, C. C. (yyyy)[1]. *Title of the software: Subtitle* (Version #.#)[2] [Computer software][3]. Publisher[4], https://URL[5]

Developer, A. A., Developer, B. B., & Developer, C. C. (yyyy). *Title of the software: Subtitle* [Computer software]. Archive Name. Retrieved Month dd, yyyy, from https://URL

Katz et al. 2021, Recognizing the value of software: a software citation guide: https://doi.org/10.12688/f1000research.26932.2

CodeMeta: metadata for software

What metadata you want from software is determined by your use case. […]

Different software repositories, software languages and scientific domains denote this information in different ways, which makes it difficult or impossible for tools to work across these different sources without losing valuable information along the way.

https://codemeta.github.io/; https://github.com/codemeta/codemeta; https://codemeta.github.io/codemeta-generator

CodeMeta: metadata for software

CodeMeta

gitter join chat | build passing | doi: 10.5063/SCHEMA/CODEMETA-2.0

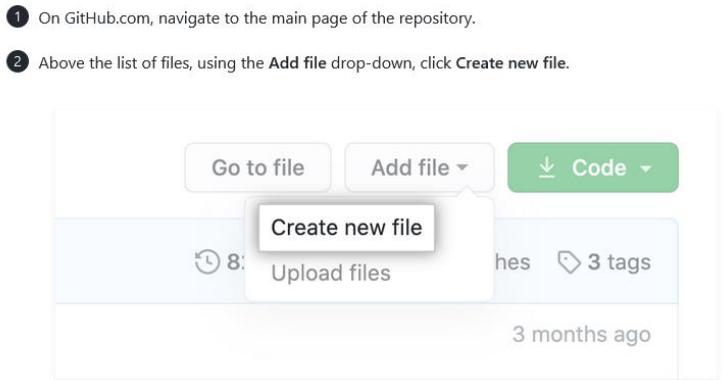
CodeMeta contributors are creating a minimal metadata schema for science software and code, in JSON and XML. The goal of CodeMeta is to create a concept vocabulary that can be used to standardize the exchange of software metadata across repositories and organizations. CodeMeta started by comparing the software metadata used across multiple repositories, which resulted in the CodeMeta Metadata Crosswalk. That crosswalk was then used to generate a set of software metadata concepts, which were arranged into a JSON-LD context for serialization.

https://codemeta.github.io/; https://github.com/codemeta/codemeta; https://codemeta.github.io/codemeta-generator

Code of conduct
Contributor guidelines

- CODE_OF_CONDUCT.md: define community standards, signal a welcoming and inclusive project, and outline procedures for handling abuse

- CONTRIBUTING.md: create guidelines to communicate how people should contribute to your project



Further resources – sharing data

openscience.no
incl. event calendar

CESSDA Data Management
Expert Guide

PhD on Track -
Open Science

ELIXIR RDM Kit

Further resources – sharing code

- RDA - FAIR for Research Software (FAIR4RS) WG
- FORCE 11 - FAIR for Research Software (FAIR4RS) WG

- Hettrick et al. 2014, UK Research Software Survey 2014
- FAIR4RS subgroup 3 – Research software definition

- Lamprecht et al. 2019, Towards FAIR principles for research software
- Katz et al. 2021, Taking a fresh look at FAIR research software
- Chue Hong et al. 2021, FAIR principles for Research Software
- Chue Hong et al. 2022, FAIR principles for Research Software (FAIR4RS Principles) (1.0)
- 4OSS recommendations
- Five recommendations for FAIR software

- Katz et al. 2021, Recognizing the value of software: a software citation guide
- Chue Hong et al. 2019, Software Citation Checklist for Developers

- EOSC Executive Board Working Group - Scholarly infrastructures for research software
- Chapter 9 RDA COVID-19 group recommendations for research software

- Software Sustainability Institute – FAIR Software
  Library Carpentry: FAIR Data and Software – Software
- LibraryCarpentry - Research Software
- SoftwareCarpentry – Open Science
- CodeRefinery - Reproducible Research: Sharing code and data

research-data@uib.no

Jenny.Ostrop@uib.no

@UiB_UB @DLabUiB @jennyostrop