

# FAIR train the trainers Methodology section FAIR tools

February 20, 2019

Judit Fazekas-Paragh

Edit Görögh

Ádám Száldobágyi



- FAIR principles
- How and where to choose the right FAIR tools
- Tools for all disciplines:
  - FAIRsharing
  - RDA Metadata Standards Directory
  - FAIR Evaluation Services
  - ARGOS
  - AMNESIA
  - B2SHARE, Zenodo
  - Checklist to evaluate FAIRness of data

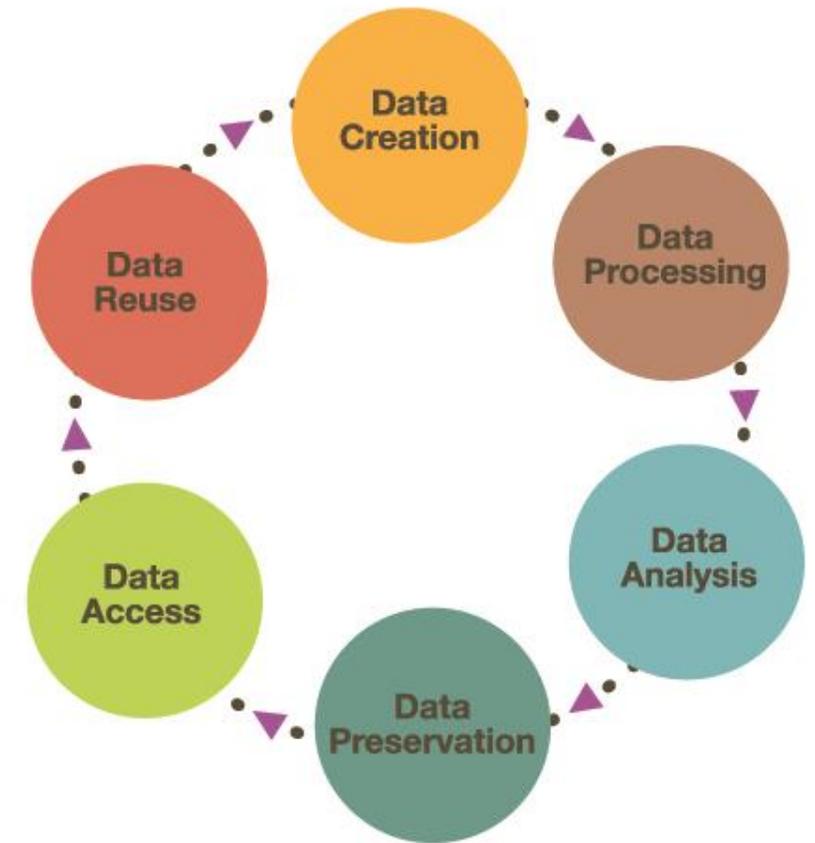
# FAIR principles I



# FAIR principles II

What aspects to consider when looking for the right tool?

- ❑ Research data life cycle
- ❑ Disciplines
- ❑ The possibilities and expectations of the given institution
- ❑ Funder and journal expectations



Picture: <https://blogs.ntu.edu.sg/lib-datamanagement/data-lifecycle/>

# How and where to choose the right FAIR tools

Databases:

For everyone:

[Open Science MOOC](#)

[Research Data Alliance](#)

[400+ Tools and innovations in scholarly communication](#)

NI4OS partners:

[NI4OS Box tools](#)

[NI4OS Box ORDМ](#)

---

# FAIRsharing

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

The screenshot displays the FAIRsharing.org website interface. At the top left is the logo "FAIRsharing.org" with the tagline "standards, databases, policies". A search bar is located at the top right, along with navigation buttons for "Standards", "Databases", "Policies", "Collections", "Add/Claim Content", "Stats", and "Log in or Register".

The main content area shows a table of records, with the first row highlighted. The table has the following columns: Registry, Name, Abbreviation, Type, Subject, Domain, Taxonomy, Related Database, Related Standard, Related Policy, In Collection/Recommendation, and Status. The first record is for "GenomeHubs" (Registry: GenomeHubs, Abbreviation: GenomeHubs, Type: Database, Subject: Life Science, Domain: Gene Prediction, Genome Annotation, Genome Assembly, Genome Visualization, Genome, Taxonomy: AI, Related Database: Ensembl, Related Standard: None, Related Policy: None, In Collection/Recommendation: None, Status: R).

On the left side, there are several filter sections:

- 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**
- Record Type**: Journal (85), Funder (23), Project (13), Society (9)
- Domains**: Chemical Entity (382)

At the bottom of the table, there is a pagination bar showing "Showing records 1 - 50 of 1475." and a set of numbered links from 1 to 30.

# RDA Metadata Standards Directory

The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data can help address infrastructure challenges.

The screenshot displays the RDA Metadata Standards Directory website. On the left is a sidebar with the following navigation options:

- Metadata (RDA | Metadata Directory)
- Edit this page
- View the standards
- View the extensions
- View the tools
- View the use cases
- Browse by subject areas
- Contribute
- Add standards
- Add extensions
- Add tools
- Add use cases
- github
- @twitter
- linkedin
- facebook

The main content area is organized into subject categories:

- Arts and Humanities**
  - [FISH Interoperability Toolkit](#) [Edit](#)  
A suite of tools using the [MIDAS](#) Heritage metadata standard to facilitate the process of moving information between the wide variety of information systems used to record the historic environment.
- Engineering**
  - [CIF2Cell](#) [Edit](#)  
A tool to generate the geometrical setup for various electronic structure codes from a [CIF](#) file.
  - [ICATLite](#) [Edit](#)  
A sister project of ICAT, consisting of a suite of [CSMD](#)-based software tools designed to support derived data management in the scientific research process.
  - [IUCr checkCIF](#) [Edit](#)  
A tool used to check the integrity and consistency of crystal structure encodings in [CIF](#) format.
  - [Software for CIF](#) [Edit](#)  
The International Union of Crystallography's list of programs and libraries available for use with [CIF](#) files.
- Life Sciences**
  - [Bio-Formats](#) [Edit](#)  
Bio-Formats reads proprietary microscopy image data and metadata, and converts them to OME-TIFF, a combination of TIFF and [OME-XML](#).
  - [Darwin Core Archive Assistant](#) [Edit](#)  
A web application that offers data publishers wishing to serve to the GBIF network an easy interface for describing data elements as basic text files, composing an appropriate XML [Darwin Core](#) descriptor file to accompany them.
  - [Darwin Core Archive Validator](#) [Edit](#)  
A tool to validate XML metadata against the [Darwin Core](#) Text Guidelines.
  - [Fiji](#) [Edit](#)  
[Fiji](#) is an image processing package that supports the [OME data model](#) for images
  - [Integrated Publishing Toolkit](#) [Edit](#)  
A software platform using [Darwin Core](#) and [EML](#) to facilitate the efficient publishing of biodiversity data on the Internet, using the GBIF network.
  - [ISA Software Suite](#) [Edit](#)  
The open source ISA metadata tracking tools facilitate [ISA-TAB](#)-compliant collection, curation, local management and reuse of datasets in an increasingly diverse set of life science domains.
  - [Metacat](#) [Edit](#)  
Metacat is a repository for data and metadata that helps scientists find, understand, and effectively use the data sets they manage or that have been created by others.
  - [MOLGENIS](#) [Edit](#)  
A software generator to rapidly build web databases and a suite of web databases for genotype, phenotype, QTL and analysis pipelines.
  - [Morpho](#) [Edit](#)  
An application for accessing and manipulating metadata and data (both locally and on the network) with various analysis metadata files using a subset of Ecological Metadata Language (EML).

ARGOS is an online tool in support of automated processes to creating, managing, sharing and linking DMPs with research artifacts they correspond to.

The screenshot displays the ARGOS web application interface. At the top left is the ARGOS logo, followed by a navigation menu with 'HOME' and a search bar. The main content area features a large banner image of a snowy mountain range with the text 'Welcome to ARGOS' and 'Create, Link, Share Data Management Plans'. Below the banner are four summary cards: 'DMPs' (1), 'Dataset Descriptions' (1), 'Grants' (1), and 'Related Organizations' (2). A 'What is ARGOS?' section provides a brief description of the service. Below this are sections for 'DATA MANAGEMENT PLANS' (1 DMP) and 'DATASETS' (1 Dataset Description). The bottom of the page includes a footer with 'Glossary', 'FAQ', and 'Terms Of Service'.

ARGOS

HOME >

SEARCH... Log in

GENERAL

- Home
- About

PUBLISHED

- Published DMPs
- Published Dataset Descriptions

Welcome to ARGOS  
Create, Link, Share Data Management Plans

DMPs 1  
VIEW ALL

Dataset Descriptions 1  
VIEW ALL

Grants 1

Related Organizations 2

What is ARGOS?

ARGOS is an open extensible service that simplifies the management, validation, monitoring and maintenance and of Data Management Plans. It allows actors (researchers, managers, supervisors etc) to create actionable DMPs that may be freely exchanged among infrastructures for carrying out specific aspects of the Data management process in accordance with the intentions and commitment of Data owners.

DATA MANAGEMENT PLANS  
1 DMPs  
VIEW ALL

DMP For NEANIAS Project  
just a summary  
HORIZON 2020

DATASETS  
1 Dataset Descriptions  
VIEW ALL

Horizon 2020 Dataset Description

Glossary  
Terms Of Service  
FAQ

# AMNESIA

Amnesia is a data anonymization tool, that allows to remove identifying information from data. Amnesia not only removes direct identifiers like names, SSNs etc but also transforms secondary identifiers like birth date and zip code so that individuals cannot be identified in the data.

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

k-anonymization



ID	Age	Zipcode	Diagnosis
1	[20-30]	130**	Heart Disease
2	[20-30]	130**	Heart Disease
3	[20-30]	130**	Viral Infection
4	[20-30]	130**	Viral Infection
5	[40-60]	148**	Cancer
6	[40-60]	148**	Heart Disease
7	[40-60]	148**	Viral Infection
8	[40-60]	148**	Viral Infection
9	[30-40]	13***	Cancer
10	[30-40]	13***	Cancer
11	[30-40]	13***	Cancer
12	[30-40]	13***	Cancer

B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store and publish small-scale research data from diverse contexts.

The screenshot displays the B2SHARE website interface. At the top, there is a navigation bar with the B2SHARE and EUDAT logos, a search bar with the placeholder text "Search records for...", and a "SEARCH" button. Below the search bar are links for "HELP", "COMMUNITIES", "UPLOAD", and "CONTACT", along with a "Login" button. The main content area features the heading "Store and publish your research data" and a sub-heading "Search in public datasets or register as a user to upload and publish your data!". Below this, there is a note: "Please use <https://trng-b2share.eudat.eu> for testing or training." and links for "Login or Register". A prominent orange button labeled "Create a new record" is positioned in the center. The "Latest Records" section lists several recent uploads with their titles, dates, and authors. The records listed are:

- SNIC Storage Infrastructure and FAIR**  
9 Feb 2020 by Vitlacil, Dejan.  
"SNIC Storage Infrastructure and FAIR" is a presentation given by Vitlacil Dejan on importance of FAIR principles for Open Science from national service provider perspective.
- FLASH\_H2020\_Electrical\_characterization\_Dataset**  
7 Feb 2020 by Persichetti, Luca.  
Electrical Characterization Dataset for the FLASH project- Far-infrared Lasers Assembled using Silicon Heterostructures (756719) supported by the European Union research and innovation programme Hori
- LTER Mar Piccola phytoenhthos biomass from 2011 to 2019**  
3 Feb 2020 by Petrocelli, Antonella; Portacci, Giuseppe; Cecere, Ester.  
This data set includes biomass values of phytoenhthos species, mainly seaweeds, recorded in the Mar Piccolo of Taranto in the years 2011-2019. Seasonal collection campaigns have been carried out in fo
- A pure shift experiment with increased sensibility and superior performance for strongly coupled systems**  
28 Jan 2020 by Julian Ilgen, Lukas Kaltschnee, Christina Marie Thiele.  
Motivated by the persisting need for enhanced resolution in solution state NMR spectra, pure shift techniques such as Zangger-Sterk decoupling have recently attracted widespread interest. These techni
- Specificity, crosstalk and redundancy between the chloroplast thioeredoxin systems in regulation of photosynthesis during the development of Arabidopsis rosettes (1 of 2)**  
26 Jan 2020 by Guinea, Diaz, Manuel; Nikkanen, Lauri; Himanen, Kristina; Toivola, Jouni; Rintamäki, Eevi.  
Light is a substrate of photosynthetic reactions in plants but its intensity constantly fluctuates in the field. Optimization of photosynthetic production under fluctuating light conditions needs stri
- FLASH\_H2020\_Structural\_characterization\_Dataset**  
7 Feb 2020 by Persichetti, Luca.  
Structural Characterization Dataset for the FLASH project- Far-infrared Lasers Assembled using Silicon Heterostructures (756719) supported by the European Union research and innovation programme Hori
- Dataset for "Terahertz Absorption-Saturation and Emission in Electron-doped Germanium Quantum Wells"**  
6 Feb 2020 by Persichetti, Luca.  
Raw data acquired during beamtime at the FELBE FEL light source discussed in C. Ciano, M. Virgilio, L. Bagolini, L. Baldassarre, A. Pashkin, M. Helm, M. Montanari, L. Persichetti, L. Di Gaspare, G. Ca
- Low-noise transfer of the spectral purity of an optical comb line using a feedforward scheme**  
3 Feb 2020  
Data sets of the paper entitled "Low-noise transfer of the spectral purity of an optical comb line using a feedforward scheme" PBROCHARD, et al., published in the journal JEOS.
- Specificity, crosstalk and redundancy between the chloroplast thioeredoxin systems in regulation of photosynthesis during the development of Arabidopsis rosettes (2 of 2)**  
26 Jan 2020 by Guinea, Diaz, Manuel; Nikkanen, Lauri; Himanen, Kristina; Toivola, Jouni; Rintamäki, Eevi.  
Light is a substrate of photosynthetic reactions in plants but its intensity constantly fluctuates in the field. Optimization of photosynthetic production under fluctuating light conditions needs stri
- Aeroelastic response of a multi-megawatt upwind horizontal axis wind turbine (HAWT) based on fluid-structure interaction simulation**  
26 Jan 2020 by Guinea, Diaz, Manuel; Nikkanen, Lauri; Himanen, Kristina; Toivola, Jouni; Rintamäki, Eevi.

Zenodo is an open repository for all scholarship, enabling researchers from all disciplines to share and preserve their research outputs.

The screenshot displays the Zenodo website interface. At the top, there is a blue navigation bar with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right side of the navigation bar, there are 'Log in' and 'Sign up' buttons. Below the navigation bar, the main content area is divided into several sections:

- Recent uploads:** A list of recent uploads with details for each:
  - DWD European Weather:** Family name, given names. opendata.dwd.de - OpenData by Deutscher Wetter Dienst Conditions: https://www.dwd.de/EN/service/copyright/copyright\_node.html dates and times are UTC. Uploaded on April 3, 2019. 164 more version(s) exist for this record.
  - Pyndl: Naive discriminative learning in python:** Konstantin Spring, Marc Weitz, David-Elias Künstle, Lennart Schneider. fixes encoding error in long description of setup.py a lot of small code improvements improvements in the maintenance code. Uploaded on April 2, 2019. 17 more version(s) exist for this record.
  - scikit-hep/particle: 0.4.1:** Eduardo Rodriguez, Henry Schreiner. Enhancements to Particle class: Particles in .dec decay files dealt with, see Particle.from\_dec(...). Loading tables made nicer, with more documentation. Particle charge is an entry of CSV files again, so that user particles are better dealt with. Bug fix for corner cases of using the... Uploaded on April 2, 2019. 6 more version(s) exist for this record.
  - Discovery of tandem and interspersed segmental duplications using high throughput sequencing:** Uploaded on March 27, 2019 (v1). Dataset. Open Access.
- Zenodo now supports usage statistics!** (with a line graph icon): Read more about it, in our newest blog post.
- Using GitHub?** (with a GitHub logo icon): Just Log in with your GitHub account and click here to start preserving your repositories.
- Zenodo in a nutshell:** A list of features:
  - Research. Shared.** – all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
  - Citeable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
  - Communities** – create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
  - Funding** – identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
  - Flexible licensing** – because not everything is under Creative Commons.
  - Safe** – your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.Read more about Zenodo and its features.

# Checklist to evaluate FAIRness of data

This checklist helps you assess the quality (FAIRness) of your dataset(s) and the trustworthiness of the repository that you have chosen.

## Checklist to evaluate FAIRness of data(sets)

You would like to deposit one or several dataset(s) at a digital repository but you are not sure whether the information you provide is sufficient and in line with the principles of FAIR (Findable, Accessible, Interoperable, Reusable)? This checklist helps you assess the quality (FAIRness) of your dataset(s) and the trustworthiness of the repository that you have chosen.

The assessment will cover four levels:

1. The data repository you are planning to use
2. The metadata with which you describe your dataset
3. The dataset itself
4. The data files of which your dataset consists

This checklist, furthermore, draws upon two core concepts: that of the trustworthy repository and that of FAIR data. The CoreTrustSeal (CTS) Data Repository Certification (<https://www.coretrustseal.org/>) is taken as an example for certified trustworthy repositories. Repositories with such a certification are to a large degree already compliant with the FAIR principles. A list of CTS-certified repositories can be found here: <https://www.coretrustseal.org/why-certification/certified-repositories/>  
More information about FAIR and the principles per character is provided on the website of the Go-FAIR initiative: <https://www.go-fair.org/fair-principles/>

The checklist consists of 7 sections including a feedback section at the end.

The structure of the questions per letters will be as follows:

- Data repository: 1 question
- Findability (F): 3 questions
- Accessibility (A): 1 question
- Interoperability (I): 2 questions
- Reusability (R): 3 questions
- Additional question: 1 question

Thank you for your attention!