Guide to FAIR BioImage Data

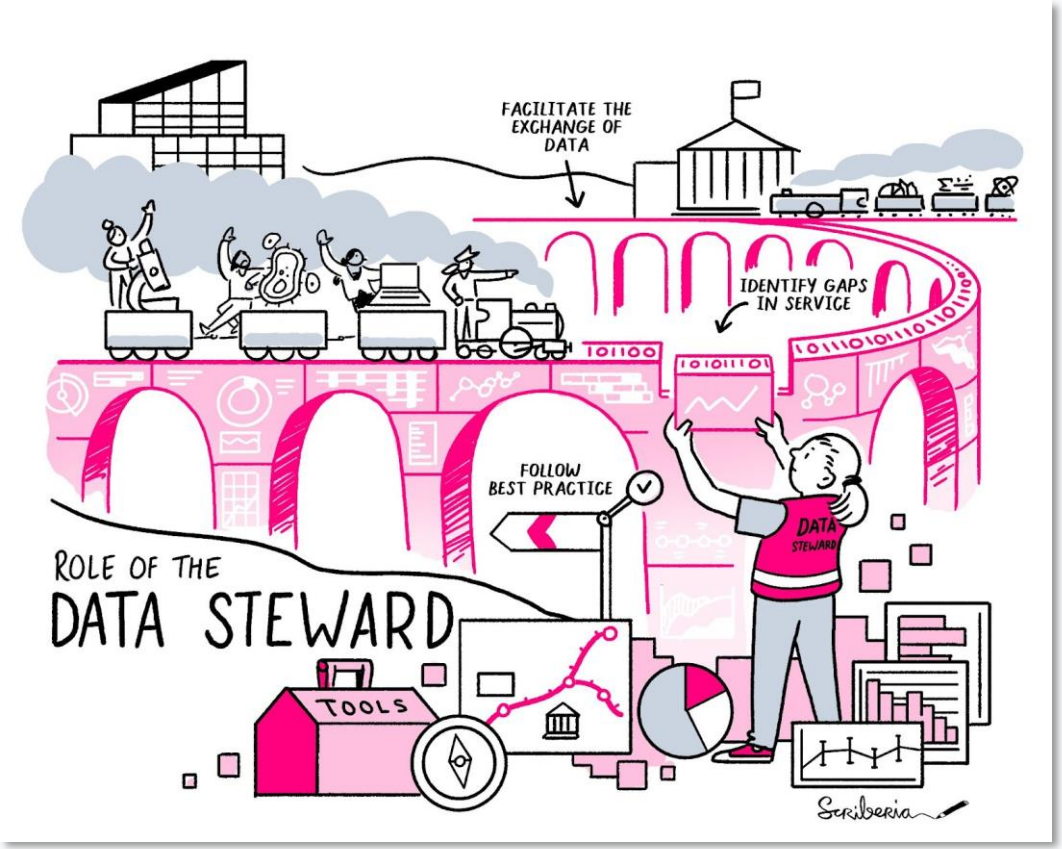
Isabel Kemmer

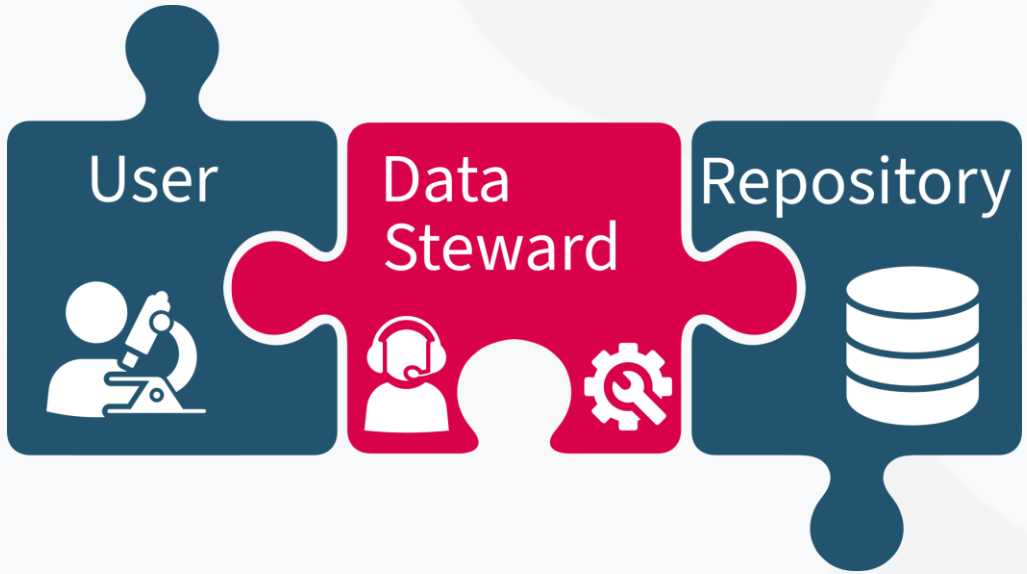
FAIR Image Data Steward
Euro-BioImaging Bio-Hub

Funded by the European Union

BY-COVID beyond covid

Image Data Stewardship





fairdata@eurobioimaging.eu

Benefits of FAIR data …

… for the scientific community

- Improves research transparency and reproducibility
- Enhances robustness and quality of results
- Increases the value of scientific data
- Boosts collaboration within the scientific community
- Accelerates method development and scientific progress

… for the individual researcher

- Additional citations on published articles[1]

[1] Colavizza, Giovanni et al. "The citation advantage of linking publications to research data." *PloS one* vol. 15,4 e0230416. 22 Apr. 2020, doi:10.1371/journal.pone.0230416

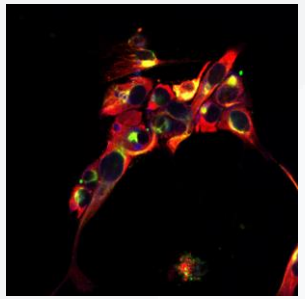
- Increased visibility and discoverability of your work
- A secure place to house the dataset
- Time saving in the long run (data organization, paper writing, peer review)
- Qualifiying criteria for journals and funders

In the future, FAIR will no longer be a choice but a **requirement**!

FAIR across disciplines

Structural biology

Bioimaging

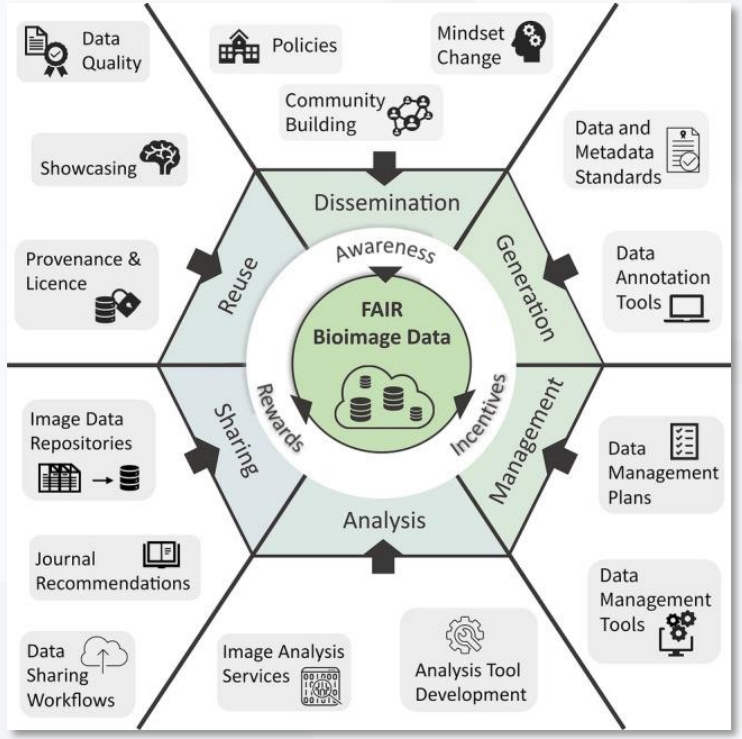


FAIR is a spectrum – start somewhere!

Life Science-communities are at different stages of FAIR

Special challenges for Bioimaging

- *Large volumes of data (several TB)*
- *Various (proprietary) file formats*
- *No widely-adopted standards*
- *Diverse methodologies*
- *Complex, multidimensional data*



Kemmer, Isabel et al. "Building a FAIR image data ecosystem for microscopy communities." Histochemistry and cell biology vol. 160,3 (2023): 199-209. doi:10.1007/s00418-023-02203-7

F A I R

FINDABLE

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

ACCESSIBLE

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

INTEROPERABLE

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

REUSABLE

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

Repository

FAIR tip: Depositing datasets in repositories is a crucial step towards FAIR data

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

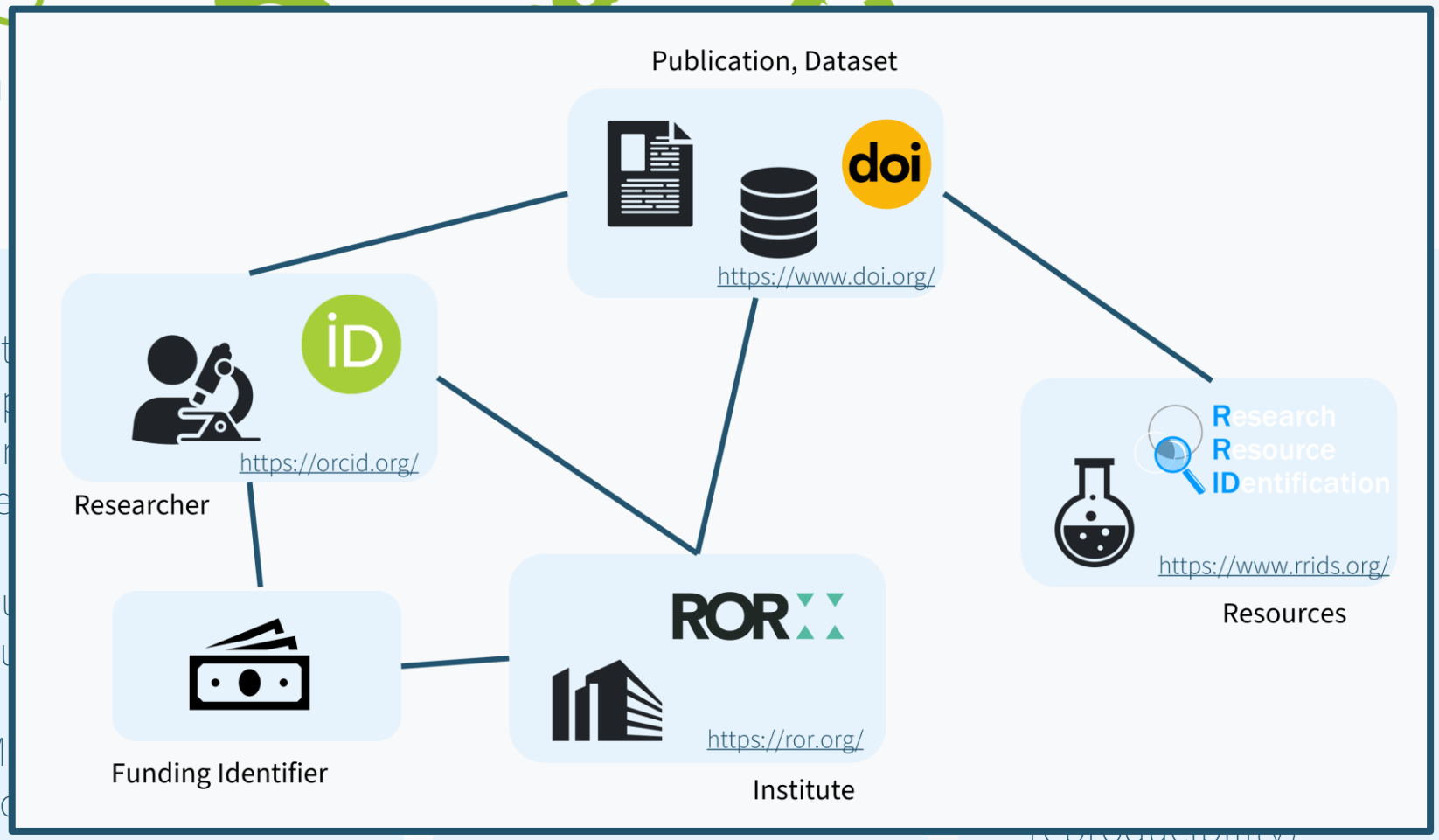
- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

Repository

F A I R

FIN

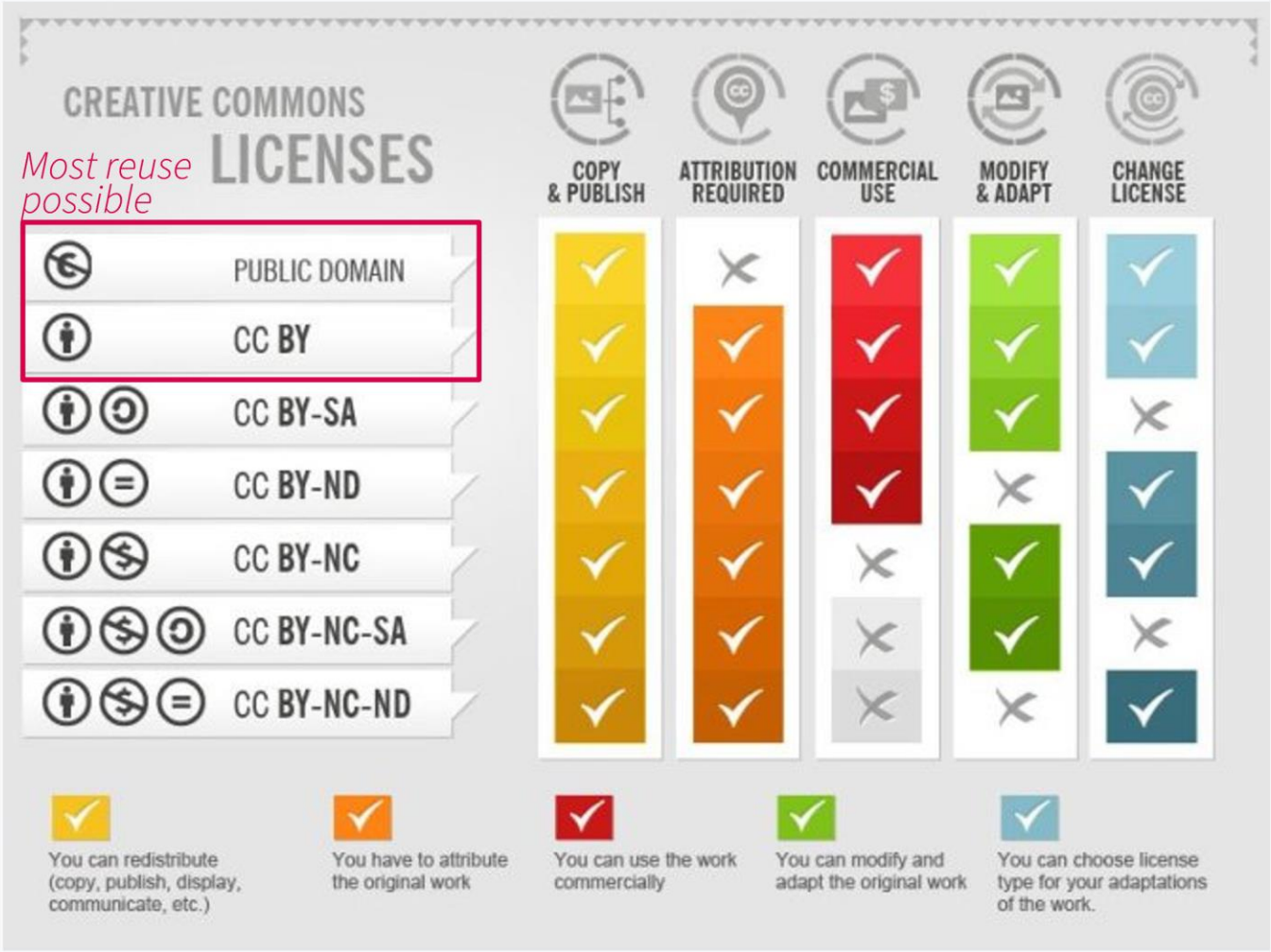
- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata



Rep

- Data
  sear
- Mac
  pers
- Des
  met

For open science resources (and open access publishing) the Creative Commons (CC) licenses are most commonly used



Creative Commons licenses by Foter (CC-BY-SA)

R

REUSABLE

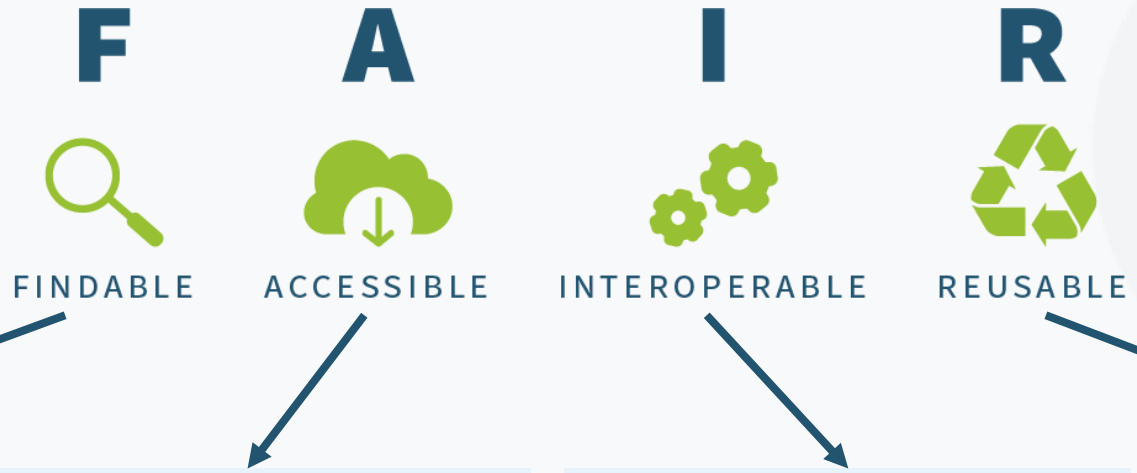
ndard file

cabulary

nces to

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

Repository



- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

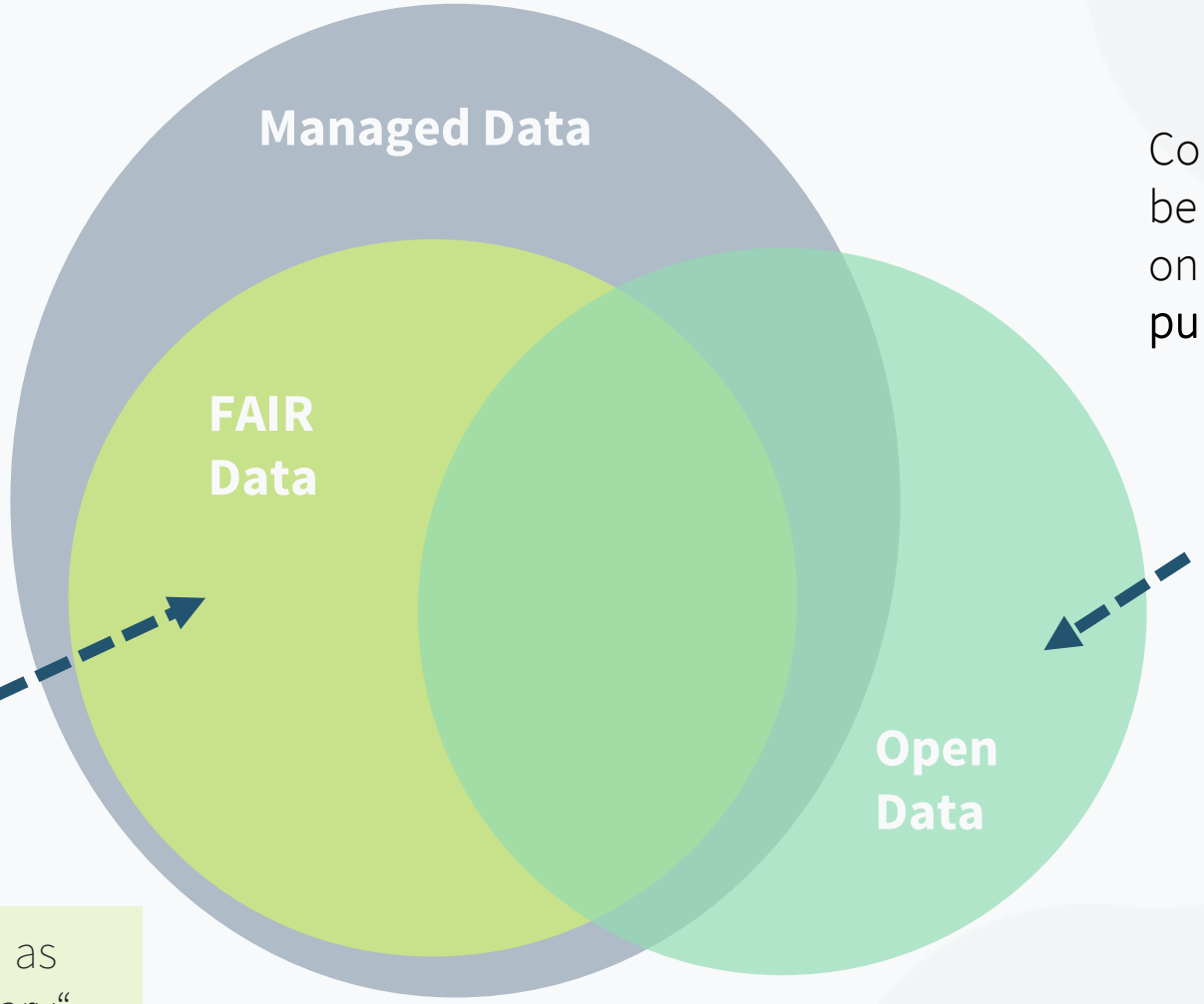
- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

FAIR data ≠ Open data

Always managed and shared **respecting** any ethical, legal or contractual **restrictions.**



Completely open data can be used, shared and built-on by **anyone for any purpose**

FAIR data may not always be open

Example: a dataset can be accessible only to a closed group, but still be findable, reusable, and interoperable.

Open Data are not necessarily FAIR

Example: Undocumented data dump on a personal homepage is neither findable, reusable nor interoperable.

FAIR tip: Make data „as open as possible, as closed as neccesary“

Different types of repositories

General purpose repositories

Examples:

zenodo

DRYAD

figshare

BioStudies.

Institutional repositories

Example:

SciLifeLab

Domain-specific repositories

Open repositories

Examples:

EMPIAR

IDR

Restricted-access repositories

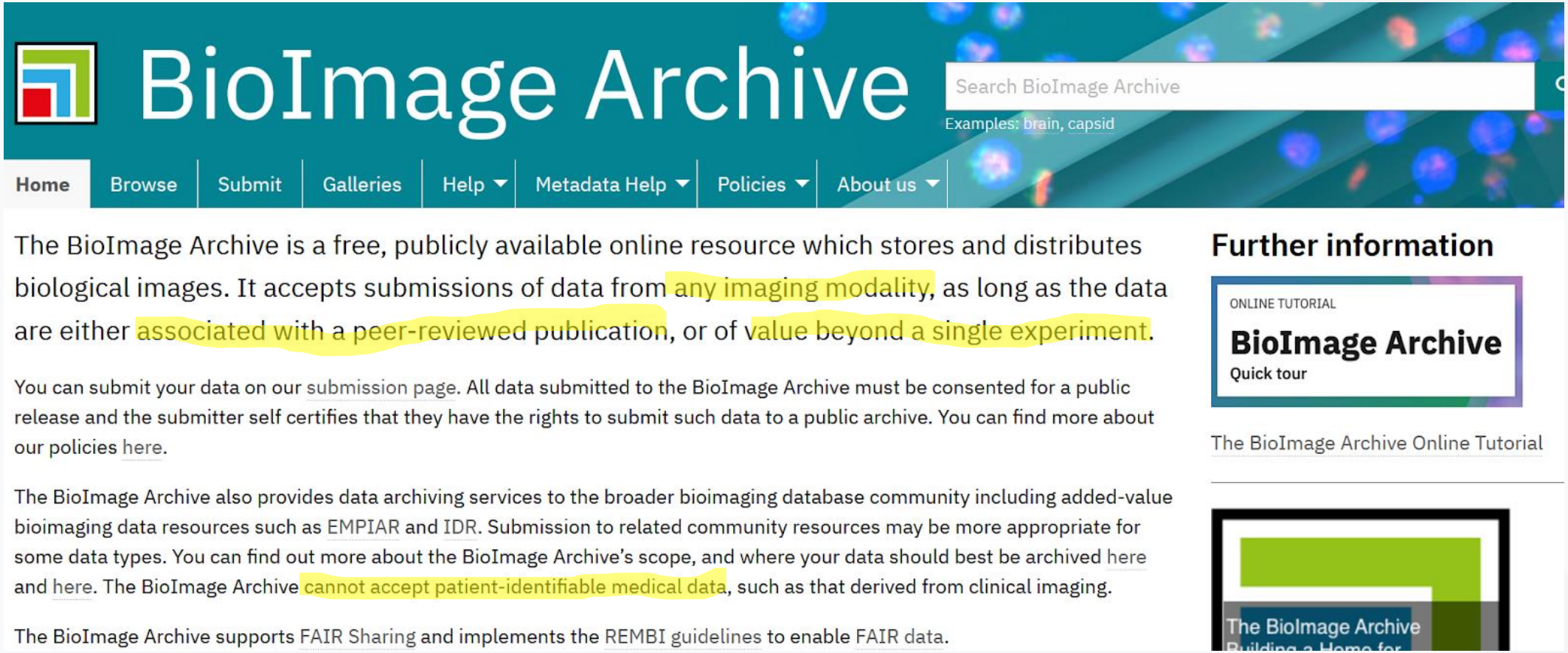
Example:

CANCER IMAGING ARCHIVE

Metadata repositories

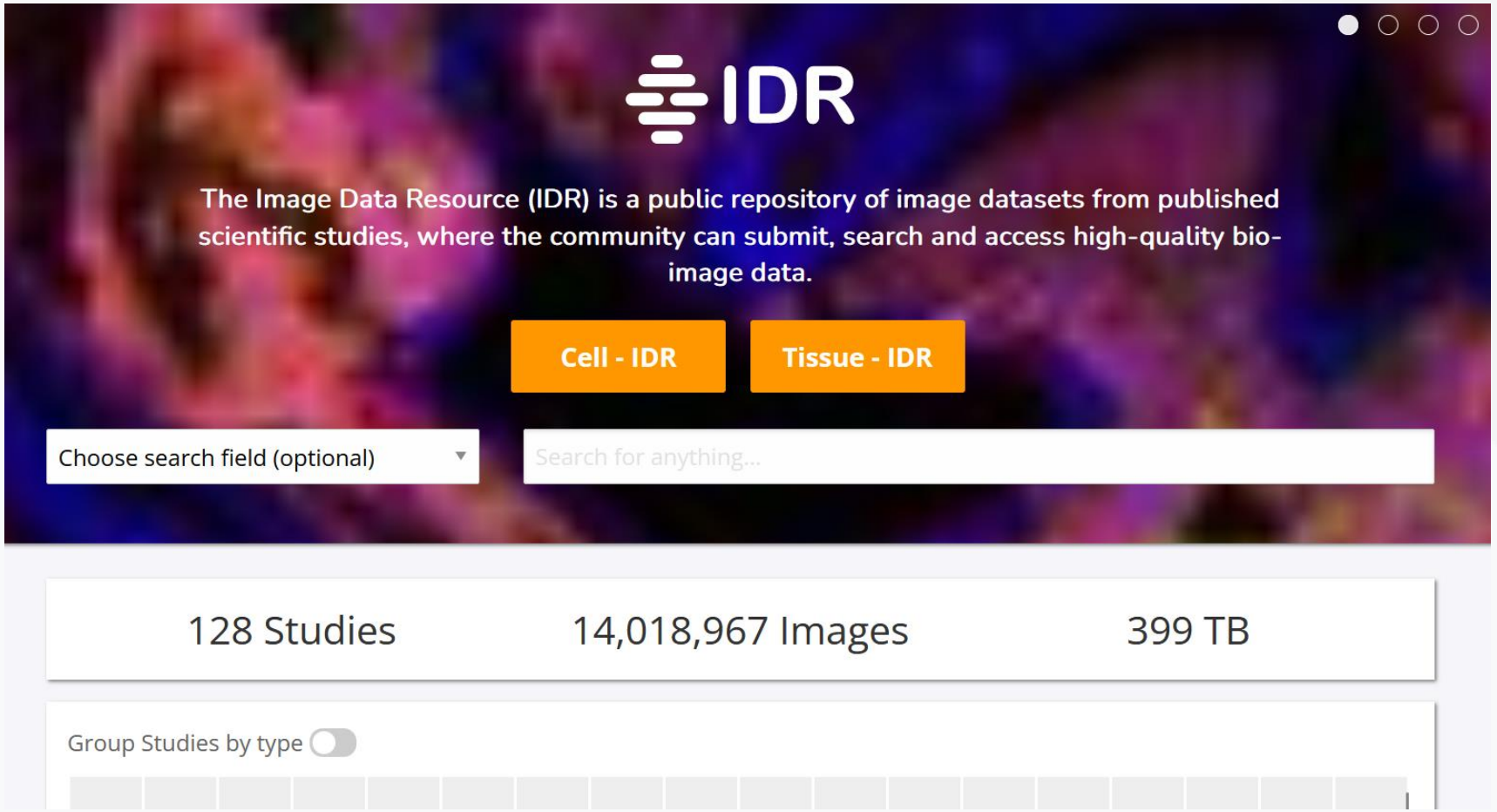
Example:

PIDAR



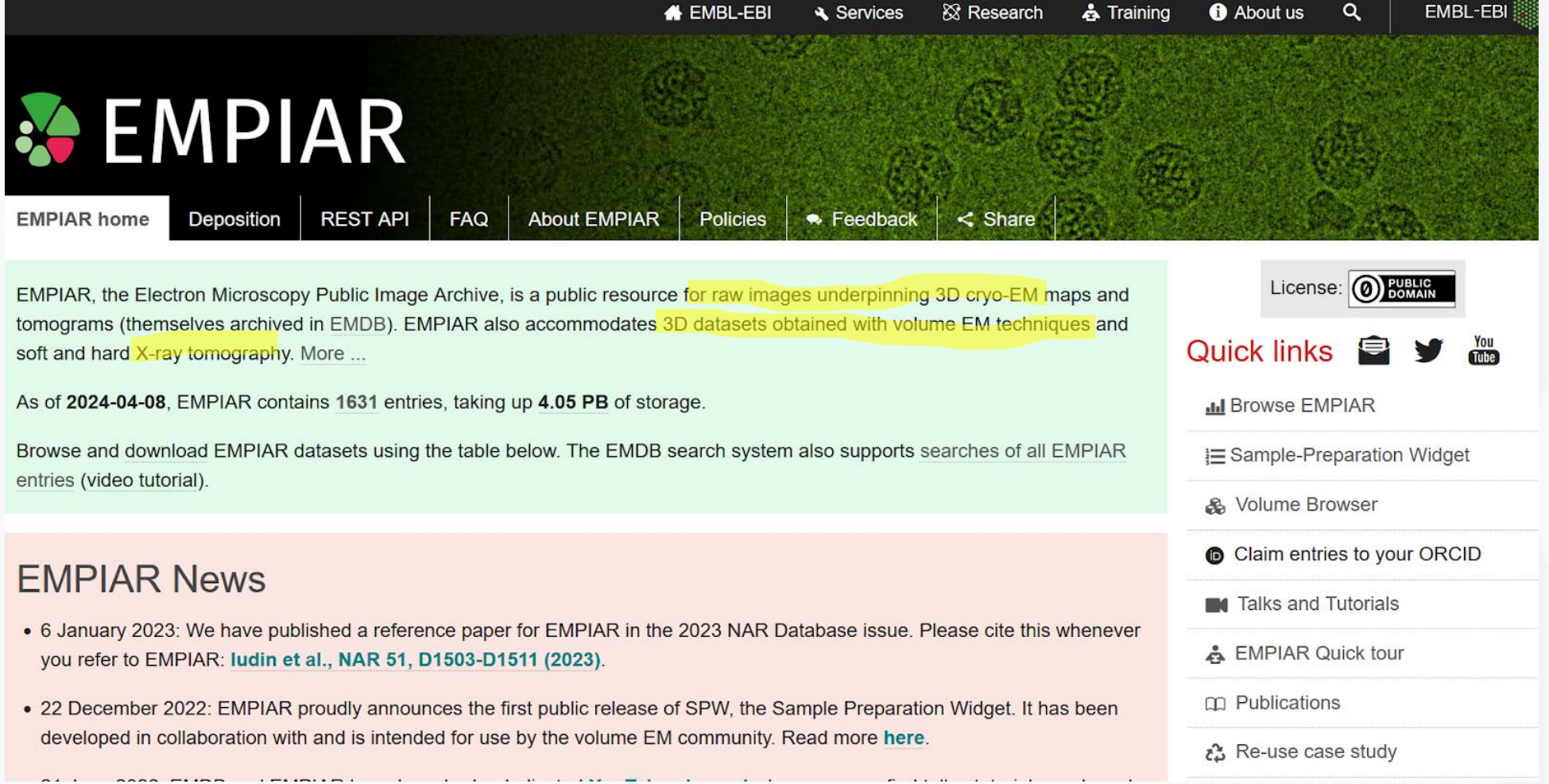
- Open archive
- Broad scope
- Low metadata requirements
- Quick deposition process

Links:
BioImage Archive
Scope
Submission Guide
Download



Publishes complete Reference datasets containing molecular and functional annotations, associated with an existing or upcoming publication with curated metadata and integration with other datasets/studies.

Links:
IDR
Submission Overview
API access



Links:

EMPIAR

Deposition Manual

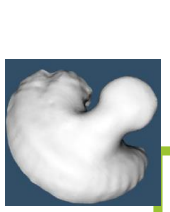
Sample preparation Widget

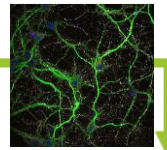
API

Open BioImage Repositories

EURO BIOIMAGING

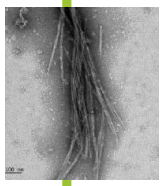
Where should I deposit my (open) bioimaging data?



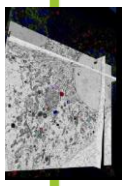


Which imaging method did you perform?

- Other imaging technique
  - *(including, but not limited to, AFM/ ultrasound imaging/ photoacoustic imaging…)*
- Light/fluorescence microscopy
  - *(including but not limited to confocal, SPIM, super resolution techniques…)*
- Electron microscopy
  - 2D EM



  - 3D/volume EM



  - Cryo-EM/ET

    - Raw (2D) images
      - *(particle-stacks, micrographs, tilt-series)*
    - processed 3D volumes and tomograms
- X-ray tomography

Can your dataset be regarded as a complete reference * dataset?

- Yes → Is extensive, per-image metadata present and are you willing for manual image curation
  - Yes → IDR
  - No → BioImage Archive
- No → BioImage Archive

Light microscopy part of correlative studies

EMDB (Electron Microscopy Data Bank)

EMPIAR

* - **reference dataset:** mainly datasets related to cells and tissues where a comprehensive study across a number of genes, small molecule compounds and/or drugs or other perturbations are executed

F A I R

Metadata

FINDABLE

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

ACCESSIBLE

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

INTEROPERABLE

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

REUSABLE

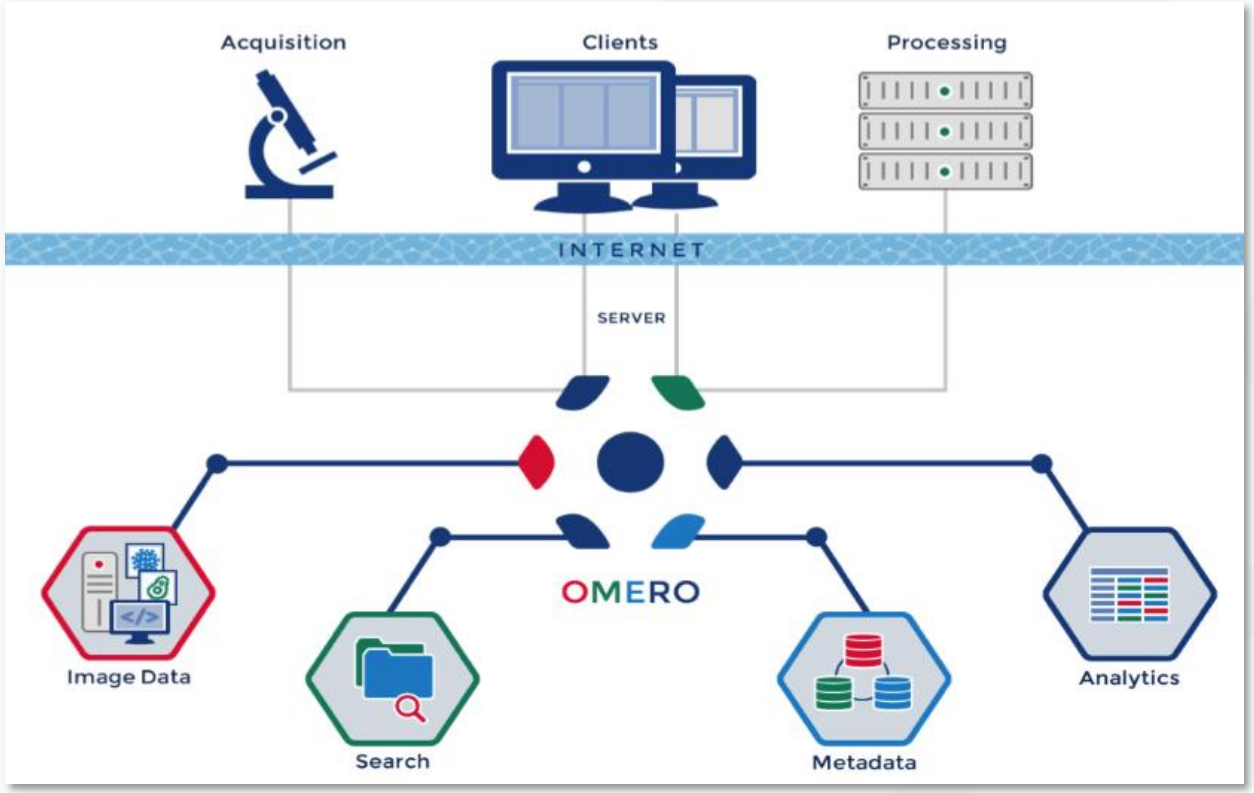
- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

Image management software - OMERO

FAIR tip: and consider using an image management software

Visualize, manage, annotate, and share scientific image data from anywhere in the centralized, **client-server platform** OMERO.

Stores acquisition metadata and user-added annotations (like ROI & Segmentations) in relational database



https://www.openmicroscopy.org/omero/

Community Tools for Metadata

Creates a **virtual representation** of your microscope and all of its parts, and allows you to export a microscope.json file with all of its specifications.

https://www.nature.com/articles/s41592-021-01315-z

https://www.nature.com/articles/s41592-021-01290-5

MethodsJ2

Plugin for Fiji

Generates **methods text for publication** using image metadata (via bioformats) and microscopy metadata (via Mirco-Meta App)

Visualize, manage, annotate, and share scientific image data from anywhere in the centralized, **client-server platform** OMERO.

Stores acquisition metadata and user-added annotations (like ROI & Segmentations) in relational database

OMERO

https://www.openmicroscopy.org/omero/

MDEmic

MetaData Editor for microscopy

https://www.nature.com/articles/s41592-021-01288-z

As part of OMERO installation package, MDEmic reads technical metadata from image files and presents it in the OME data model to **allow editing of that metadata.**

F A I R

Metadata

FINDABLE

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

ACCESSIBLE

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

INTEROPERABLE

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

REUSABLE

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

The problem with names

Different words for the same concepts
➡ difficult search

Light-sheet microscopy

Light-sheet fluorescence microscopy (LSFM)
Lattice Light-sheet Microscopy (LLSM)
Spherical aberrations assisted Extended Depth-of-field Lightsheet Microscopy
Bessel Beam Lightsheet Microscopy
single objective Selective Plane Illumination Microscopy
Orthogonal-plane fluorescence optical sectioning (OPFOS)
Selective plane illumination microscopy (SPIM)
OCPI
Dual-View inverted SPIM (diSPIM)
inverted SPIM (iSPIM)
Thin Sheet Laser Imaging Microscopy (TSLIM)
Multidirectional SPIM
Objective-Coupled Planar Illumination Microscopy
LLS
OPTiSPIM
Multiview selective plane illumination microscopy (MuViSPIM)
DSLM
mesoSPIM
Oblique Plane Microscopy (OPM)
Digitally scanned Laser Light-sheet Microscopy
Swept confocally-aligned planar excitation microscopy (SCAPE)
soSPIM
COLM
SPED
Clarity Optimized Light-sheet Microscopy
eSPIM
mSPIM

The same words for different concepts
➡ ambigous and confusing

palm

palm

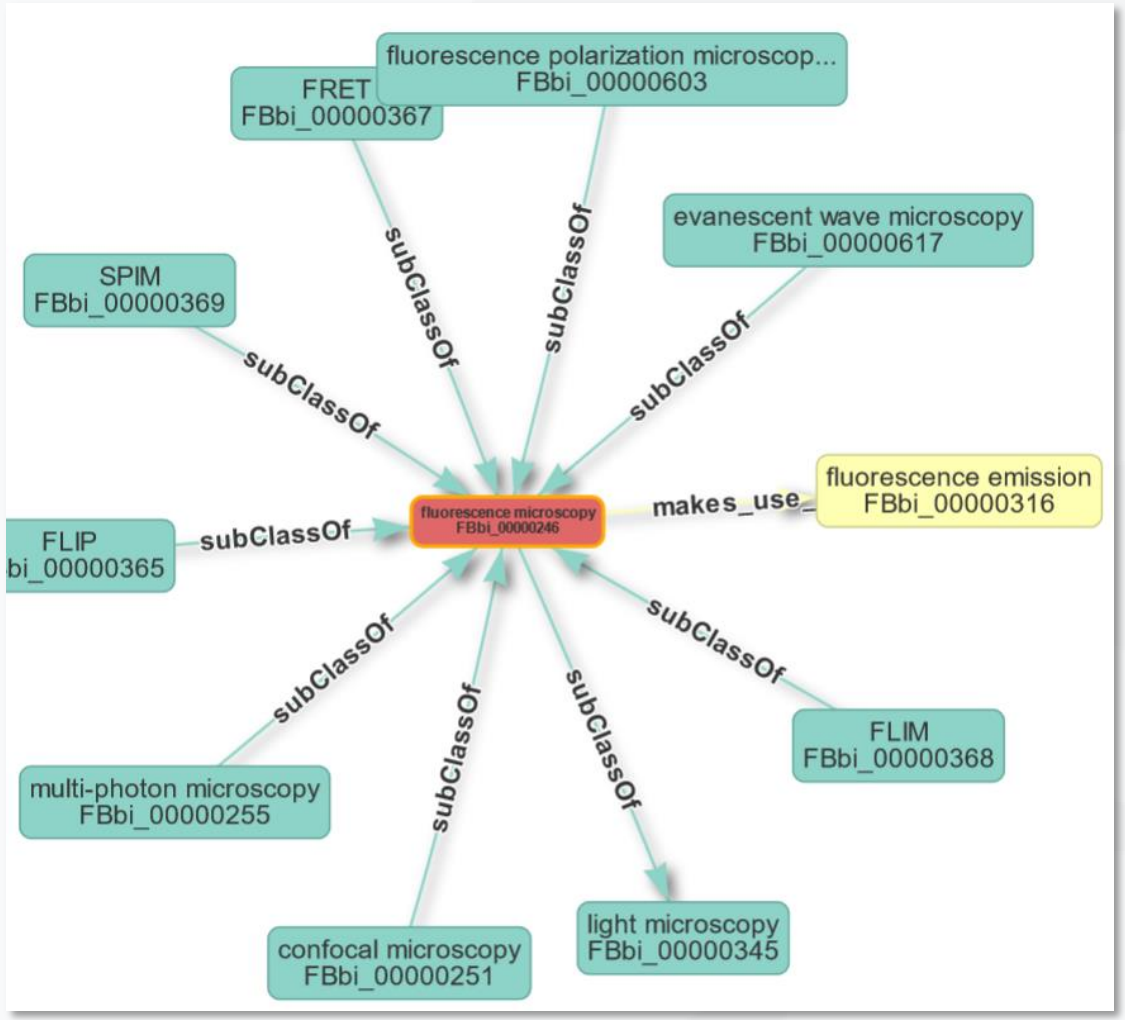
Goal: Make sure we are all speaking the same language

The solution to the problem with names

- **Ontology:** describes the categories of objects, their relationships, and the relationships between categories.

- The labels used to describe the objects can be used to create **a controlled vocabulary**

Example graph from FBbi Ontology



Find a suitable ontology with the Ontology-lookup service: https://www.ebi.ac.uk/ols4

The solution to the problem with names

- **Ontology:** describes the categories of objects, their relationships, and the relationships between categories.

- The labels used to describe the objects can be used to create **a controlled vocabulary**

Some example identifiers and ontologies

Name	Example ontology	link
Organism	NCBI Taxonomy	https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/
Gene	Ensembl	https://www.ensembl.org/index.html
Protein	Uniprot	https://www.uniprot.org/
Compound/ Drug	PubChem/ ChEMBL	https://pubchem.ncbi.nlm.nih.gov/
Imaging Method	FBBI /EDAM-BioImaging	https://www.ebi.ac.uk/ols4/ontologies/fbbi
Study Type	Experimental Factor Ontology	https://www.ebi.ac.uk/efo/
Phenotypes	Human Phenotype Ontology	https://hpo.jax.org/app/
…	…	…

Find a suitable ontology with the Ontology-lookup service: https://www.ebi.ac.uk/ols4

FAIR tips:

- Crosslink to as much other resources as possible
- Search for ontologies that are used in repository and collect data accordingly

F A I R

Metadata

FINDABLE

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

ACCESSIBLE

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

INTEROPERABLE

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

REUSABLE

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

REcomended Metadata for Biological Images

8 categories of metadata with sub keywords

Partly stored in the image itself

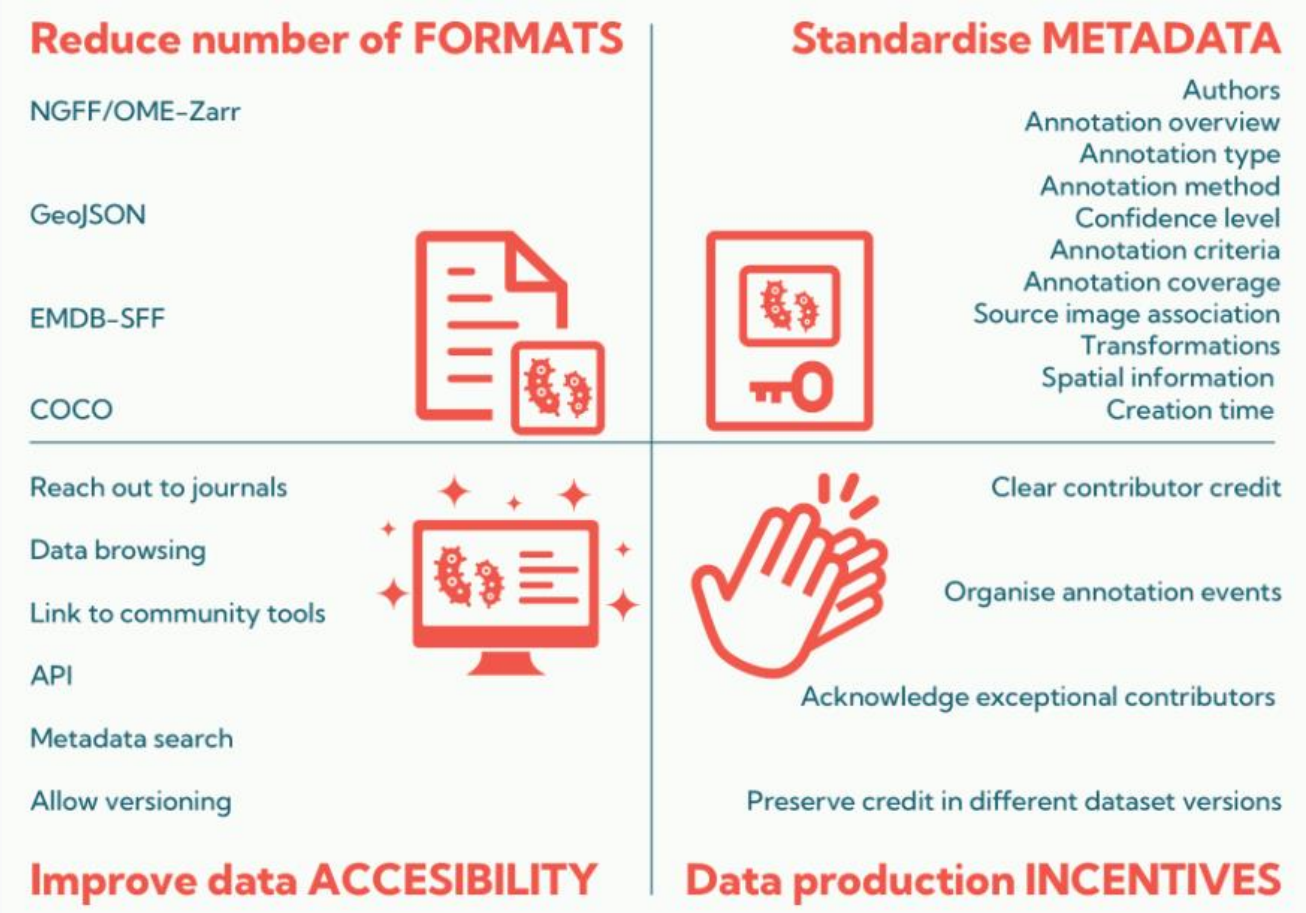
FAIR tips:

- Collect metadata for each of the REMBI categories where applicable
- Write a README-file in highest-level project folder
- Record experimental metadata in an Electronic Lab Notebook

*Sarkans, U., Chiu, W., Collinson, L. et al. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. Nat Methods **18**, 1418–1422 (2021). https://doi.org/10.1038/s41592-021-01166-8*

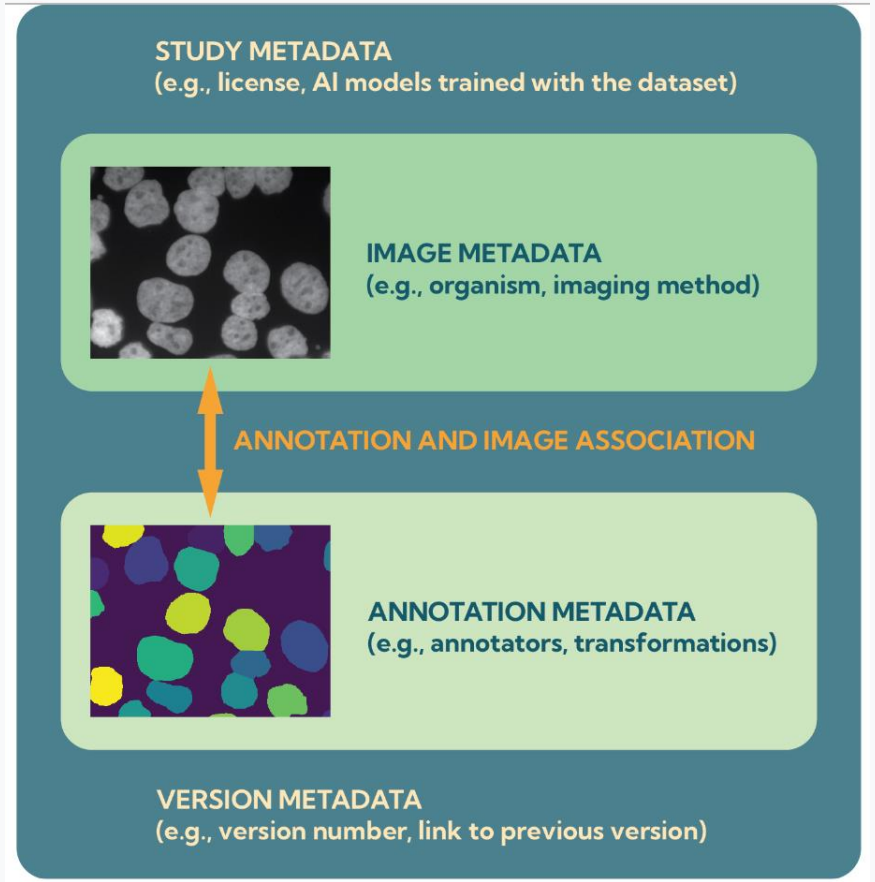
MIFA – output from AI4Life

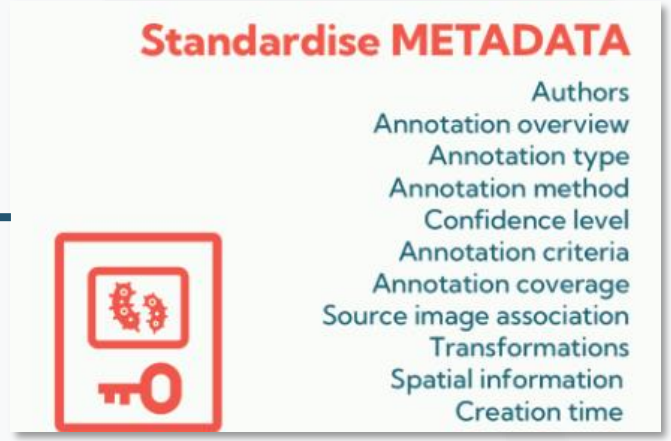
MIFA guidelines are a result of a workshop where community experts came together to develop guidelines, in order to improve the reuse of bioimages and annotations for AI applications.



Zulueta-Coarasa, Teresa et al. "MIFA: Metadata, Incentives, Formats, and Accessibility guidelines to improve the reuse of AI datasets for bioimage analysis." (2023).

MIFA – adopted by BioImage Archive





Zulueta-Coarasa, Teresa et al. "MIFA: Metadata, Incentives, Formats, and Accessibility guidelines to improve the reuse of AI datasets for bioimage analysis." (2023).

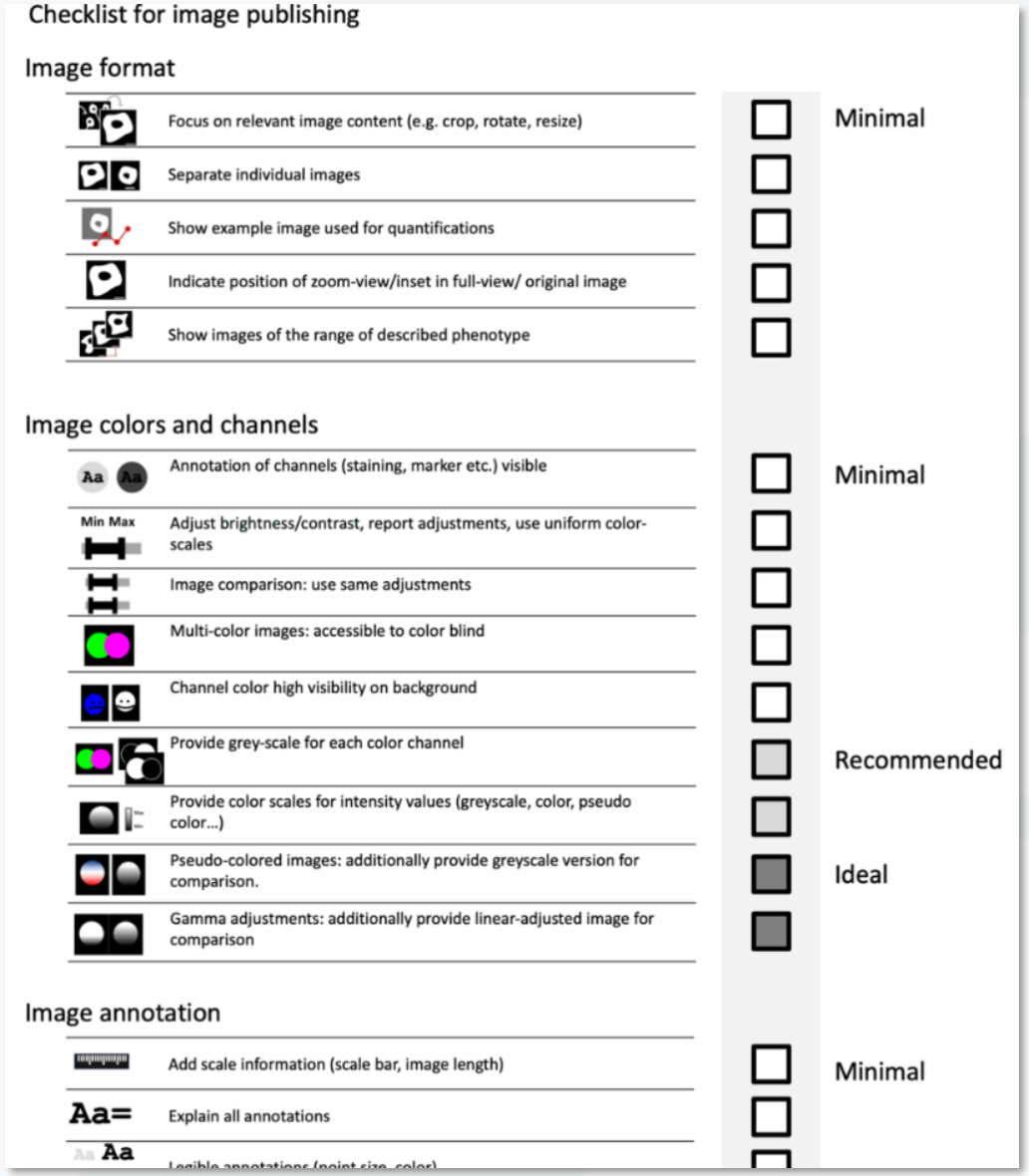
Different metadata standards for bioimaging

- REMBI
- OME data model
- 4DN-BINA-OME-QUAREP
- MITI (Minimum Information for Highly Multiplexed Tissue Images)
- DICOM (Digital Imaging and Communications in Medicine)
- MIHCSME (Minimum Information for High Content Screening Microscopy Experiments)
- …

FAIR tip: Adhere to metadata recommendations/ requirements of repository

Community-developed metadata checklist for image publishing

Schmied, Christopher et al. "Community-developed checklists for publishing images and image analyses." ArXiv arXiv:2302.07005v2. 14 Sep. 2023 Preprint.



Checklist for image publishing

Image format

Focus on relevant image content (e.g. crop, rotate, resize)	☐	Minimal
Separate individual images	☐	
Show example image used for quantifications	☐	
Indicate position of zoom-view/inset in full-view/ original image	☐	
Show images of the range of described phenotype	☐	

Image colors and channels

Annotation of channels (staining, marker etc.) visible	☐	Minimal
Adjust brightness/contrast, report adjustments, use uniform color-scales	☐	
Image comparison: use same adjustments	☐	
Multi-color images: accessible to color blind	☐	
Channel color high visibility on background	☐	
Provide grey-scale for each color channel	☐	Recommended
Provide color scales for intensity values (greyscale, color, pseudo color…)	☐	
Pseudo-colored images: additionally provide greyscale version for comparison.	☐	Ideal
Gamma adjustments: additionally provide linear-adjusted image for comparison	☐	

Image annotation

Add scale information (scale bar, image length)	☐	Minimal
Explain all annotations	☐	
Legible annotations (point size, color)	☐	

Data Management

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

F A I R

FINDABLE ACCESSIBLE INTEROPERABLE REUSABLE

Data Management

proprietary

.lif

.czi

and > 150 more

FAIR tip: Convert images to open formats where possible

open

OME.tiff

https://www.openmicroscopy.org/bio-formats/

Software tool with application interface for programs like ImageJ, CellProfiler and Icy, OMERO and MATLAB.

Open and standard file formats

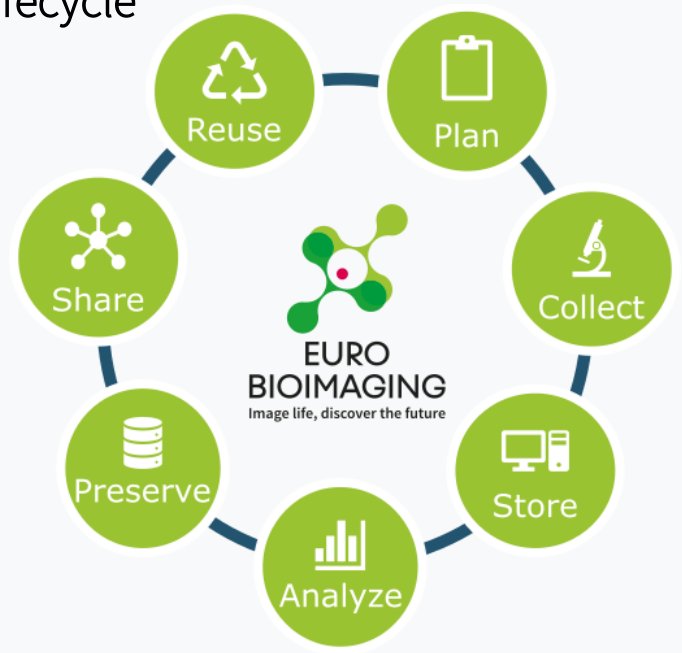
Metadata use controlled vocabulary

Include references to other data

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

Data Management Plans

Research data lifecycle



Describes the journey of data through the data lifecycle

Ideally created in project planning phase, updated regularly throughout and contains all relevant information on the data upon project completion

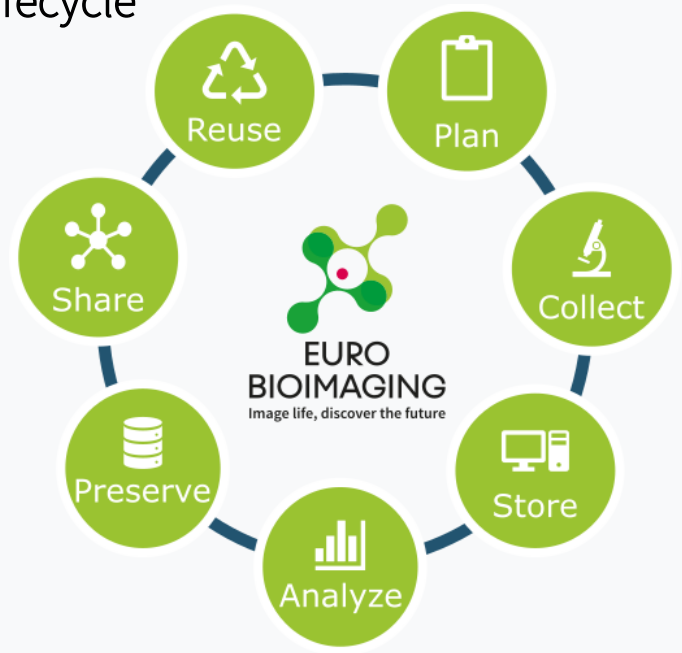
Increasingly required by funders and instituts

Sections of a DMP:

- General project description
- Roles & responsibilites
- Data used & collected
- Data & metadata documentation
- Data storage & preservation
- Data access & sharing
- Ethics & legal issues

Data Management Plans

Research data lifecycle



Describes the journey of data through the data lifecycle

Ideally created in project planning phase, updated regularly throughout and contains all relevant information on the data upon project completion

Euro-BioImaging DMP Template

10.5281/zenodo.11473802

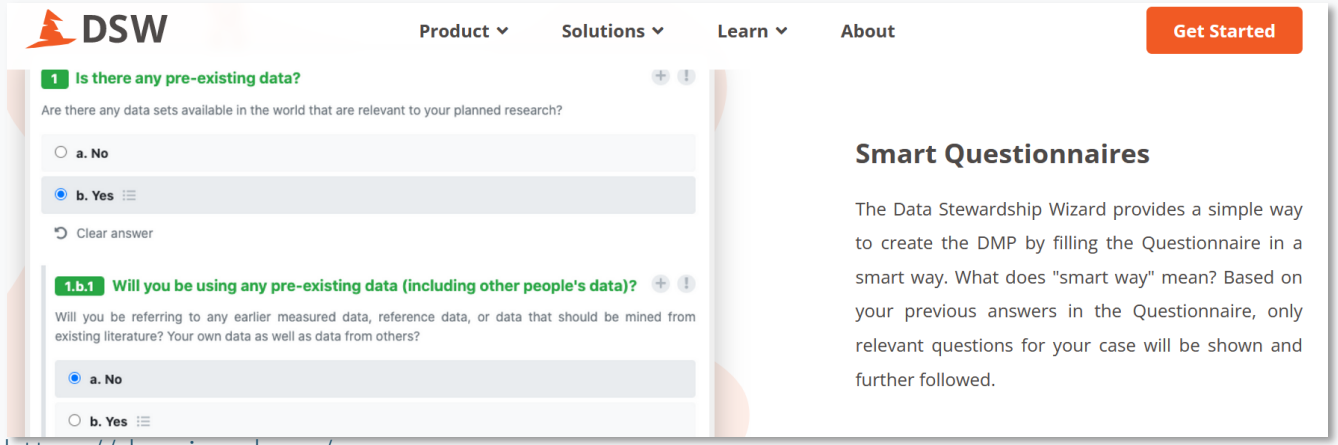
Increasingly required by funders and instituts

Sections of a DMP:

- General project description
- Roles & responsibilites
- Data used & collected
- Data & metadata documentation
- Data storage & preservation
- Data access & sharing
- Ethics & legal issues

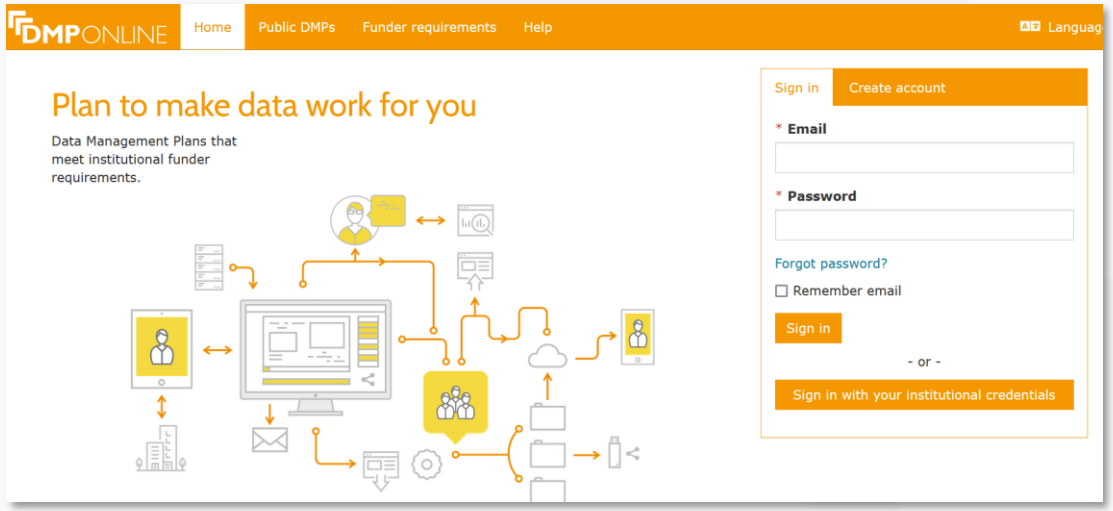
Tools to facilitate writing DMPs

Data Stewardship Wizard



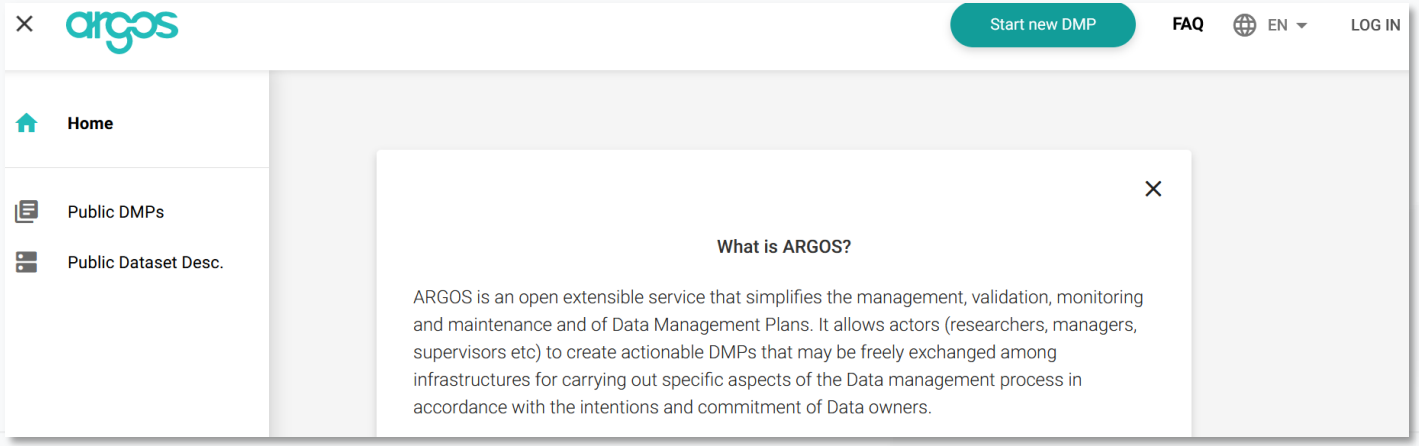
https://ds-wizard.org/

DMP Online



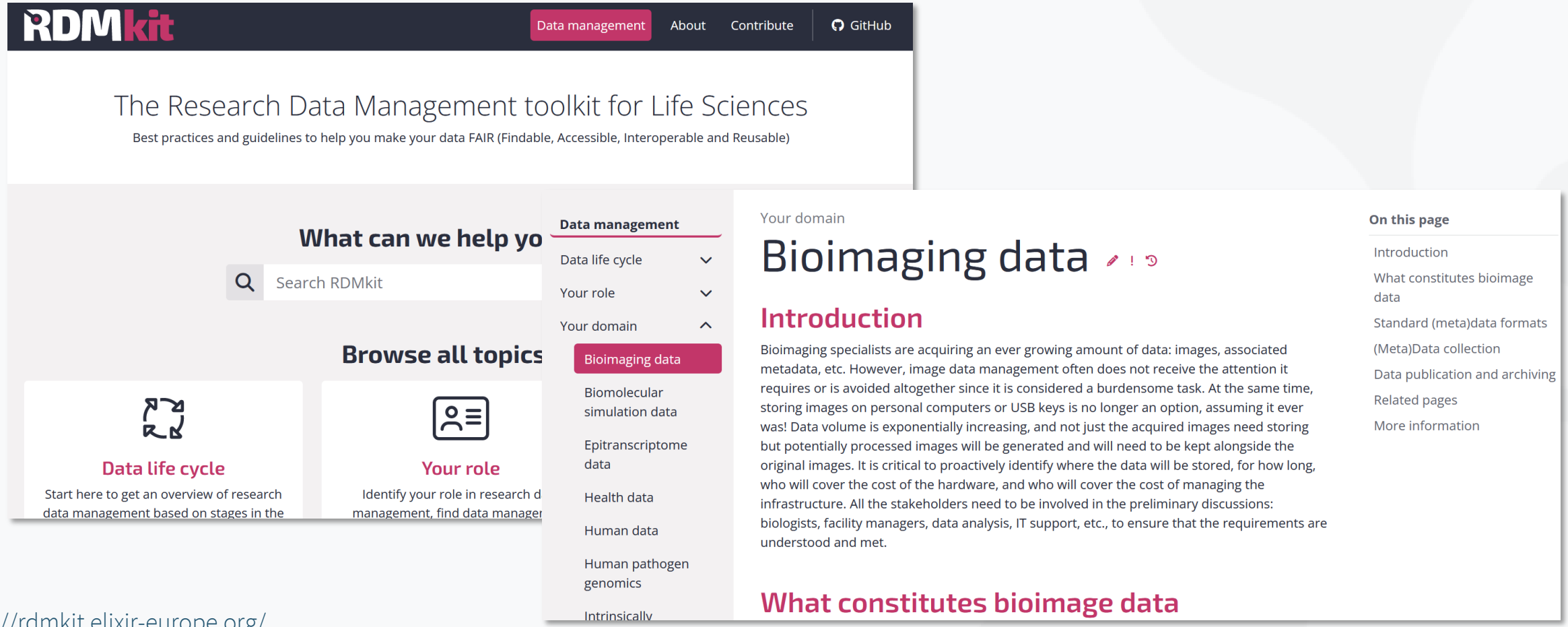
https://dmponline.dcc.ac.uk/

Argos



https://argos.openaire.eu/splash/

RDMkit
User-oriented guide to the FAIR RDM practices in life sciences



https://rdmkit.elixir-europe.org/

Repository

Metadata

F A I R

FINDABLE

ACCESSIBLE

INTEROPERABLE

REUSABLE

Data Management

- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

- Standardized and open communication protocol for data retrieval
- Authentication and authorization possible
- Metadata are longer accessible than data

- Open and standard file formats
- Metadata use controlled vocabulary
- Include references to other data

- Clear and accessible license
- Metadata according to community standards
- Well-described provenance information (documentation of reproducibility)

F A I R

FINDABLE ACCESSIBLE INTEROPERABLE REUSABLE

fairdata@eurobioimaging.eu