

# FAIR Principles & How-to

Cunliang Geng

NPLinker eScience Workshop

2023-03-22 WUR

netherlands  
eScience center



# Netherlands eScience Center

We are the national centre for innovative software solutions in academic research.

It was founded in 2012 as an independent foundation by NWO and SURF.



# VISION

a robust research community, in which all investigators in all domains are able to exploit advanced digital technologies to answer curiosity-driven questions, keeping the Netherlands at the forefront of cutting-edge international research.

## MISSION

empowering researchers across all disciplines through innovative research software.

**AMBITION 1**  
collaboratively designing software for research

**AMBITION 2**  
building digital expertise

## STRATEGIC PRIORITIES

- ✓ institutional alignment
- ✓ software sustainability

open calls for eScience domain research

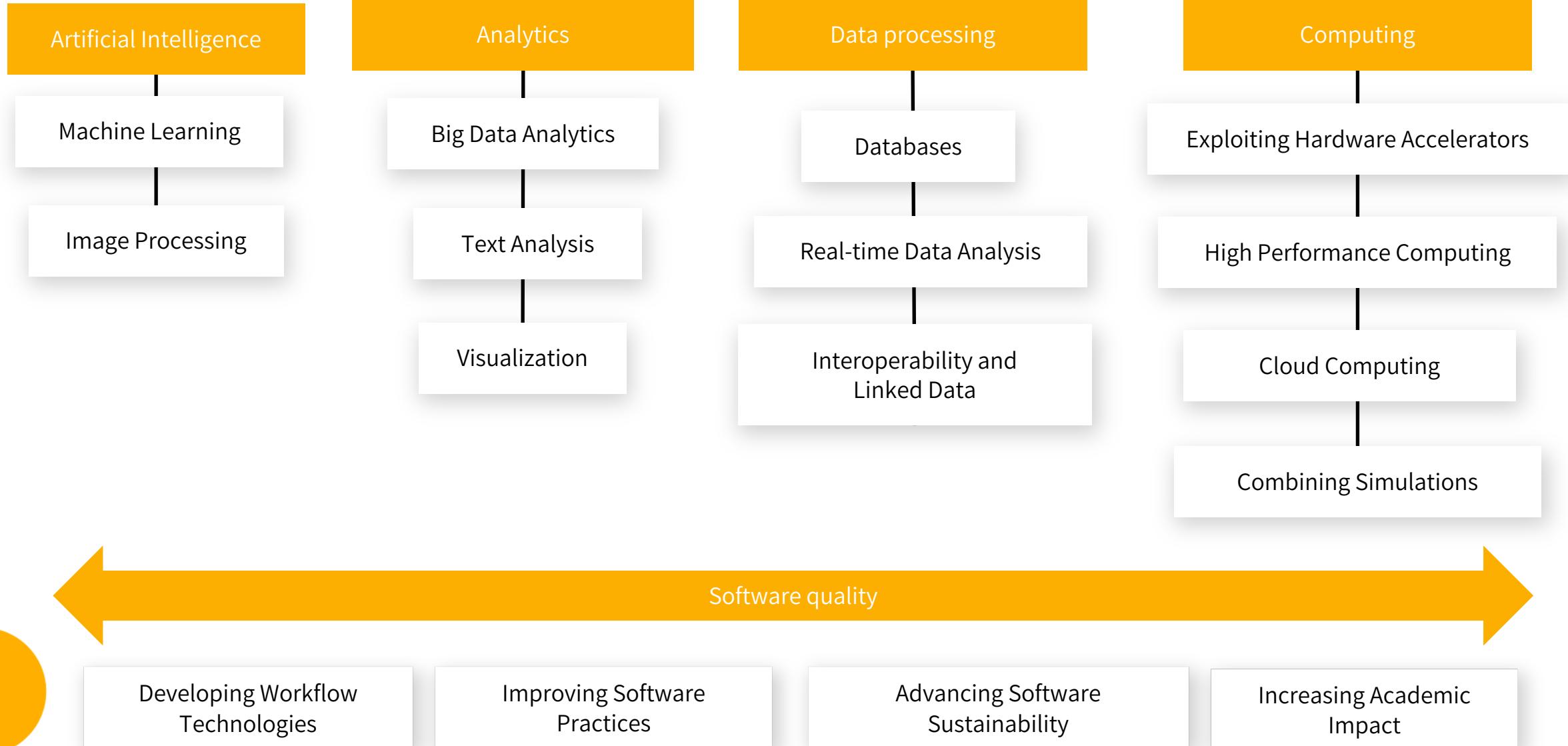
collaborations in advanced eScience technologies

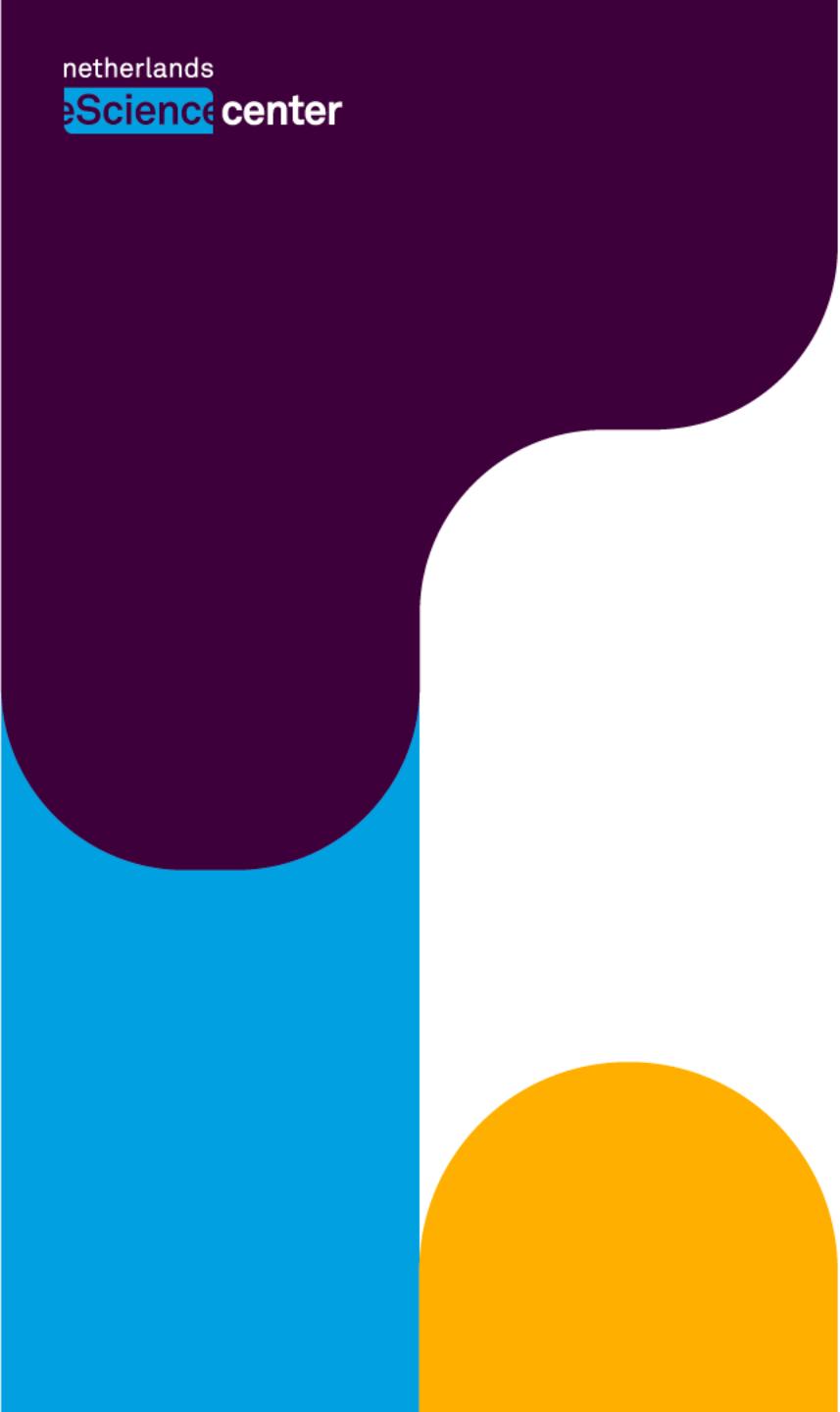
workshops & training

community building & networking



# Our Technological expertise





F A I R

F indable



A ccessible



I nteroperable



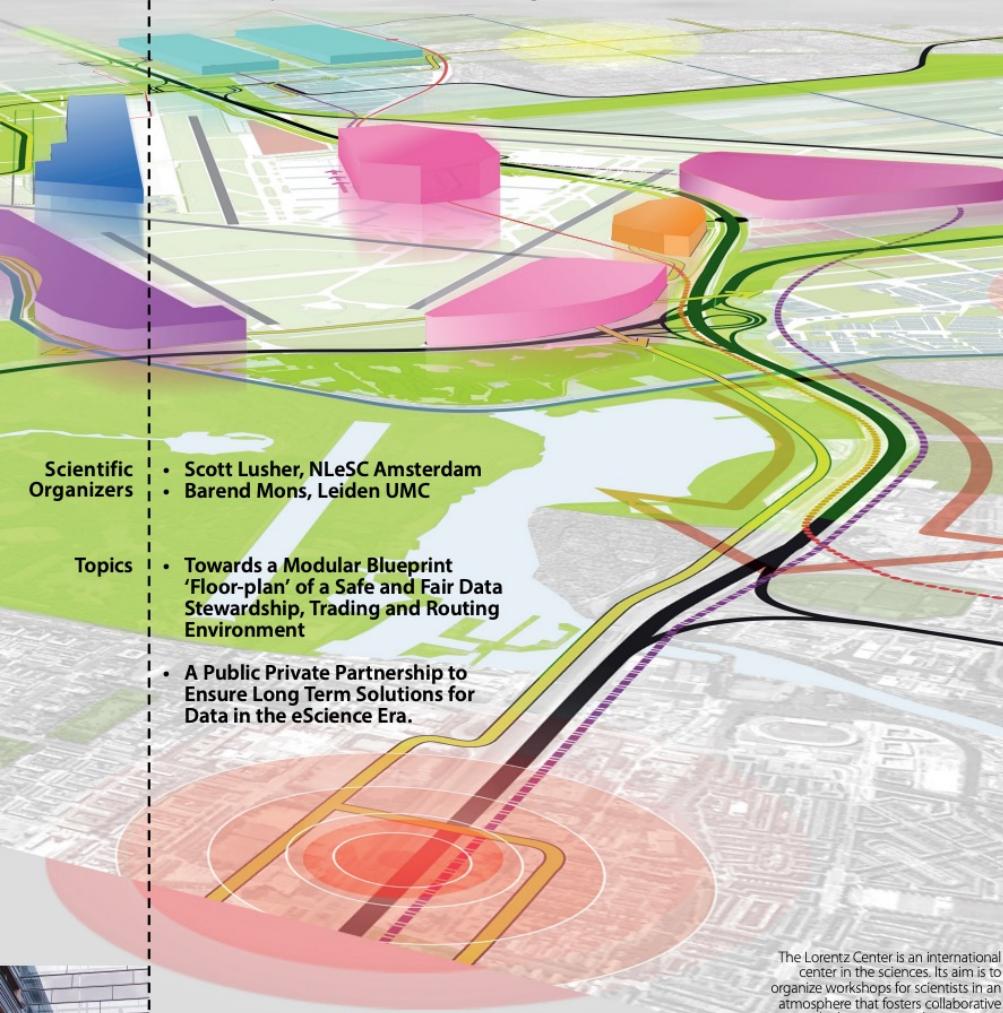
R eusable





## Jointly Designing a Data FAIRPORT

Workshop: 13 - 16 January 2014, Leiden, the Netherlands



The image shows a detailed aerial map of Schiphol Airport, featuring various runways, terminals, and surrounding urban areas. Overlaid on the map are several large, semi-transparent colored shapes: pink, orange, and red. A vertical dashed line on the left side of the map separates the main map area from a sidebar.

**Scientific Organizers**

- Scott Lusher, NLeSC Amsterdam
- Barend Mons, Leiden UMC

**Topics**

- Towards a Modular Blueprint 'Floor-plan' of a Safe and Fair Data Stewardship, Trading and Routing Environment
- A Public Private Partnership to Ensure Long Term Solutions for Data in the eScience Era.

The Lorentz Center is an international center in the sciences. Its aim is to organize workshops for scientists in an atmosphere that fosters collaborative work, discussions and interactions. For registration see: [www.lorentzcenter.nl](http://www.lorentzcenter.nl).

Image: Structure Plan Schiphol Airport by KCAP Architects&Planners.  
Poster design: SuperNova Studios - NL.

**Logos:** Universiteit Leiden, FOM, STW, elixir, DTL, eScience center, NWO, Lorentz center

[www.lorentzcenter.nl](http://www.lorentzcenter.nl)

Goal:

***Increase transparency, reproducibility,  
and reusability of research***



Open Access | Published: 15 March 2016

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons  + Show authors

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

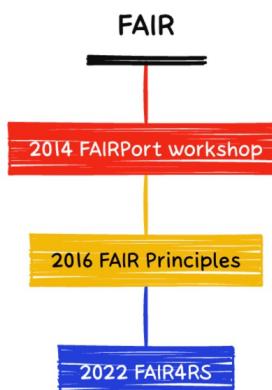
539k Accesses | 5312 Citations | 2071 Altmetric | [Metrics](#)

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

**Scientific data** refers to all digital research objects, including data (in the conventional sense), algorithms, tools, workflows...

However, **software** has some specific characteristics:

- Executability
- Composite nature
- Continuous evolution and versioning



Article | Open Access | Published: 14 October 2022

## Introducing the FAIR Principles for research software

Michelle Barker [✉](#), Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez & Tom Honeyman

*Scientific Data* 9, Article number: 622 (2022) | [Cite this article](#)

9524 Accesses | 2 Citations | 244 Altmetric | [Metrics](#)

<https://doi.org/10.1038/s41597-022-01710-x>

**Research software** is defined as

*source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose.*



## Take a closer look at

F indable



A ccessible



I nteroperable



R eusable



FAIR Guiding Principles (2016)	FAIR4RS Principles (2022)
<b>F. Findable</b>	
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.	Software, and its associated metadata, is easy for both humans and machines to find.
F1. (Meta)data are assigned a globally unique and persistent identifier	F1. Software is assigned a globally unique and persistent identifier.  F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.  F1.2. Different versions of the software are assigned distinct identifiers.
F2. Data are described with rich metadata (defined by R1 below)	F2. Software is described with rich metadata.
F3. Metadata clearly and explicitly include the identifier of the data they describe	F3. Metadata clearly and explicitly include the identifier of the software they describe.
F4. (Meta)data are registered or indexed in a searchable resource	F4. Metadata are FAIR, searchable and indexable.



FAIR Guiding Principles (2016)	FAIR4RS Principles (2022)
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.	Software, and its metadata, is retrievable via standardized protocols.
A. Accessible	
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	A1. Software is retrievable by its identifier using a standardized communications protocol.
A1.1. The protocol is open, free, and universally implementable	A1.1. The protocol is open, free, and universally implementable.
A1.2. The protocol allows for an authentication and authorization procedure, where necessary	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	A2. Metadata are accessible, even when the software is no longer available.

**FAIR ≠ Open**

***“As open as possible, as closed as necessary”***



FAIR Guiding Principles (2016)	FAIR4RS Principles (2022)
<b>I. Interoperable</b>	
The data usually needs to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.	Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
I2. (Meta)data use vocabularies that follow FAIR principles	Now split between F4 and I1.
I3. (Meta)data include qualified references to other (meta)data	I2. Software includes qualified references to other objects.



FAIR Guiding Principles (2016)	FAIR4RS Principles (2022)
R. Reusable	
The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.	Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	R1. Software is described with a plurality of accurate and relevant attributes.
R1.1. (Meta)data are released with a clear and accessible data usage license	R1.1. Software is given a clear and accessible license.
R1.2. (Meta)data are associated with detailed provenance	R1.2. Software is associated with detailed provenance.
R1.3. (Meta)data meet domain-relevant community standards	R3. Software meets domain-relevant community standards.
	R2. Software includes qualified references to other software.



## How to make software

F indable



A ccessible



I nteroperable



R eusable

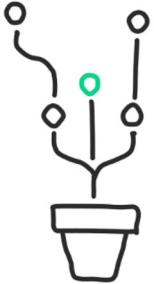


# FAIR-software.eu



# FAIR-software.eu

#1  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



## Repos

- **GitHub.com**
- GitLab.com
- Bitbucket.org

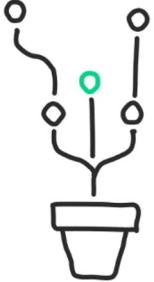
## Version control

- **Git**
- Mercurial
- Bazaar
- CVS



# FAIR-software.eu

**#1**  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



**#2**  
**ADD A LICENSE**



Permissive licenses

- **Apache-2.0**
- MIT licenses

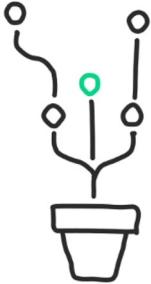
Find more on

<https://choosealicense.com>



# FAIR-software.eu

**#1**  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



**#2**  
**ADD A LICENSE**



**#3**  
**REGISTER YOUR  
CODE IN A  
COMMUNITY  
REGISTRY**

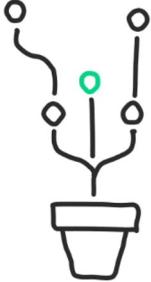


Register Python packages in PyPI (<https://pypi.org/>).

Find more registries on  
<https://github.com/NLeSC/awesome-research-software-registries>

# FAIR-software.eu

#1  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



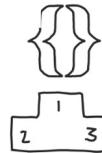
#2  
**ADD A LICENSE**



#3  
**REGISTER YOUR  
CODE IN A  
COMMUNITY  
REGISTRY**



#4  
**ENABLE CITATION  
OF THE  
SOFTWARE**



Citation File Format (CITATION.cff)  
<https://citation-file-format.github.io/>

Easy to create with CFFInit website  
<https://citation-file-format.github.io/cff-initializer-javascript>

```
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/1234-5678-9101-1121
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2021-08-11
```

Search or jump to... Pull requests Issues Codespaces Marketplace Explore

citation-file-format / citation-file-format Public Watch 16 Fork 102 Star 358

Code Issues 82 Pull requests 9 Actions Security Insights

main 14 branches 3 tags Go to file Add file Code

jspaaks Merge pull request #429 from ol... 393ac52 3 weeks ago 1,157 commits

.github	Create cffconvert.yml	2 years ago
examples	Rename schema_poc to validator	2 years ago
tests	Make error message more useful	2 years ago
.bumpversion.cfg	Bump to version 1.2.0	2 years ago
.gitignore	Ignore VSCode metadata	2 years ago
.mlc-config.json	refs #243; local links are not checked reliably, s...	2 years ago
.zenodo.json	Update Zenodo metadata with related identifiers	2 years ago
CHANGELOG.md	Update changelog	3 months ago
CITATION.cff	Fix name spelling	3 weeks ago
CODE_OF_CONDUCT....	Move CoC to root	
CONTRIBUTING.md	Inform users which branch to fork from for whi	
LICENSE	Fix #87: Change license file format	
README.md	Merge pull request #423 from citation-file-for	
pytest.ini	added validator for snippets in schema-guide.m	
requirements.txt	added bumpversion requirement	
schema-guide.md	Remove my real ORCID from examples and ter	
schema.json	Remove unneeded hash from schema	2 years ago

About

The Citation File Format lets you provide citation metadata for software or datasets in plaintext files that are easy to read by both humans and machines.

citation-file-format.github.io

attribution format citation credit  
research-software-engineering  
citation-files software-sustainability  
wssspe

Readme CC-BY-4.0 license Code of conduct Cite this repository

Cite this repository

If you use this software in your work, please cite it using the following metadata. [Learn more](#)

APA BibTeX

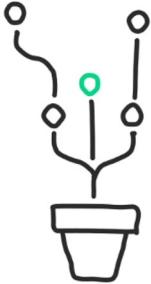
Druskat, S., Spaaks, J. H., Chue Hong, N.,

View citation file



# FAIR-software.eu

**#1**  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



**#2**  
**ADD A LICENSE**



**#3**  
**REGISTER YOUR  
CODE IN A  
COMMUNITY  
REGISTRY**



**#4**  
**ENABLE CITATION  
OF THE  
SOFTWARE**



**#5**  
**USE SOFTWARE  
QUALITY  
CHECKLIST**



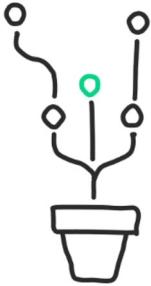
Checklists help you write good quality software

- OpenSSF Best Practices Badge Program  
<https://bestpractices.coreinfrastructure.org/>



# FAIR-software.eu

**#1**  
**USE A PUBLICLY  
ACCESSIBLE  
REPOSITORY  
WITH VERSION  
CONTROL**



**#2**  
**ADD A LICENSE**



**#3**  
**REGISTER YOUR  
CODE IN A  
COMMUNITY  
REGISTRY**



**#4**  
**ENABLE CITATION  
OF THE  
SOFTWARE**



**#5**  
**USE SOFTWARE  
QUALITY  
CHECKLIST**



The image shows a dark-themed mobile application screen. At the top, there is a navigation bar with the word "FAIR" and numbers 1 through 5. To the right of the numbers is a button labeled "ENDORSE". Below the navigation bar, the text "FIVE RECOMMENDATIONS FOR FAIR SOFTWARE" is displayed in large white capital letters. Underneath this text is a green button labeled "LET'S GO! →". To the right of the text, there is a hand icon with dashed green lines emanating from it, suggesting action or endorsement. At the bottom of the screen, there are two buttons: "WHAT'S FAIR" and "ABOUT US".



# Check how FAIR is your software

Python package **howfairis**

Check compliance with *fair-software.eu* recommendations for the repository at URL

```
> pip install howfairis
> howfairis https://github.com/<owner>/<repo>
```

```
Checking compliance with fair-software.eu...
url: https://github.com/fair-software/badge-test
(1/5) repository
  ✓ has_open_repository
(2/5) license
  ✓ has_license
(3/5) registry
  ✗ has_ascl_badge
  ✗ has_bintray_badge
  ✗ has_conda_badge
  ✗ has_cran_badge
  ✗ has_crates_badge
  ✗ has_maven_badge
  ✗ has_npm_badge
  ✓ has_pypi_badge
  ✗ has_rsd_badge
  ✗ is_on_github_marketplace
(4/5) citation
  ✗ has_citation_file
  ✗ has_citationcff_file
  ✗ has_codemeta_file
  ✓ has_zendodo_badge
  ✗ has_zendodo_metadata_file
(5/5) checklist
  ✓ has_core_infrastructures_badge
```



# Check how FAIR is your software

## howfairis-github-action

```
name: fair-software

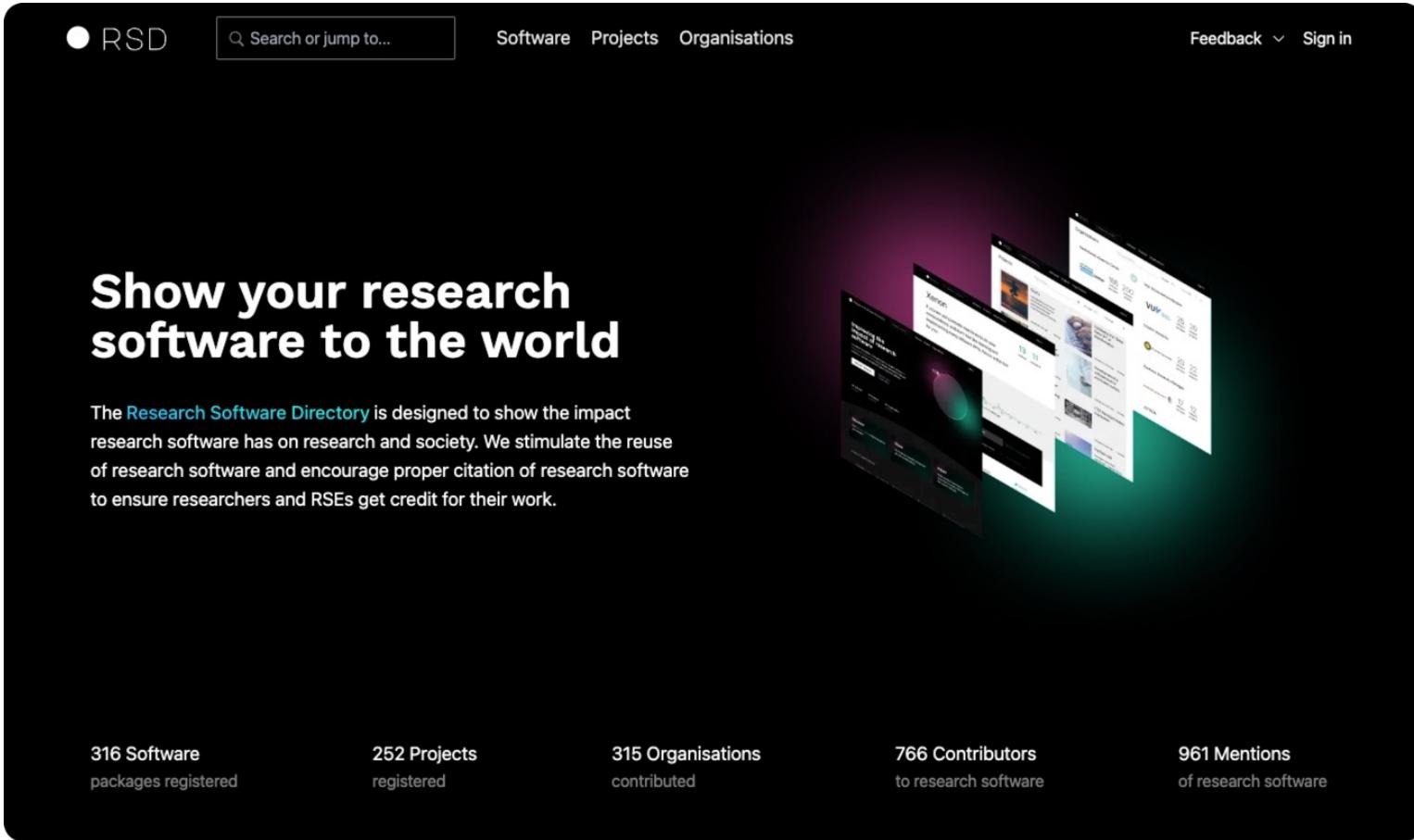
on: push

jobs:
  verify:
    name: "fair-software"
    runs-on: ubuntu-latest
    steps:
      - uses: fair-software/howfairis-github-action@0.2.1
        name: Measure compliance with fair-software.eu recommendations
        env:
          PYCHARM_HOSTED: "Trick colorama into displaying colored output"
        with:
          MY_REPO_URL: "https://github.com/{{ github.repository }}"
```



# Research-Software-Directory.org

To promote the impact, re-use and citation of research software



The screenshot shows the homepage of the Research-Software-Directory.org website. At the top, there is a navigation bar with a logo, a search bar, and links for Software, Projects, Organisations, Feedback, and Sign in. The main content area features a large call-to-action button with the text "Show your research software to the world". Below this, a paragraph explains the directory's purpose: "The Research Software Directory is designed to show the impact research software has on research and society. We stimulate the reuse of research software and encourage proper citation of research software to ensure researchers and RSEs get credit for their work." To the right of this text, there is a graphic showing three floating mobile device screens displaying different parts of the directory's interface. At the bottom of the page, there are five summary statistics: 316 Software packages registered, 252 Projects registered, 315 Organisations contributed, 766 Contributors to research software, and 961 Mentions of research software.

● RSD

Search or jump to...

Software Projects Organisations

Feedback Sign in

## Show your research software to the world

The [Research Software Directory](#) is designed to show the impact research software has on research and society. We stimulate the reuse of research software and encourage proper citation of research software to ensure researchers and RSEs get credit for their work.

316 Software packages registered

252 Projects registered

315 Organisations contributed

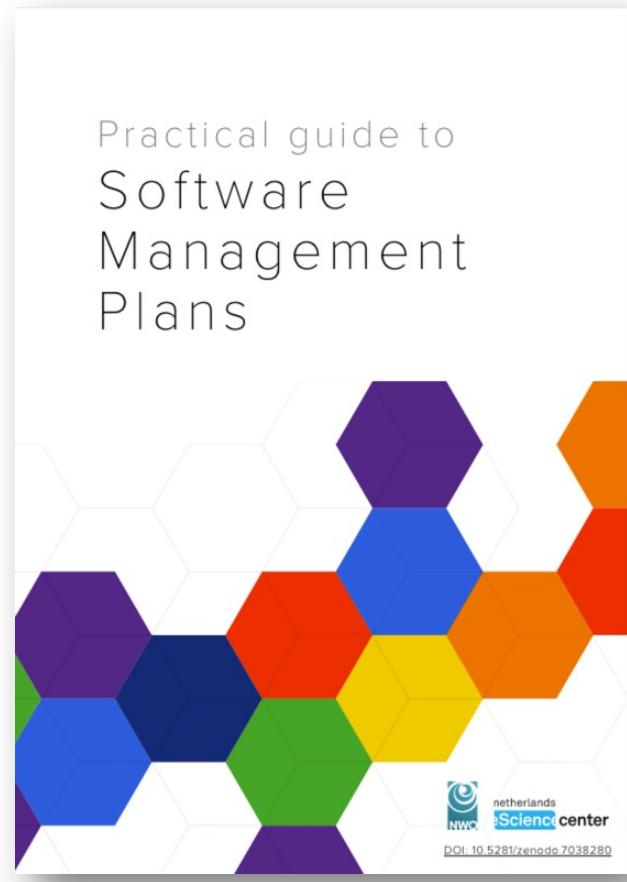
766 Contributors to research software

961 Mentions of research software



# Software Management Plans

SMP help you (project manager) manage software before, during and after the project, leading to FAIR and sustainable software



***SMP requirements grouped by their focus***

## How to make data

F indable



A ccessible



I nteroperable



R eusable



# Data Management Plans

<https://dmponline.dcc.ac.uk/>

**DMPONLINE**

Home    Public DMPs    Funder requirements    Help    Language ▾

## Plan to make data work for you

Data Management Plans that meet institutional funder requirements.

The diagram illustrates the DMPonline platform as a central hub connecting various data management components. At the center is a computer monitor displaying a dashboard. Surrounding the monitor are several interconnected icons representing different data management activities and storage locations:

- A person icon connected to a database icon.
- A person icon connected to a search bar icon.
- A person icon connected to a document icon.
- A person icon connected to a cloud storage icon.
- A person icon connected to a laptop icon.
- A person icon connected to a smartphone icon.
- A person icon connected to a server tower icon.
- A person icon connected to an envelope icon.
- A person icon connected to a gear icon.
- A person icon connected to a folder icon.
- A person icon connected to a bottle icon.

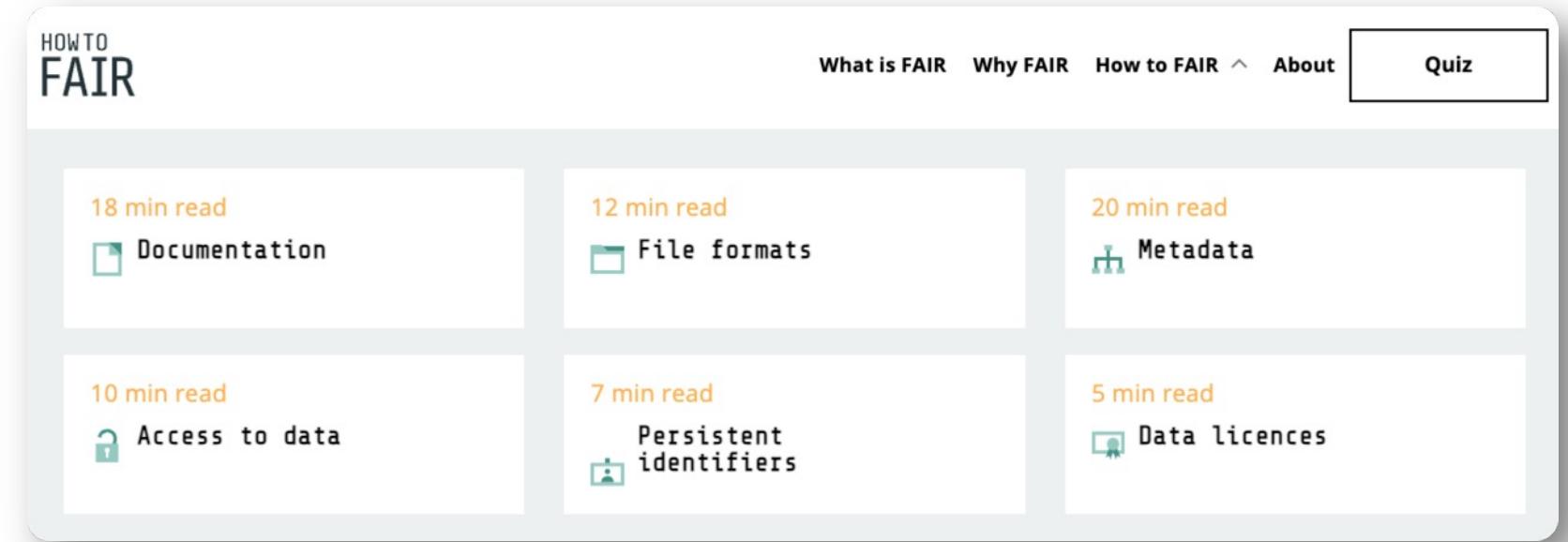
DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

# Resources

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



<https://howtofair.dk/>

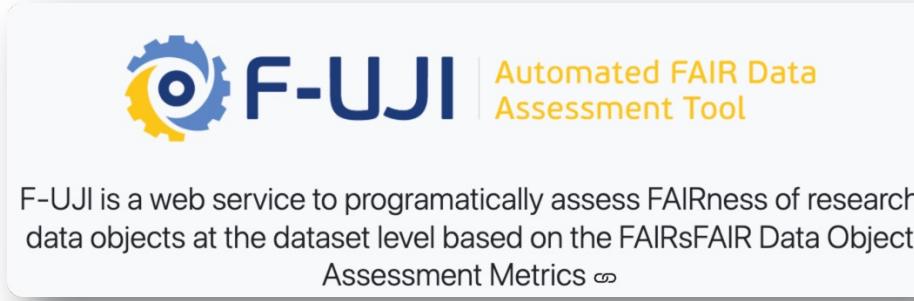


The screenshot shows the homepage of the HOWTO FAIR website. The header includes the title "HOWTO FAIR" and navigation links for "What is FAIR", "Why FAIR", "How to FAIR", "About", and "Quiz". Below the header are six cards arranged in two rows of three. Each card contains an icon, a title, and a "min read" time estimate.

Icon	Title	Time Estimate
Documentation	Documentation	18 min read
File formats	File formats	12 min read
Metadata	Metadata	20 min read
Access to data	Access to data	10 min read
Persistent identifiers	Persistent identifiers	7 min read
Data licences	Data licences	5 min read

# Tools

<https://www.f-iji.net/>

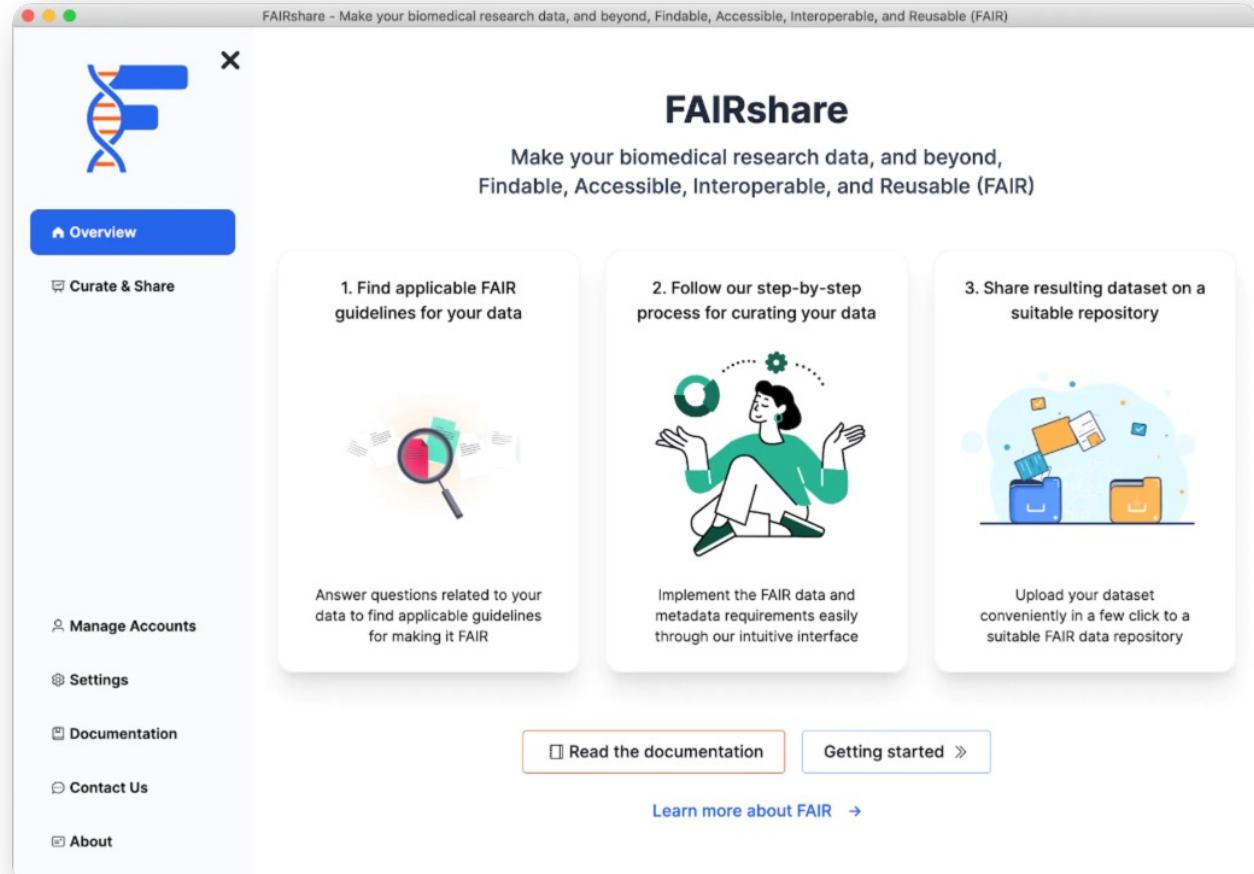


F-UJI | Automated FAIR Data Assessment Tool

F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object

Assessment Metrics ↗

<https://fairdataihub.org/fairshare>



FAIRshare - Make your biomedical research data, and beyond, Findable, Accessible, Interoperable, and Reusable (FAIR)

## FAIRshare

Make your biomedical research data, and beyond, Findable, Accessible, Interoperable, and Reusable (FAIR)

1. Find applicable FAIR guidelines for your data
2. Follow our step-by-step process for curating your data
3. Share resulting dataset on a suitable repository

Answer questions related to your data to find applicable guidelines for making it FAIR

Implement the FAIR data and metadata requirements easily through our intuitive interface

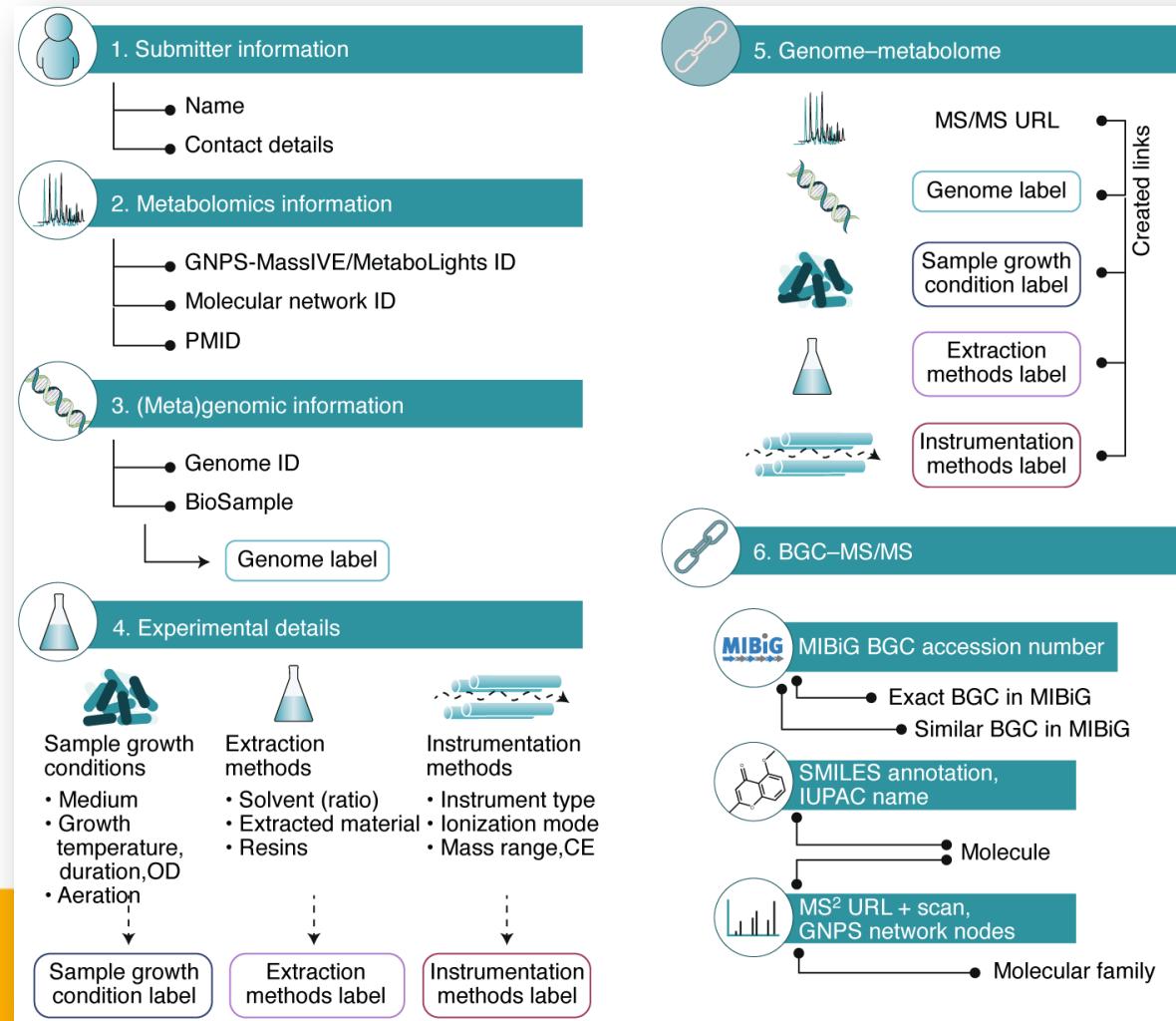
Upload your dataset conveniently in a few click to a suitable FAIR data repository

Read the documentation →      Getting started ➞

Learn more about FAIR →

Manage Accounts    Settings    Documentation    Contact Us    About

# Paired Omics Data Platform



PODP provides a community-based standard to document links between genomic and metabolomics data in a FAIR way.

**F:**

- Each project has an unique PID on PODP
- Human/machine-readable metadata

**A:**

- Publicly available on PODP website (<http://>)

**I:**

- A community-based standard (metadata)
- JSON format
- Reference to other data (GNPS, Genome ID, etc)

**R:**

- Detailed provenance

...

# Take Home Messages

- Your research data/software are valuable
- FAIR promotes (re)usability of data/software
- FAIRification is not an either/or process
- There is no one-size-fits-all solution of FAIRification



**Let's stay  
in touch**

 [www.eScienceCenter.nl](http://www.eScienceCenter.nl)

 [c.geng@esciencecenter.nl](mailto:c.geng@esciencecenter.nl)

 [@CunliangGeng](https://github.com/CunliangGeng)