

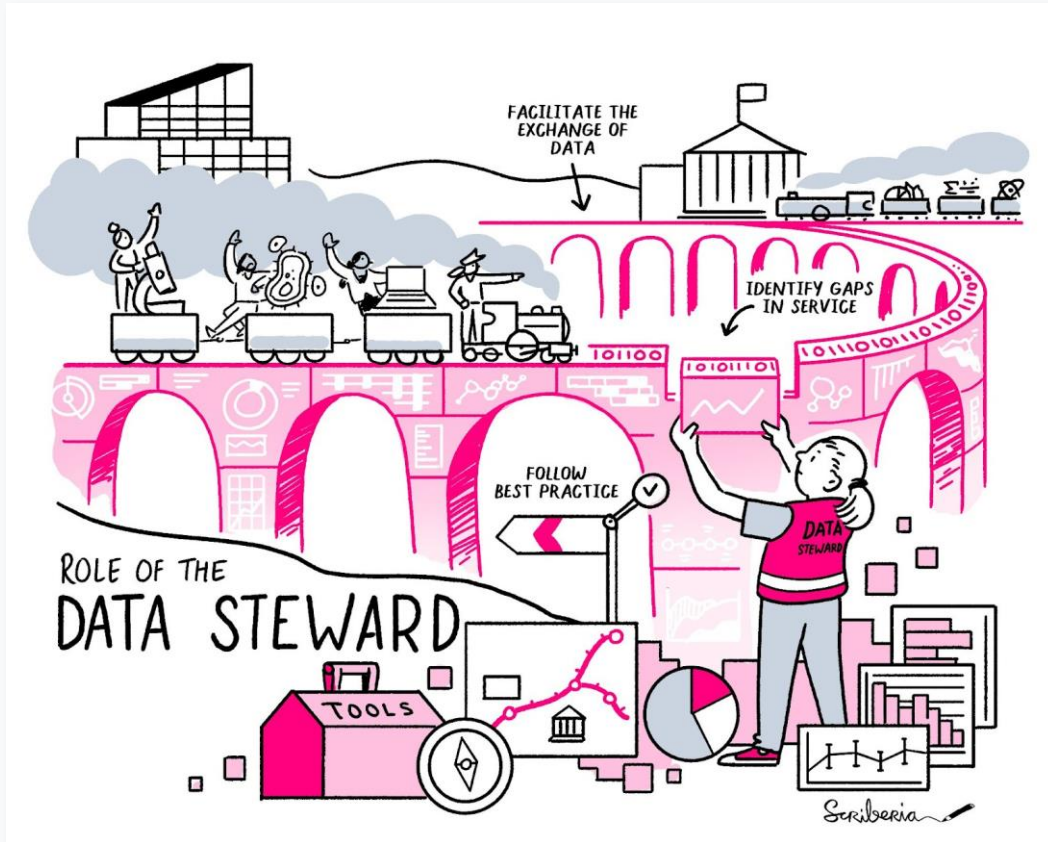
## *Guide to FAIR Biolmage Data*



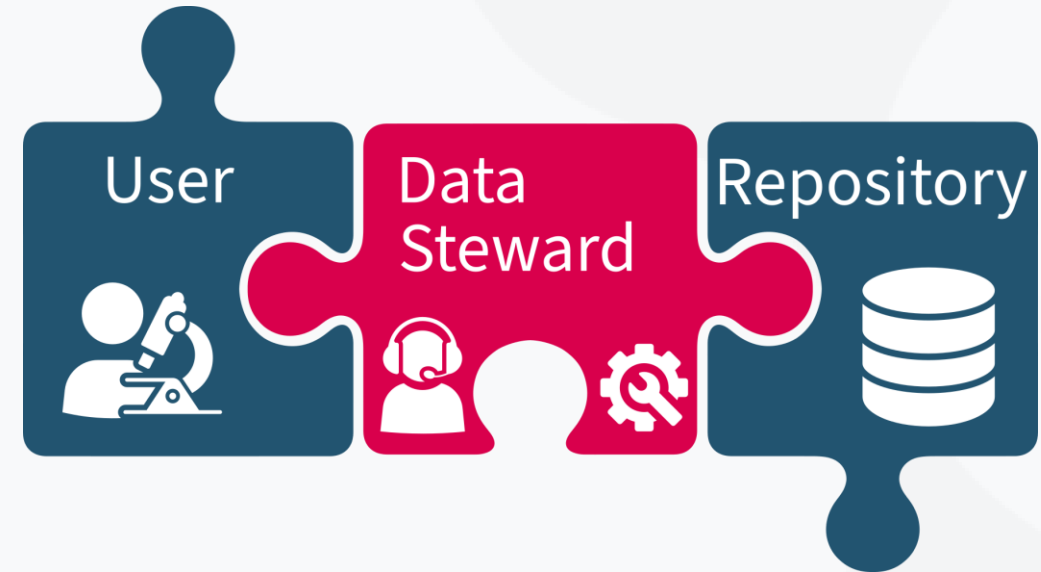
Isabel Kemmer

FAIR Image Data Steward  
Euro-Biolmaging Bio-Hub

# Image Data Stewardship



This illustration is created by Scriberia with The Turing Way community, used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)







[fairdata@eurobioimaging.eu](mailto: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<sup>1</sup>  
<sup>1</sup> 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)
-  Qualifying criteria for journals and funders

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

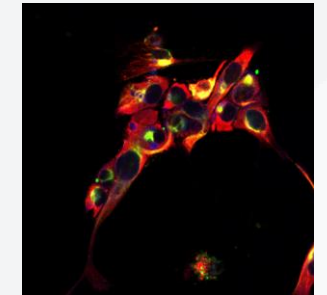
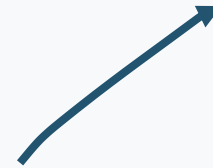
# FAIR across disciplines

Structural biology

Bioimaging



FAIR is a spectrum – start somewhere!



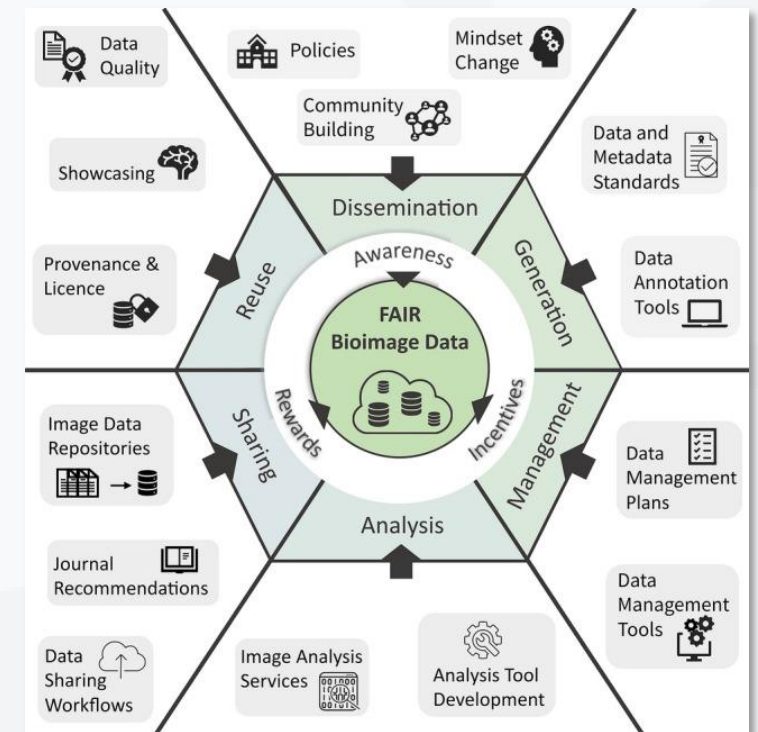
Life Science-communities are at different stages of FAIR

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



## FINDABLE

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

# A



## ACCESSIBLE

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

# I



## INTEROPERABLE

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

# R



## REUSABLE

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

# Repository

## F



FINDABLE

## A



ACCESSIBLE

## I



INTEROPERABLE

## R



REUSABLE

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)

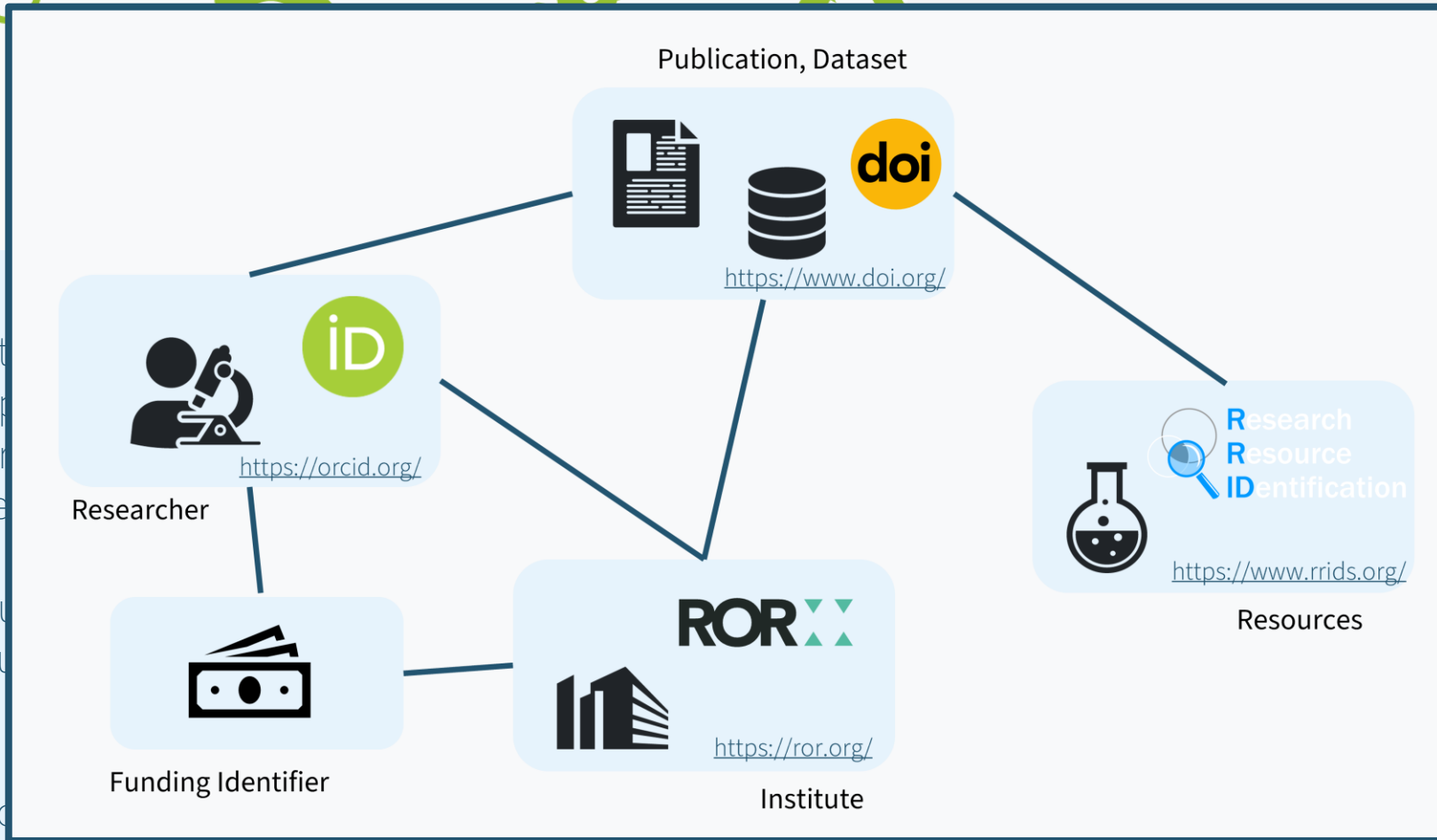
# F A I R

Repository

FIN

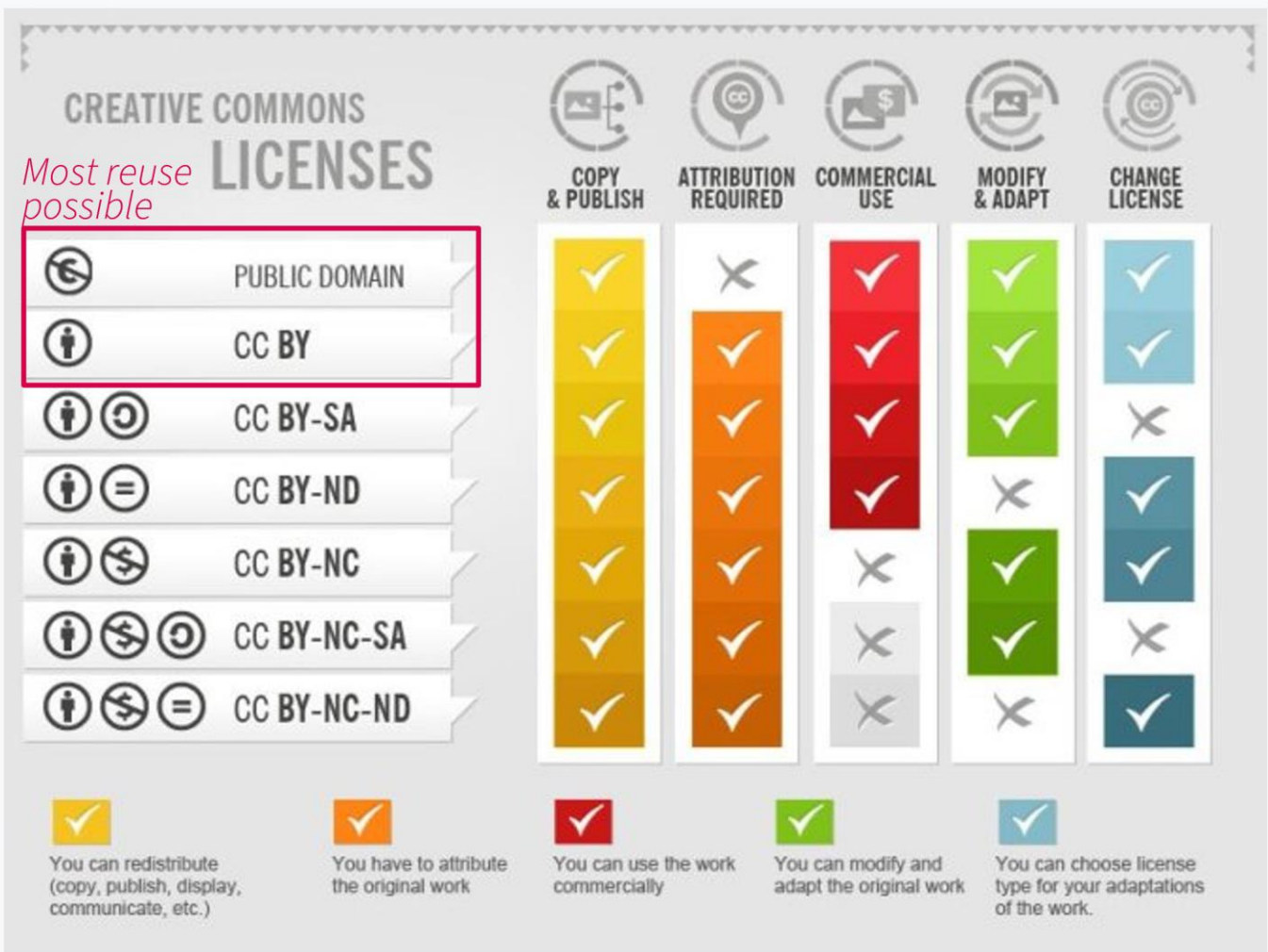
- Data are registered in a searchable resource
- Machine-readable, persistent identifiers
- Described by rich metadata

- St
- op
- pr
- re
- Au
- at
- M
- ac



Re

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

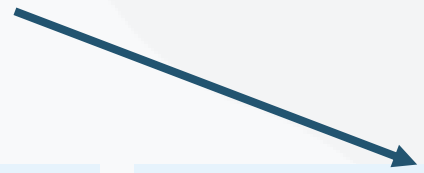


Creative Commons licenses by Foter (CC-BY-SA)

R



REUSABLE



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

- Data search
- Metadata
- Description

Standard file  
 Vocabulary  
 Access to



# Repository

## F



FINDABLE

## A



ACCESSIBLE

## I



INTEROPERABLE

## R



REUSABLE

- 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)

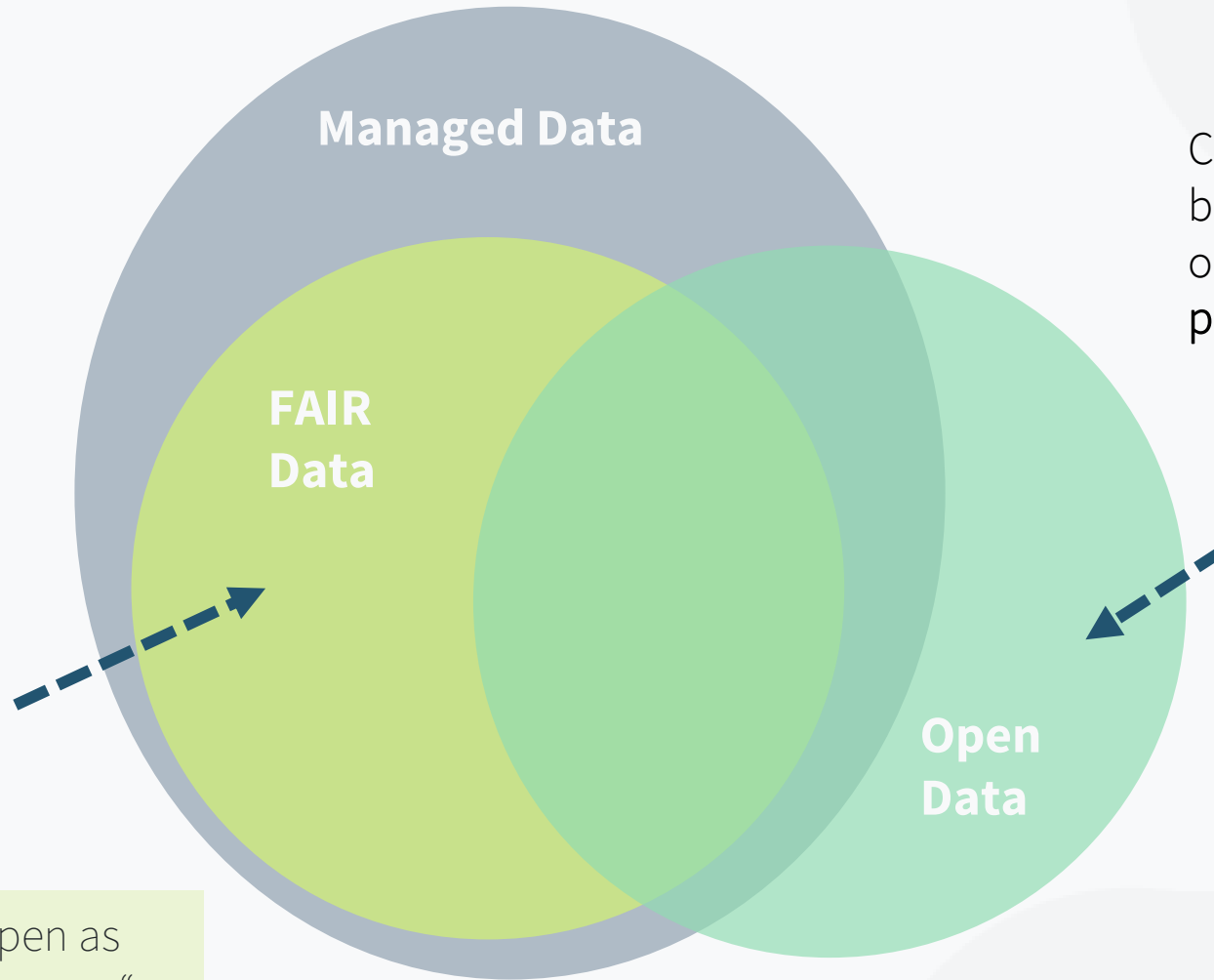
# FAIR data $\neq$ Open data

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

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.*

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



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

Open Data are not necessarily FAIR

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

# Different types of repositories

## General purpose repositories

Examples:



## Institutional repositories

Example:



## Domain-specific repositories

### Open repositories

Examples:



### Restricted-access repositories

Example:



### Metadata repositories

Example:





# BioImage Archive

Search BioImage Archive

Examples: [brain](#), [capsid](#)

[Home](#)

[Browse](#)

[Submit](#)

[Galleries](#)

[Help](#) ▾

[Metadata Help](#) ▾

[Policies](#) ▾

[About us](#) ▾

The BioImage Archive is a free, publicly available online resource which stores and distributes biological images. It accepts submissions of data from **any imaging modality**, as long as the data are either **associated with a peer-reviewed publication**, or of **value beyond a single experiment**.

You can submit your data on our [submission page](#). All data submitted to the BioImage Archive must be consented for a public release and the submitter self certifies that they have the rights to submit such data to a public archive. You can find more about our policies [here](#).

The BioImage Archive also provides data archiving services to the broader bioimaging database community including added-value bioimaging data resources such as EMPIAR and IDR. Submission to related community resources may be more appropriate for some data types. You can find out more about the BioImage Archive's scope, and where your data should best be archived [here](#) and [here](#). The BioImage Archive **cannot accept patient-identifiable medical data**, such as that derived from clinical imaging.

The BioImage Archive supports [FAIR Sharing](#) and implements the [REMBI guidelines](#) to enable FAIR data.

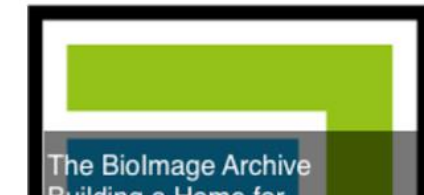
## Further information

ONLINE TUTORIAL

### BioImage Archive

Quick tour

[The BioImage Archive Online Tutorial](#)



- 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](#)



EMPIAR, the Electron Microscopy Public Image Archive, is a public resource for raw images underpinning 3D cryo-EM maps and tomograms (themselves archived in EMDB). EMPIAR also accommodates 3D datasets obtained with volume EM techniques and soft and hard X-ray tomography. More ...

As of 2024-04-08, EMPIAR contains 1631 entries, taking up 4.05 PB of storage.

Browse and download EMPIAR datasets using the table below. The EMDB search system also supports searches of all EMPIAR entries (video tutorial).

### EMPIAR News

- 6 January 2023: We have published a reference paper for EMPIAR in the 2023 NAR Database issue. Please cite this whenever you refer to EMPIAR: [ludin et al., NAR 51, D1503-D1511 \(2023\)](#).
- 22 December 2022: EMPIAR proudly announces the first public release of SPW, the Sample Preparation Widget. It has been developed in collaboration with and is intended for use by the volume EM community. Read more [here](#).



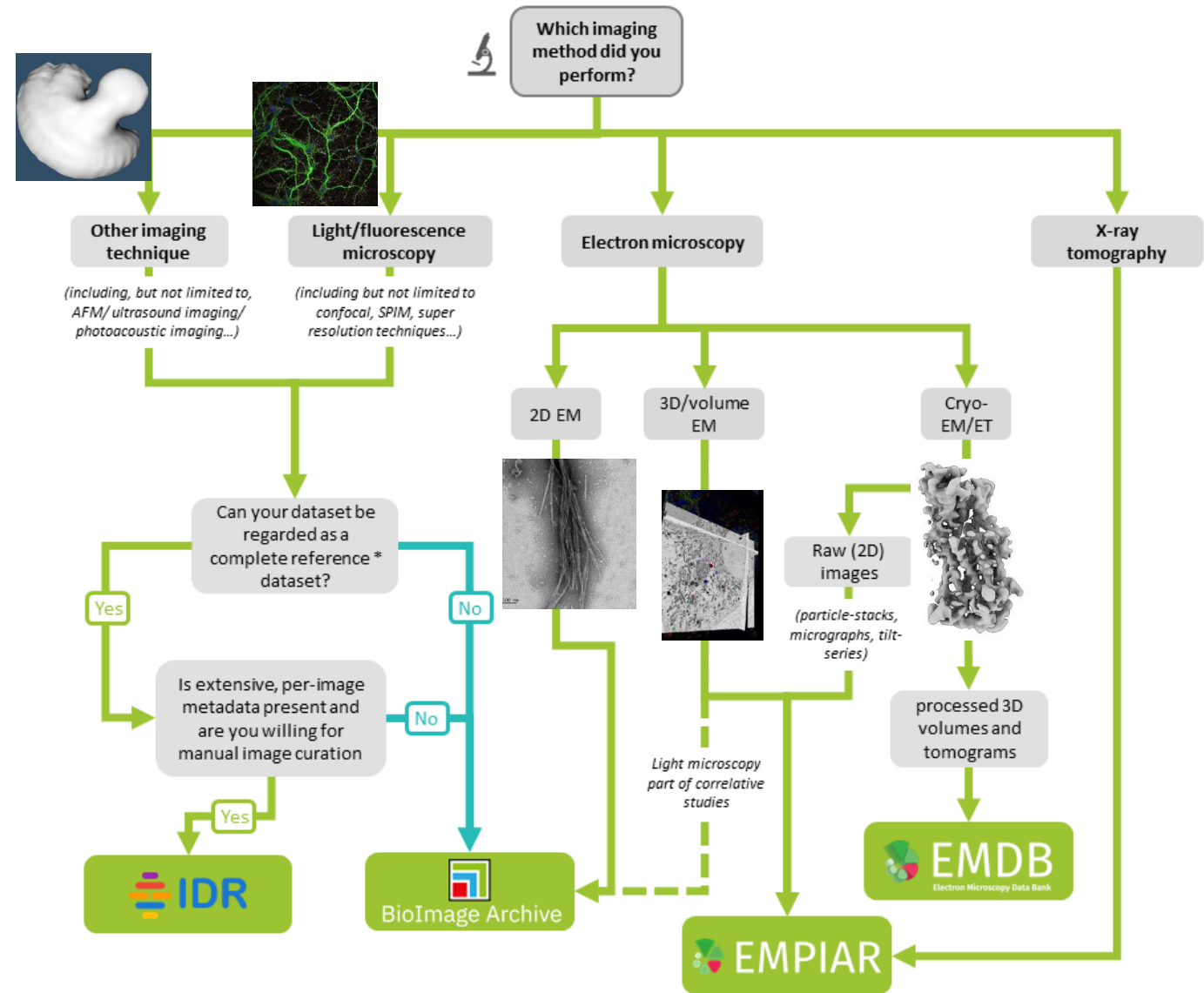
### Quick links

- Browse EMPIAR
- Sample-Preparation Widget
- Volume Browser
- Claim entries to your ORCID
- Talks and Tutorials
- EMPIAR Quick tour
- Publications
- Re-use case study

Links:

- [EMPIAR](#)
- [Deposition Manual](#)
- [Sample preparation Widget](#)
- [API](#)

# Open BioImage Repositories



\* - **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****FINDABLE****A****ACCESSIBLE****I****INTEROPERABLE****R****REUSABLE**

Metadata

- 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)

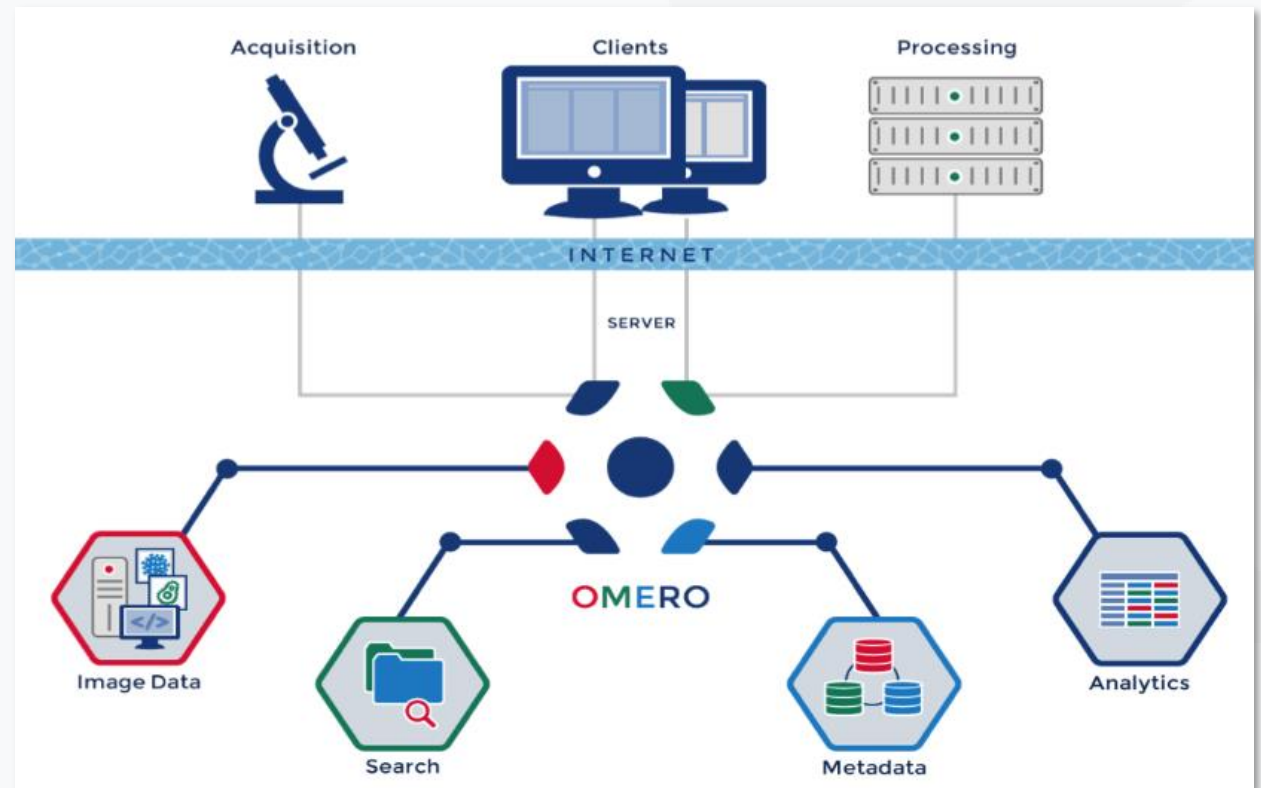


# Image management software - OMERO

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/s41592-021-01290-5>

MethodsJ2  
*Plugin for Fiji*



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

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

**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



MDEmic  
*MetaData Editor for microscopy*

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.**

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

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

**F****FINDABLE****A****ACCESSIBLE****I****INTEROPERABLE****R****REUSABLE**

Metadata

- 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)

# The problem with names

Different words for the same concepts

→ difficult search



The same words for different concepts

→ ambiguous and confusing

## 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)  
DSLIM  
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



*palm*



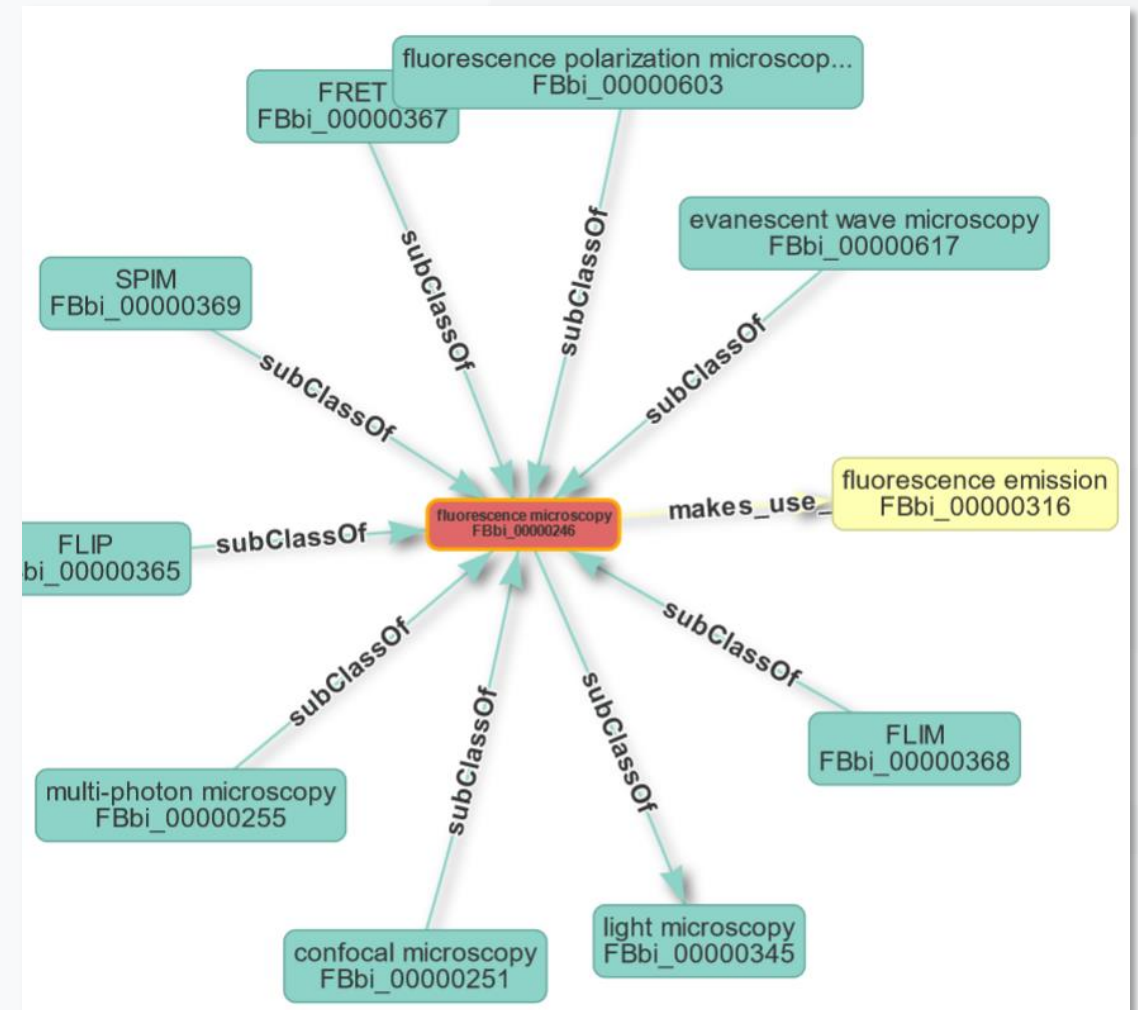
*palm*

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**



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	<a href="https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/">https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/</a>
Gene	Ensembl	<a href="https://www.ensembl.org/index.html">https://www.ensembl.org/index.html</a>
Protein	Uniprot	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>
Compound/ Drug	PubChem/ ChEMBL	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>
Imaging Method	FBBI /EDAM-BioImaging	<a href="https://www.ebi.ac.uk/ols4/ontologies/fbbi">https://www.ebi.ac.uk/ols4/ontologies/fbbi</a>
Study Type	Experimental Factor Ontology	<a href="https://www.ebi.ac.uk/efo/">https://www.ebi.ac.uk/efo/</a>
Phenotypes	Human Phenotype Ontology	<a href="https://hpo.jax.org/app/">https://hpo.jax.org/app/</a>
...	...	...

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****FINDABLE****A****ACCESSIBLE****I****INTEROPERABLE****R****REUSABLE**

Metadata

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

# REcommended Metadata for Biological Images

8 categories of metadata with sub keywords

Partly stored in the image itself



## FAIR tips:

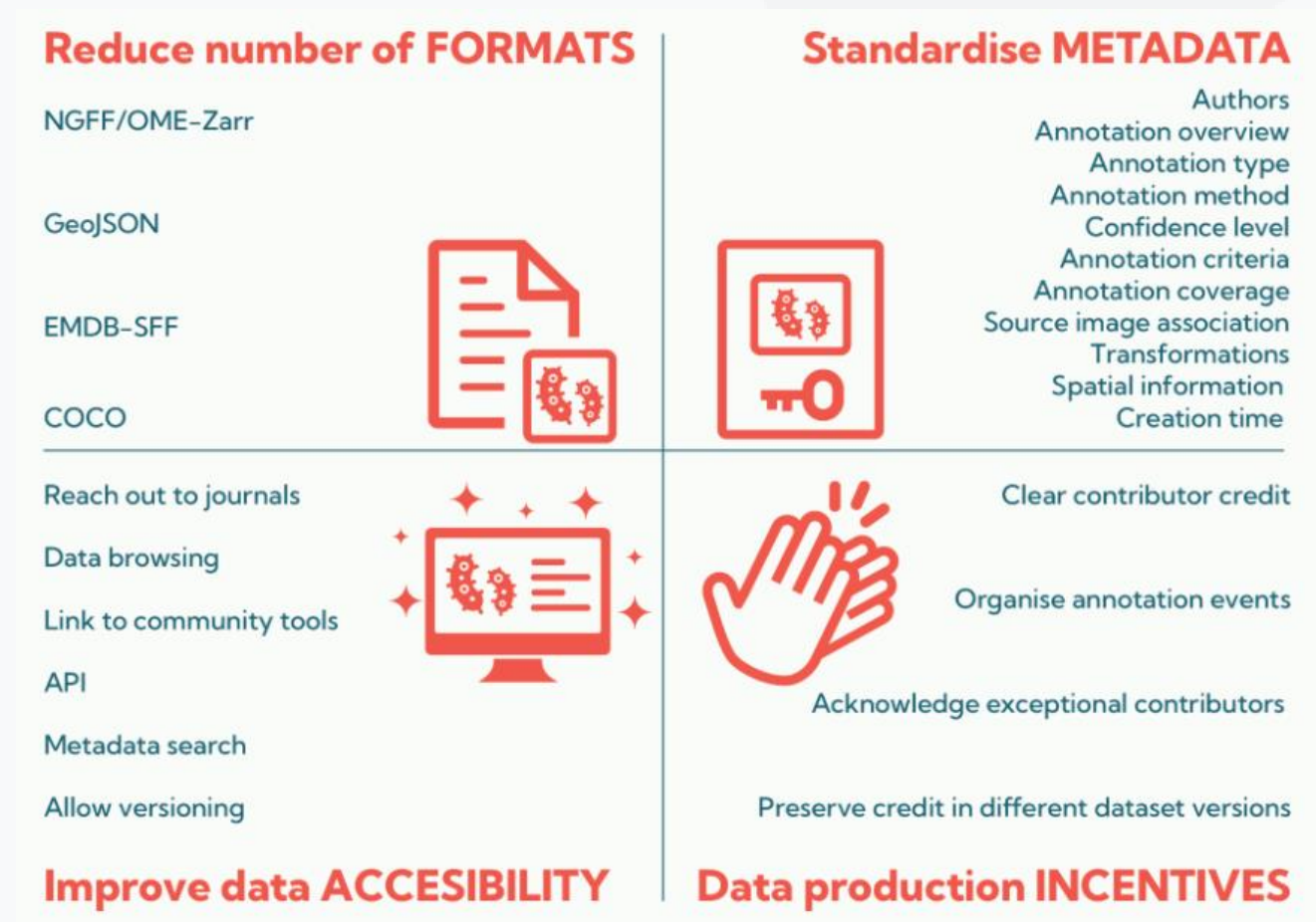
- Collect metadata for each of the REMBI categories where applicable
- Write a README-file in highest-level project folder
- Record experimental metadata in an Electronic Lab Notebook

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>



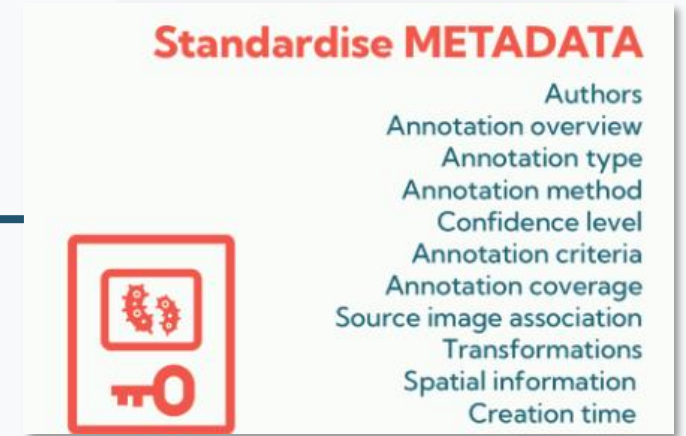
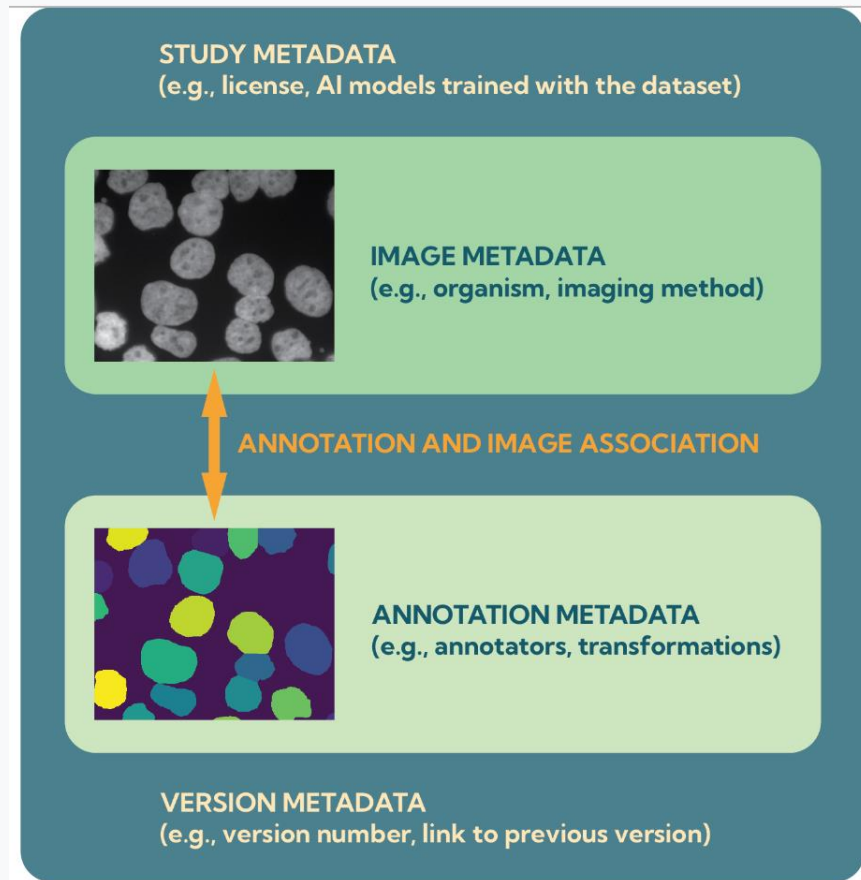
# MIFA – output from AI4Life

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)
- ...

Community-developed metadata checklist for image publishing








**FAIR tip:** Adhere to metadata recommendations/requirements of repository










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)	<input type="checkbox"/>	Minimal
	Separate individual images	<input type="checkbox"/>	
	Show example image used for quantifications	<input type="checkbox"/>	
	Indicate position of zoom-view/inset in full-view/ original image	<input type="checkbox"/>	
	Show images of the range of described phenotype	<input type="checkbox"/>	

### Image colors and channels

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

### Image annotation

	Add scale information (scale bar, image length)	<input type="checkbox"/>	Minimal
	Explain all annotations	<input type="checkbox"/>	
	Legible annotations (point size, color)	<input type="checkbox"/>	

# F



## FINDABLE

# A



## ACCESSIBLE

# I



## INTEROPERABLE

# R



## 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**



FINDABLE

**A**



ACCESSIBLE

**I**



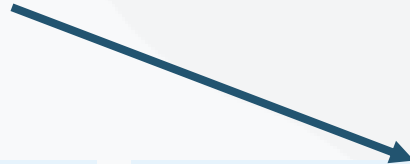
INTEROPERABLE

**R**



REUSABLE

Data Management



proprietary



and > 150 more

FAIR tip: Convert images to open formats where possible



open



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



**BIO-FORMATS**  
The solution for reading proprietary microscopy image data and metadata

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

Open and standard file formats

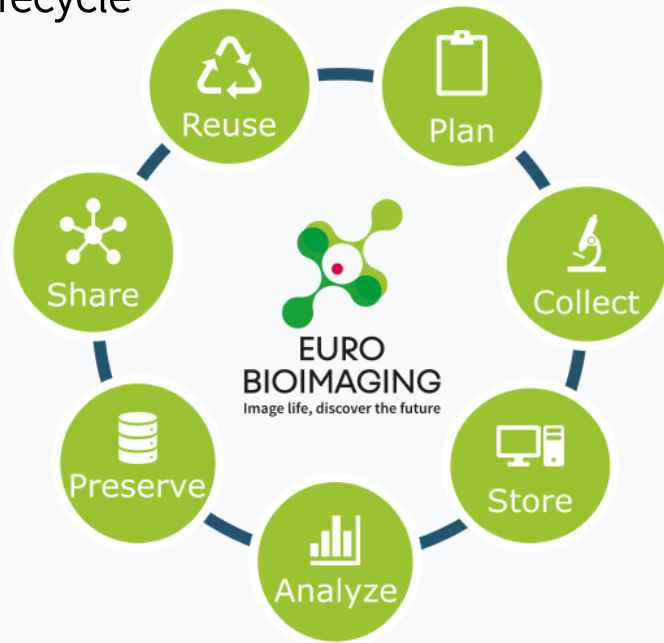
Metadata use controlled vocabulary

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



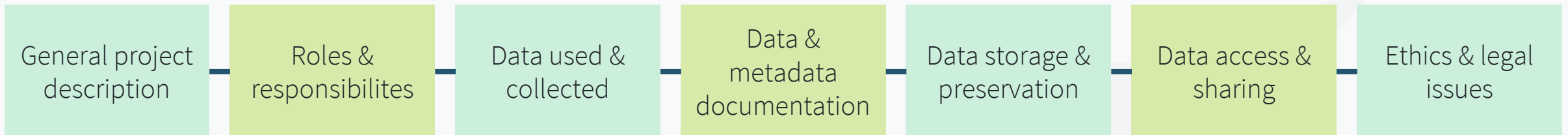
Describes the journey of data through the data lifecycle

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



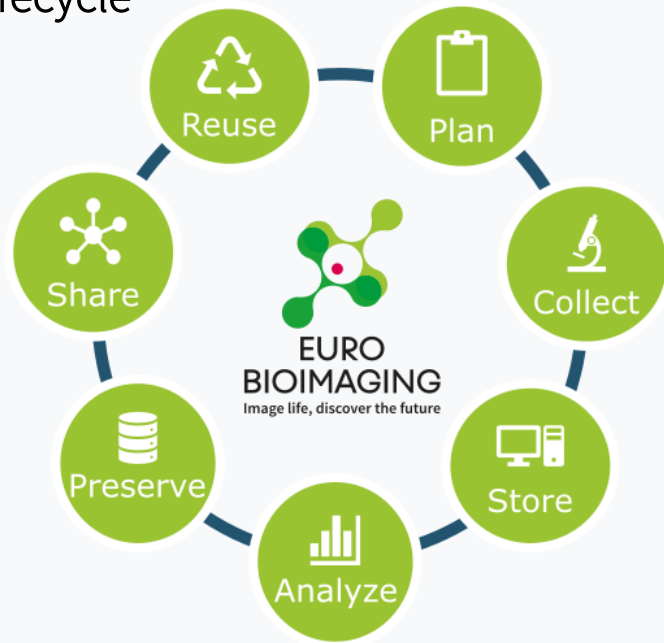
Increasingly required by funders and institutis

Sections of a DMP:



# Data Management Plans


Research data lifecycle



Describes the journey of data through the data lifecycle

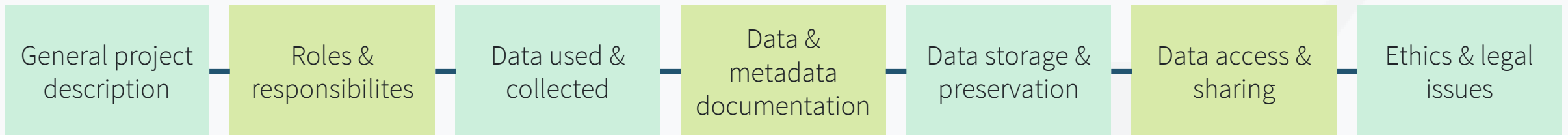
Ideally created in project planning phase, updated regularly throughout and contains all relevant information on the data upon project completion



 [10.5281/zenodo.11473802](https://zenodo.org/doi/10.5281/zenodo.11473802)  
Euro-BiolMaging  
DMP Template

Increasingly required by funders and institutis

Sections of a DMP:



# Tools to facilitate writing DMPs

DMP Online

## Data Stewardship Wizard

**DSW** Product Solutions Learn About [Get Started](#)

**1 Is there any pre-existing data?**  
Are there any data sets available in the world that are relevant to your planned research?

a. No  
 b. Yes

Clear answer

**1.b.1 Will you be using any pre-existing data (including other people's data)?**  
Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

a. No  
 b. Yes

### Smart Questionnaires

The Data Stewardship Wizard provides a simple way to create the DMP by filling the Questionnaire in a smart way. What does "smart way" mean? Based on your previous answers in the Questionnaire, only relevant questions for your case will be shown and further followed.

<https://ds-wizard.org/>

**DMP ONLINE** Home Public DMPs Funder requirements Help Language

## Plan to make data work for you

Data Management Plans that meet institutional funder requirements.

Sign in Create account

\* Email  
[Input field]

\* Password  
[Input field]

Forgot password?  
 Remember email

Sign in

- or -

Sign in with your institutional credentials

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

## Argos

**argos** Start new DMP FAQ EN LOG IN

- Home
- Public DMPs
- Public Dataset Desc.

### 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.

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



# RDMkit

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

The screenshot displays the RDMkit website interface. At the top, the RDMkit logo is on the left, and navigation links for 'Data management', 'About', 'Contribute', and 'GitHub' are on the right. The main heading reads 'The Research Data Management toolkit for Life Sciences' with a subtitle 'Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)'. Below this is a search bar and a 'Browse all topics' section with two cards: 'Data life cycle' and 'Your role'. A sidebar menu on the left lists 'Data management', 'Bioimaging data', 'Biomolecular simulation data', 'Epitranscriptome data', 'Health data', 'Human data', 'Human pathogen genomics', and 'Intrinsically'. The main content area shows the 'Bioimaging data' page with an 'Introduction' section and a 'What constitutes bioimage data' section. A right-hand sidebar titled 'On this page' lists links for 'Introduction', 'What constitutes bioimage data', 'Standard (meta)data formats (Meta)Data collection', 'Data publication and archiving', 'Related pages', and 'More information'.

**RDMkit** Data management About Contribute GitHub

### The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

#### What can we help you

Search RDMkit

#### Browse all topics

- Data life cycle**  
Start here to get an overview of research data management based on stages in the
- Your role**  
Identify your role in research data management, find data manager

#### Data management

- Data life cycle
- Your role
- Your domain

#### Bioimaging data

- Biomolecular simulation data
- Epitranscriptome data
- Health data
- Human data
- Human pathogen genomics
- Intrinsically

Your domain

## Bioimaging data

### Introduction

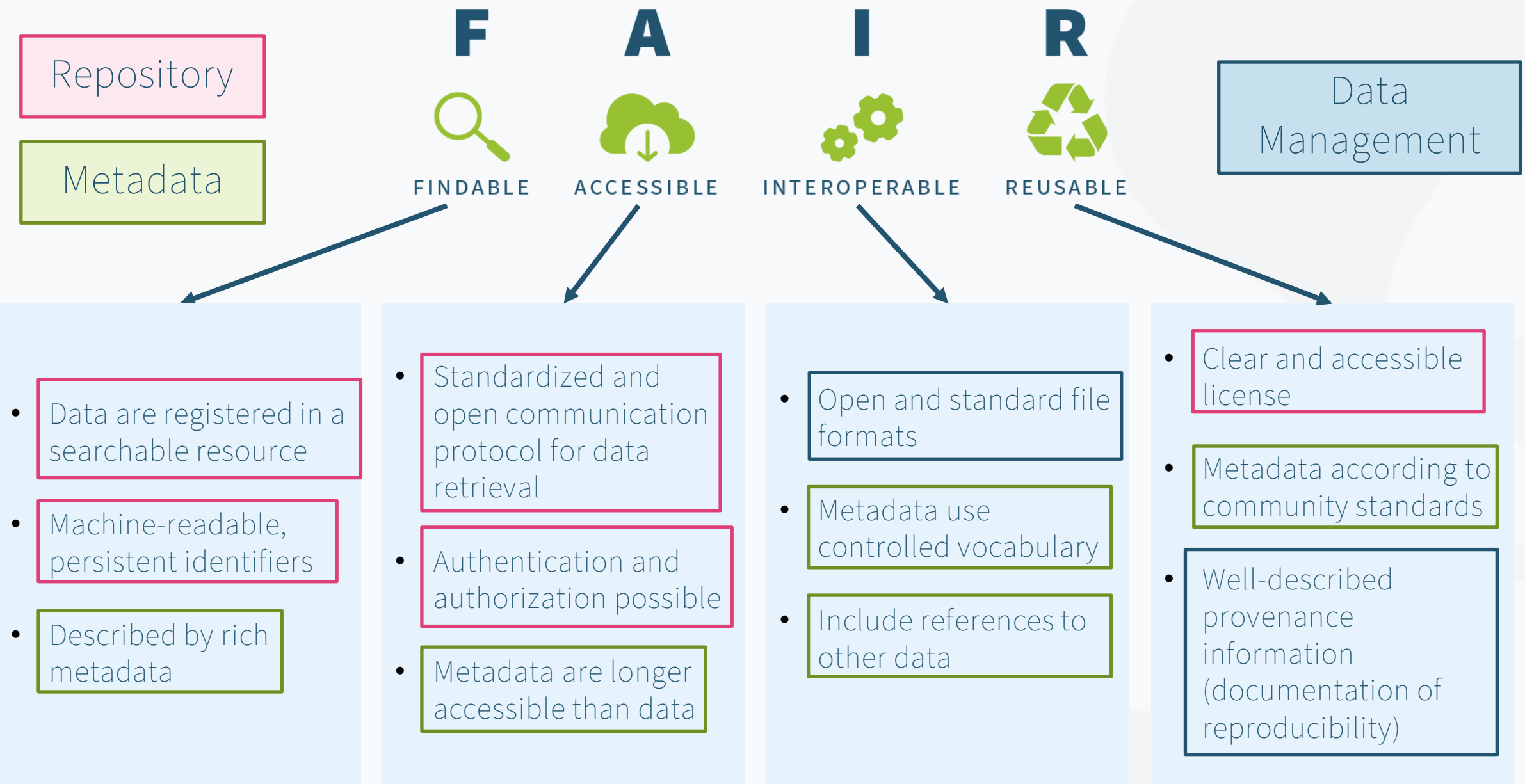
Bioimaging specialists are acquiring an ever growing amount of data: images, associated metadata, etc. However, image data management often does not receive the attention it requires or is avoided altogether since it is considered a burdensome task. At the same time, storing images on personal computers or USB keys is no longer an option, assuming it ever was! Data volume is exponentially increasing, and not just the acquired images need storing but potentially processed images will be generated and will need to be kept alongside the original images. It is critical to proactively identify where the data will be stored, for how long, who will cover the cost of the hardware, and who will cover the cost of managing the infrastructure. All the stakeholders need to be involved in the preliminary discussions: biologists, facility managers, data analysis, IT support, etc., to ensure that the requirements are understood and met.

### What constitutes bioimage data

#### On this page

- Introduction
- What constitutes bioimage data
- Standard (meta)data formats (Meta)Data collection
- Data publication and archiving
- Related pages
- More information

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





**F**



FINDABLE

**A**



ACCESSIBLE

**I**



INTEROPERABLE

**R**



REUSABLE



*fairdata@eurobioimaging.eu*

