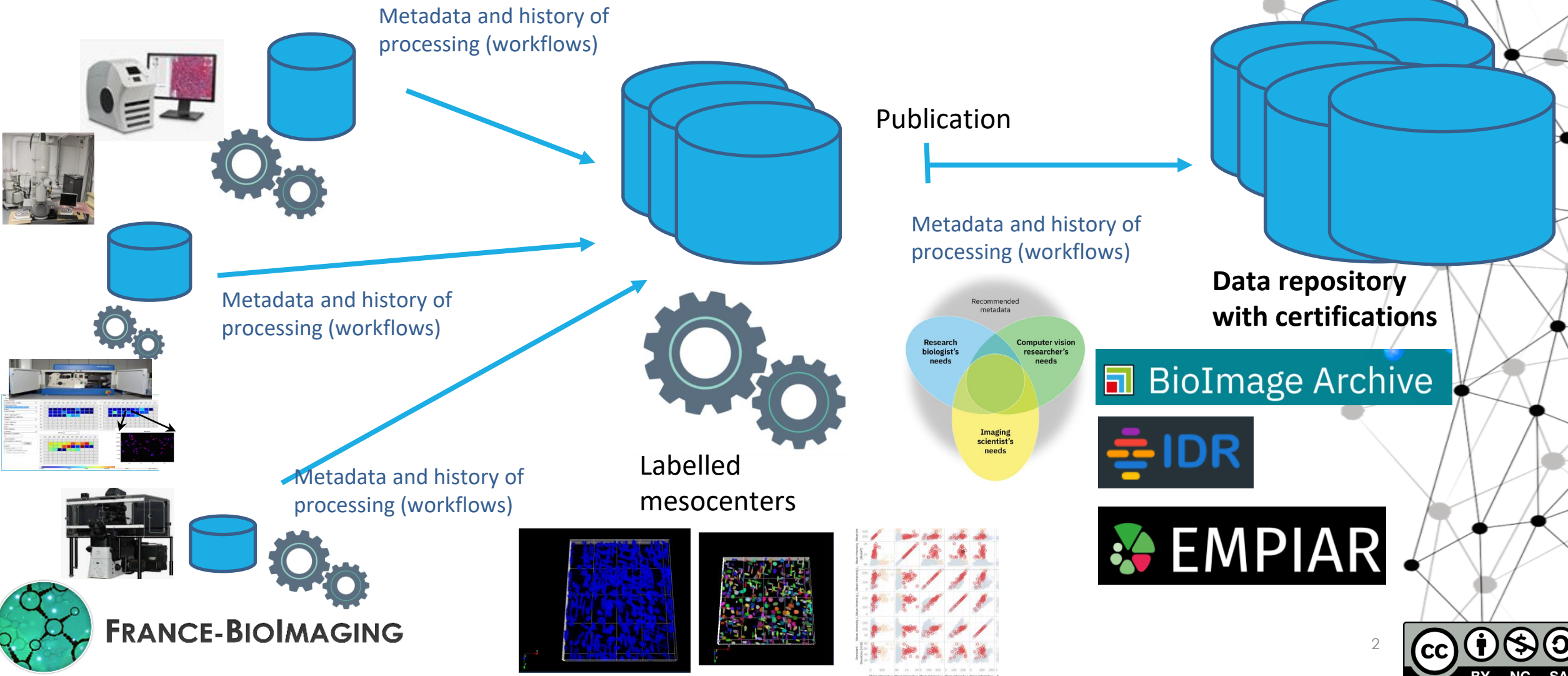

Décrire les données de microscopies en biologie: vers un consensus communautaire ou comment créer un standard de metadonnées

How to describe bioimage data? Toward a community consensus, or how to create a metadata standard?

Perrine Paul-Gilloteaux
UAR BioCore MicroPICell
FBI.data, Bretagne Loire Node of France Biolmaging

France BioImaging DMP in a nutshell



Metadata and history of processing (workflows)

Metadata and history of processing (workflows)

Metadata and history of processing (workflows)

Publication

Metadata and history of processing (workflows)

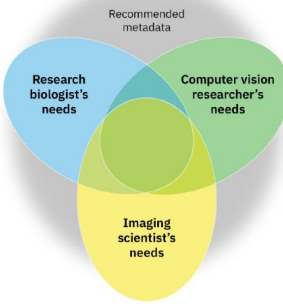
Data repository with certifications

BioImage Archive

IDR

EMPIAR

Labelled mesocenters



FRANCE-BIOIMAGING



Why Metadata?

They are everywhere in FAIR principles...

Findable

- F1 Data are assigned a globally unique and persistent identifier
- F2 Data are described with rich metadata (R1)
- F3 Metadata clearly and explicitly include the identifier of the data they describe
- F4 Data are registered or indexed in a searchable resource

Accessible

- A1 Data are retrievable by their identifier using a standardised communications protocol
- A2 Metadata are accessible, even when the data are no longer available

Interoperable

- I1 Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2 Data use vocabularies that follow FAIR principles
- I3 Data include qualified references to other (meta)data

Reusable

- R1 Data are richly described with a plurality of accurate and relevant attributes (license, provenance, relevant community standards)



Why Metadata?

They are everywhere in FAIR principles...

Findable

- F1 (Meta)data are assigned a globally unique and persistent identifier
- F2 Data are described with rich metadata (R1)
- F3 Metadata clearly and explicitly include the identifier of the data they describe
- F4 (Meta)data are registered or indexed in a searchable resource

Accessible

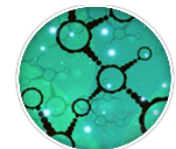
- A1 (Meta)data are retrievable by their identifier using a standardised communications protocol
- A2 Metadata are accessible, even when the data are no longer available

Interoperable

- I1 (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2 (Meta)data use vocabularies that follow FAIR principles
- I3 (Meta)data include qualified references to other (meta)data

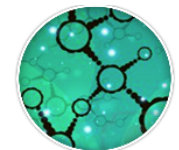
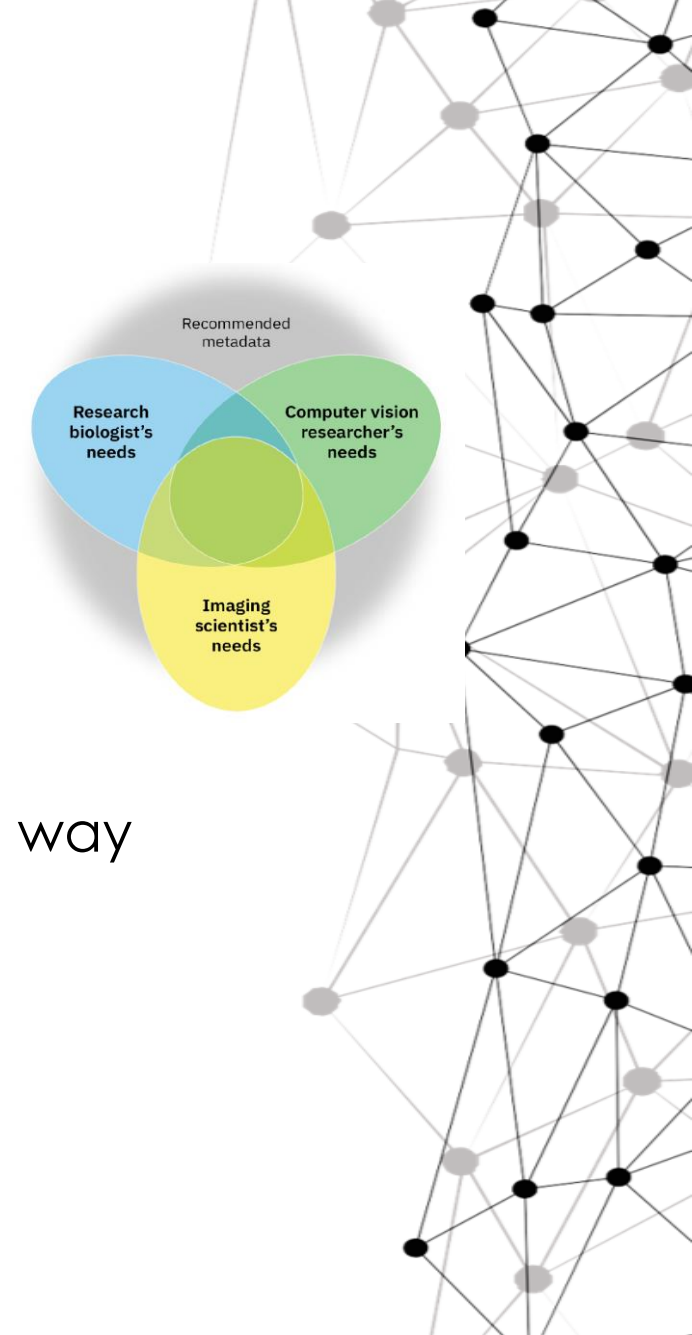
Reusable

- R1 (Meta)data are richly described with a plurality of accurate and relevant attributes (license, provenance, relevant community standards)

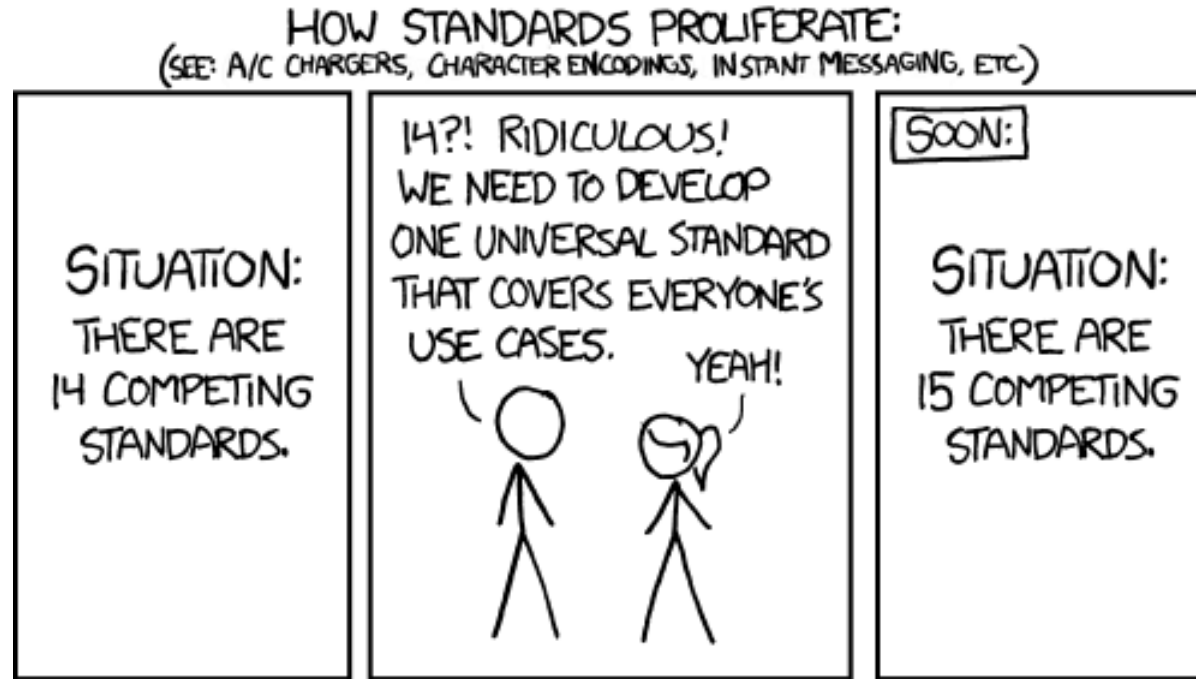


Which [standard] metadata?

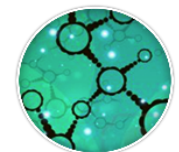
- Metadata are domain dependant:
 - What does the related community need?
 - Minimally ... « The more the better » does not always apply here.
 - They should be « easy » to get and **has a purpose**
- What is a standard? « widely accepted norm », « agreed-upon way of doing something »



Standards?

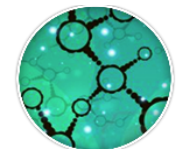
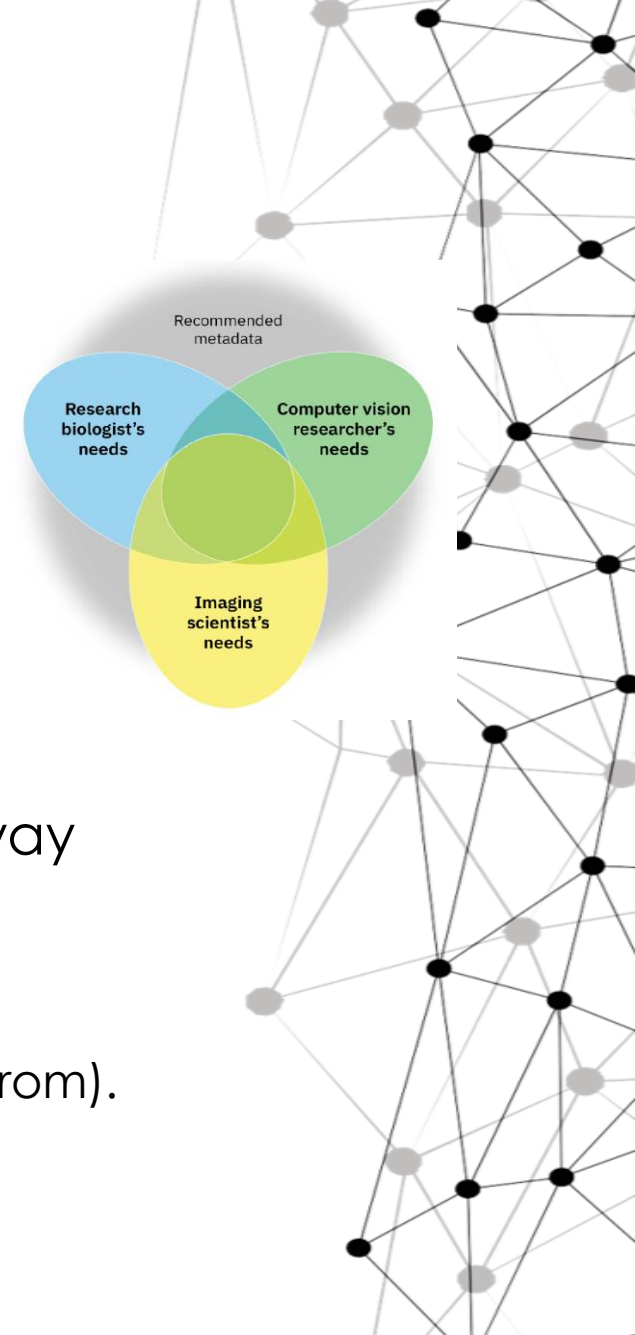


Source: xkcd a webcomic of Romance; Sarcasm, math and Language [xkcd: Standards](#)



Which [standard] metadata?

- Metadata are domain dependant:
 - What does the related community need?
 - Minimum Information Model... « The more the better » does not always here.
 - They should be « easy » to get and **has a purpose**
- What is a standard? « widely accepted norm », « agreed-upon way of doing something »
- Controlled Vocabulary VS Ontology
 - CVs is a list of terms that have been enumerated explicitly (to choose from). Each term has a definition.
 - Ontology is a Controlled Vocabulary express with a « grammar »

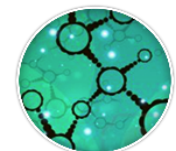
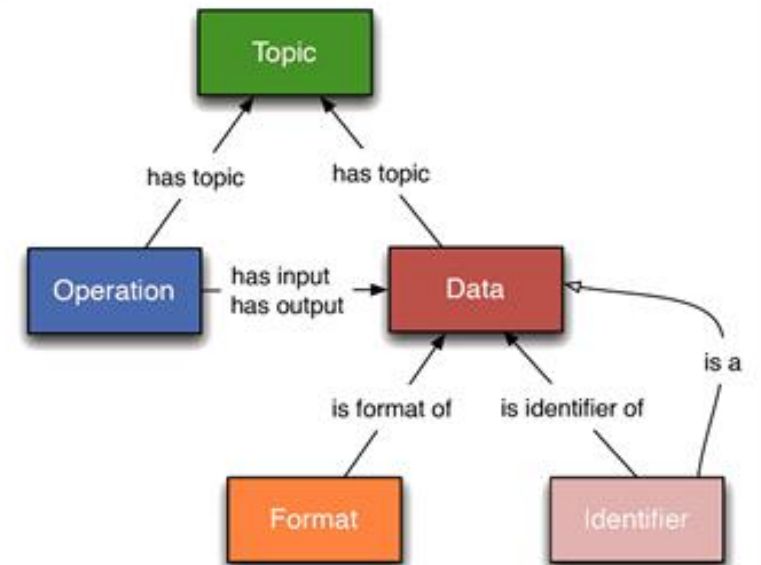


What are ontologies?

- an ontology is a formal explicit description of concepts in a domain of discourse
 - **classes** (sometimes called concepts),
 - **properties** of each concept describing various features

An ontology together with a set of individual instances of classes constitutes a **knowledge base**.
(simplified definition from WebProtégé)

STRUCTURE OF EDAM



What are ontologies?

Example from BIII

www.biii.eu

arXiv > q-bio > arXiv:2312.11256

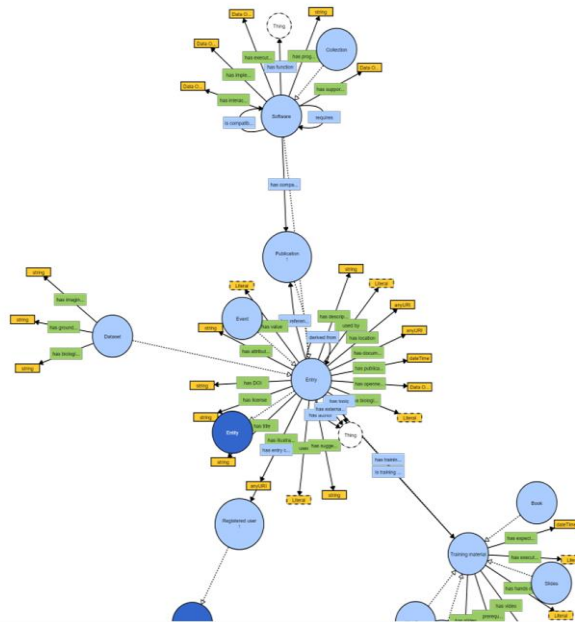
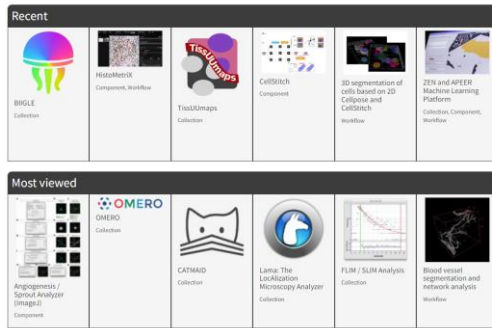
Quantitative Biology > Quantitative Methods

[Submitted on 18 Dec 2023]

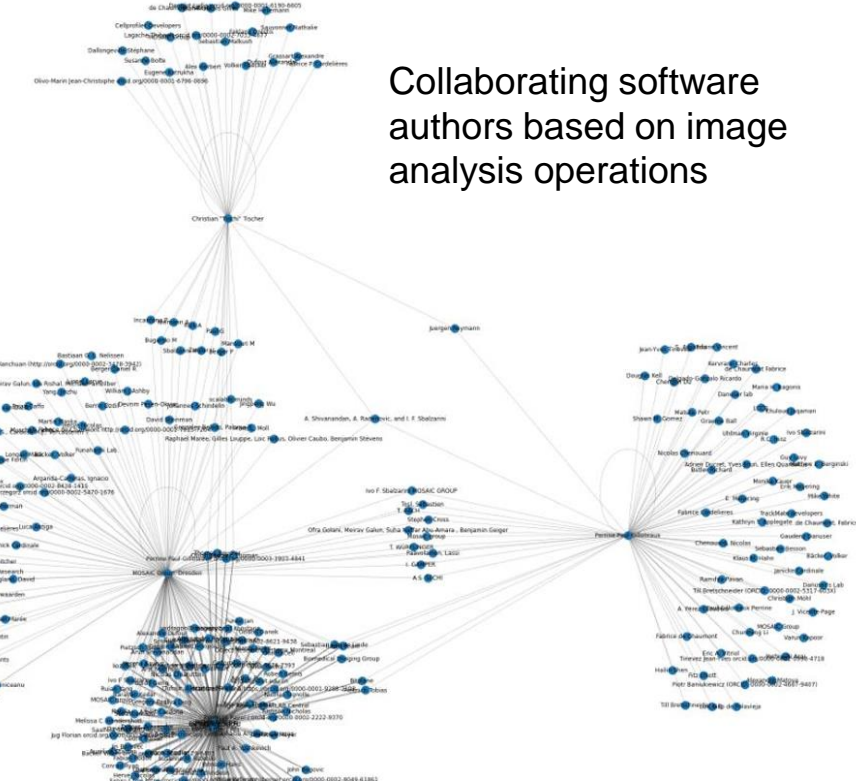
Bio-Image Informatics Index BIII: A unique database of image analysis tools and workflows for and by the bioimaging community

Chong Zhang, Alban Gaignard, Matus Kalas, Florian Levet, Felipe Delestro, Joakim Lindblad, Natasa Sladoje, Laure Plantard, Alain Latour, Robert Haase, Gabriel Martins, Paula Sampaio, Leandro Scholz, NEUBIAS taggers, Sébastien Tosi, Kota Miura, Julien Colombelli, Perrine Paul-Gilloteaux

<https://doi.org/10.48550/arXiv.2312.11256>

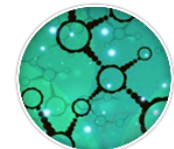


Ontologies



Collaborating software authors based on image analysis operations

Example of queries from the knowledge model

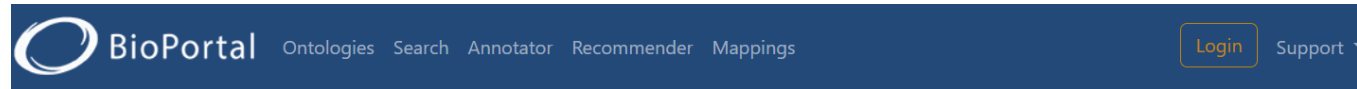


FRANCE-BIOIMAGING



Which metadata?

<https://bioportal.bioontology.org/>



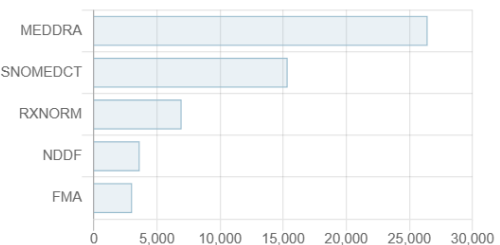
Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

[Advanced search](#)

Find an ontology

Ontology visits (March 2024)

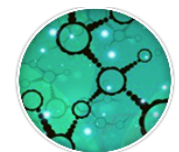


Ontology	Visits (approx.)
MEDDRA	26,000
SNOMEDCT	15,000
RXNORM	7,000
NDDF	3,000
FMA	2,000

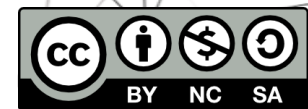
[More](#)

Statistics

Ontologies	1,095
Classes	14,814,549
Properties	36,286
Mappings	94,951,139



FRANCE-BIOIMAGING

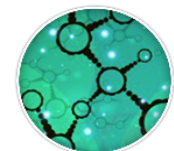


Mapping/Alignment/Matching

Can be automatized by pairing related concepts (BioPortal use LOOM for example based on names and synonyms case-insensitive)

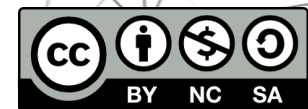
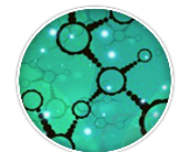
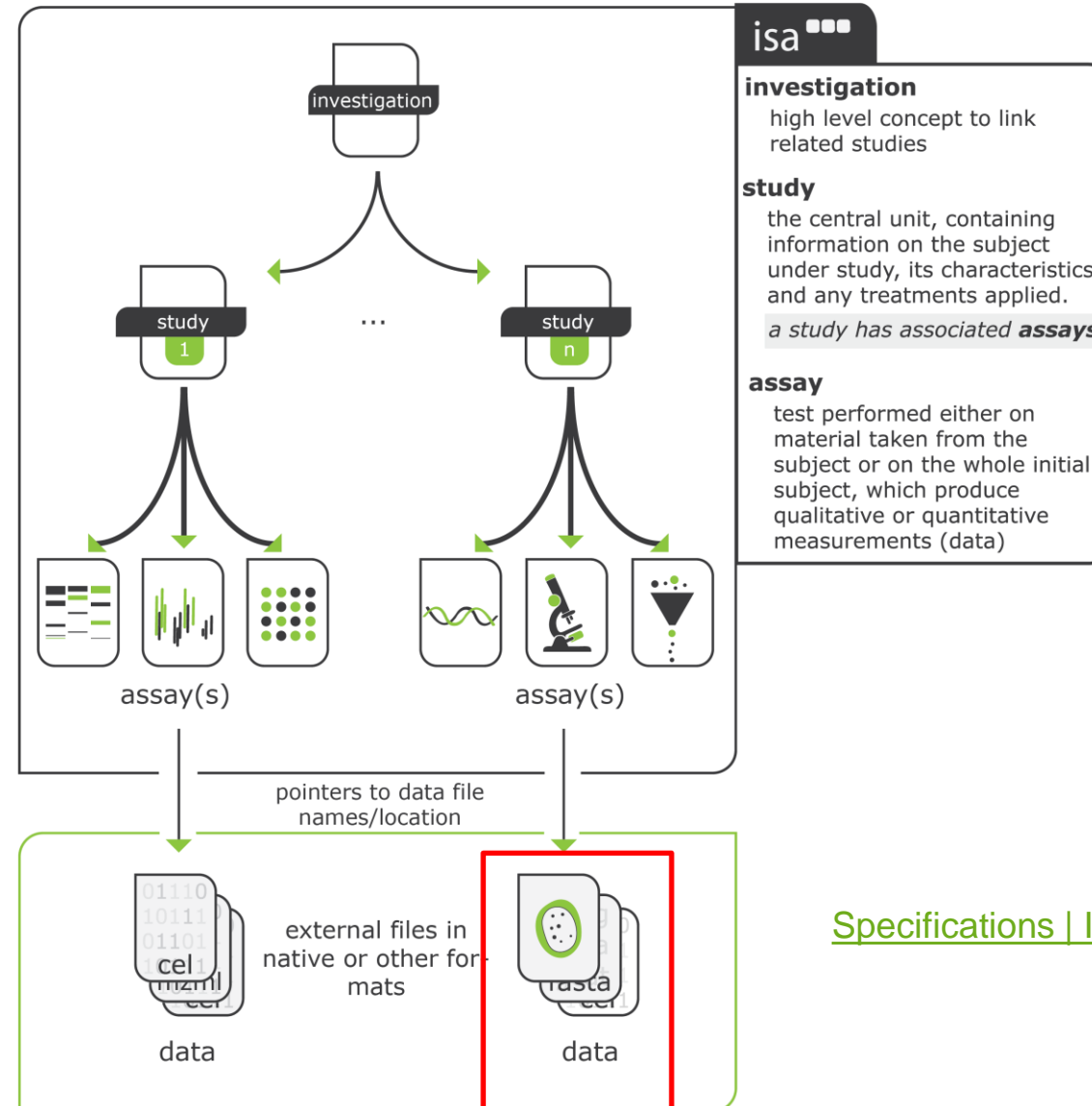
The screenshot displays the EDAM Bioimaging Ontology web interface. At the top, the title "EDAM Bioimaging Ontology" is shown with the date "Last updated: March 3, 2020". Below the title are navigation icons for download, link, home, user, and list. A horizontal menu contains tabs for "Summary", "Classes", "Properties", "Notes", "Mappings", and "Widgets". The "Mappings" tab is active, showing sub-tabs for "Details", "Visualization", "Notes (0)", and "Class Mappings (13)". A "Create mapping" button is visible above a table of mappings. On the left, a tree view shows the ontology structure, with "High-throughput screening" highlighted. The table lists mappings to various external ontologies.

MAPPING TO	ONTOLOGY	SOURCE
http://sbmi.uth.tmc.edu/ontology/ochv#C0872187	OCHV	LOOM
http://www.bioassayontology.org/bao#BAO_0010074	BAO	LOOM
http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#High_Throughput_Screening	CSEO	LOOM
http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#High_Throughput_Screening	SYN	LOOM
http://sbmi.uth.tmc.edu/ontology/ochv#51553	OCHV	LOOM
http://purl.obolibrary.org/obo/STATO_0000239	PSDO	LOOM
http://purl.obolibrary.org/obo/STATO_0000239	CHEAR	LOOM
http://purl.obolibrary.org/obo/STATO_0000239	MS	LOOM
http://purl.obolibrary.org/obo/STATO_0000239	HHEAR	LOOM



ISA: A generic framework for life science experiments

(Yes, there is more than imaging in life [sciences])



Recommended Metadata for Biological Images



Study is the highest level metadata, describing your project, including funding and publications. **Dublin Core, Datacite, Schema.org**

Study Component acts as a container that helps you organise your data, based on experiment types or samples etc. A Study Component contains one or more of the following components, or required if relevant).

Biosample is about what it is you have imaged, for example, the species of the organism that you're imaging, if you're using a particular cell line, genetic background etc. **EFO, NCBI taxonomy**

Specimen metadata describes how your sample was prepared for imaging. **OME, EFO, EDAM-BIOIMAGING, Fbbi**

Image Acquisition is about how your images were captured. **4DN, QUAREP, Fbbi, EDAM-BIOIMAGING**

Image Data contains image level metadata. This is implemented as the [File List](#) for BioImage Archive submissions, instead of a separate component in the submission form. **OME, EDAM-BIOIMAGING**

Image Correlation is optional and contains metadata about correlating images from different modalities. **EDAM-BIOIMAGING**

Analysed Data, includes **Image Analysis** metadata; information about how you analysed your images, if applicable. **EDAM BioIMAGING, OME**



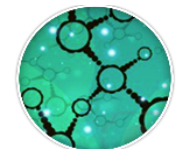
Comment | Published: 21 May 2021

REMBI: Recommended Metadata for Biological Images —enabling reuse of microscopy data in biology

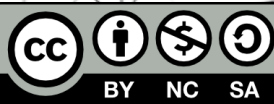
Ugis Sarkans [✉], Wah Chiu, Lucy Collinson, Michele C. Darrow, Jan Ellenberg, David Grunwald, Jean-Karim Hériché, Andrii Iudin, Gabriel G. Martins, Terry Meehan, Kedar Narayan, Ardan Patwardhan, Matthew Robert Geoffrey Russell, Helen R. Saibil, Caterina Strambio-De-Castilla, Jason R. Swedlow, Christian Tischer, Virginie Uhlmann, Paul Verkade, Mary Barlow, Omer Bayraktar, Ewan Birney, Cesare Catavittello, Christopher Cawthorne, Stephan Wagner-Conrad, Elizabeth Duke, Perrine Paul-Gilloteaux, Emmanuel Gustin, Maria Harkiolaki, Pasi Kankaanpää, Thomas Lemberger, Jo McEntyre, Josh Moore, Andrew W. Nicholls, Shuichi Onami, Helen Parkinson, Maddy Parsons, Marina Romanchikova, Nicholas Sofroniew, Jim Swoger, Nadine Utz, Lenard M. Voortman, Frances Wong, Peijun Zhang, Gerard J. Kleywegt [✉] & Alvis Brazma [✉] [Show fewer authors](#)

[Nature Methods](#) **18**, 1418–1422 (2021) | [Cite this article](#)

12k Accesses | 55 Citations | 43 Altmetric | [Metrics](#)

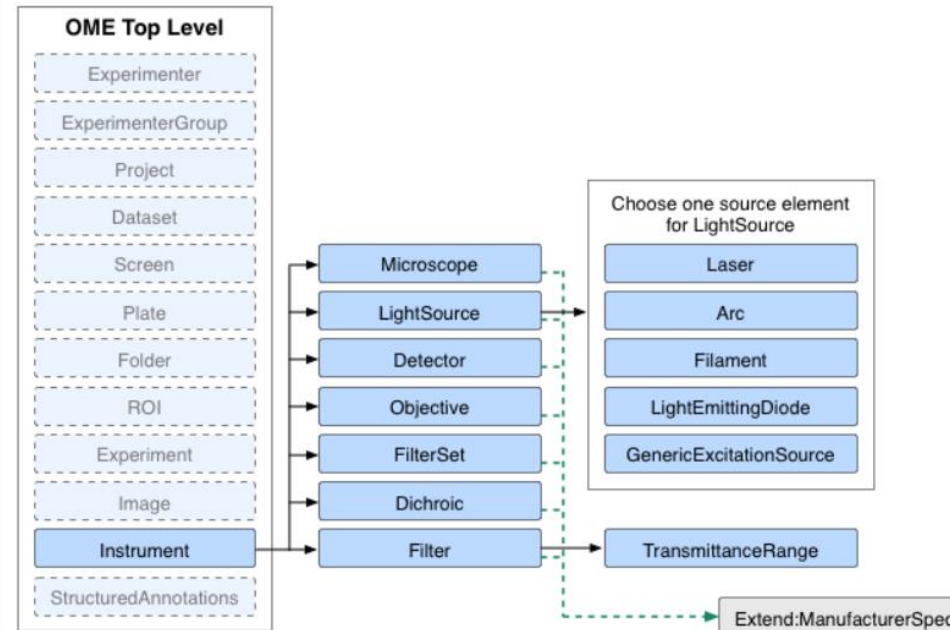
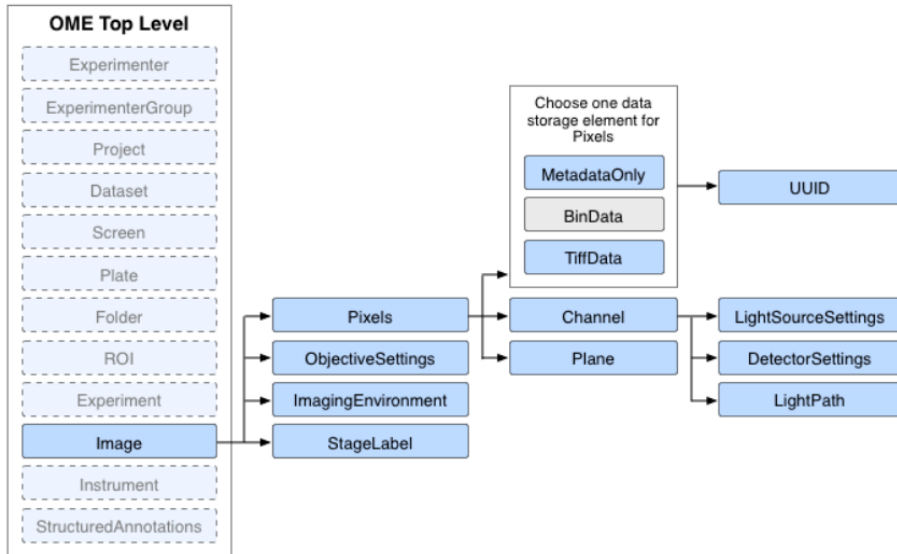


FRANCE-BIOIMAGING



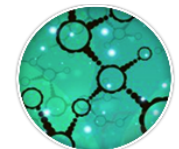
OME : standard metadata encoded in image file format (our closest to « DICOM » for Connoisseurs...) BioFormat/Ome-Tif/Ome-zarr/Ngff

Used in most bio image software



Instrument branch of the OME Schema

[Current Data Model overview — OME Data Model and File Formats documentation \(ome-model.readthedocs.io\)](#)

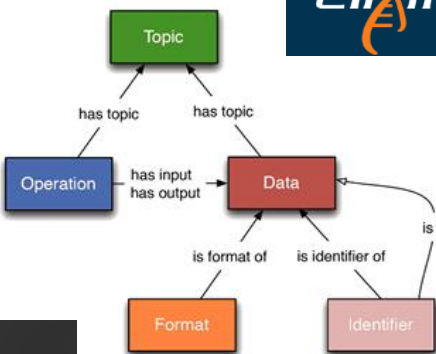


FRANCE-BIOIMAGING



Edam Bioimaging

STRUCTURE OF EDAM

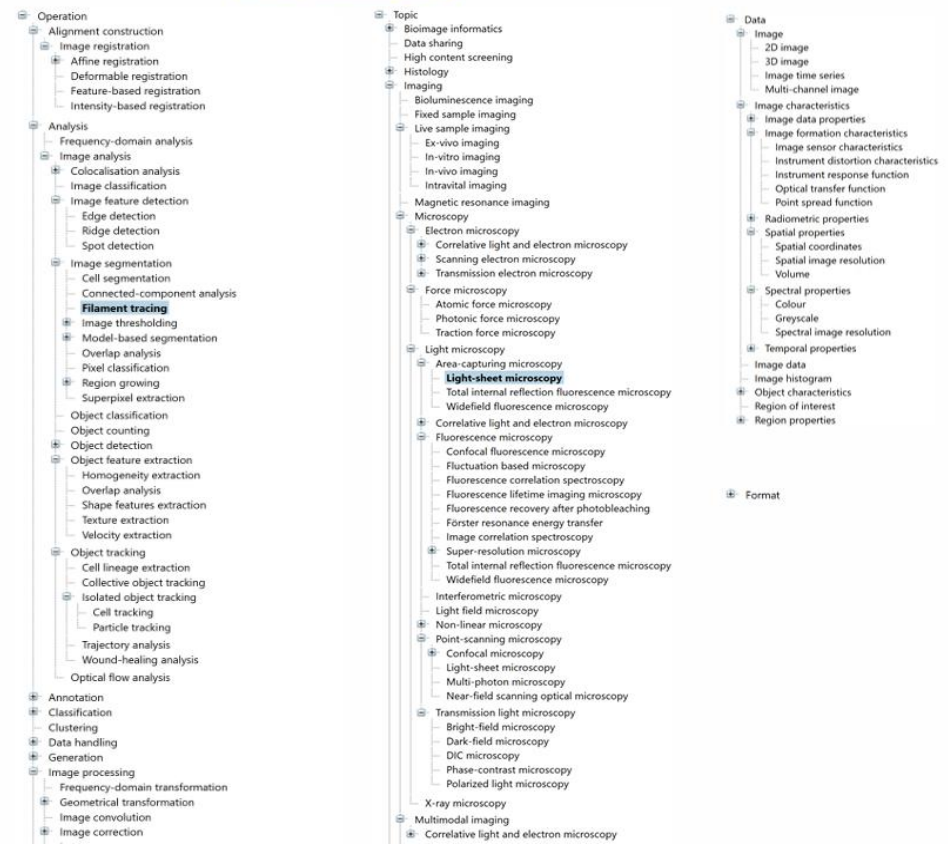


EXAMPLE CONCEPTS

Preferred Name	Light-sheet microscopy
Definition	Fluorescence microscopy technique in which the plane of illumination is orthogonal to the axis of observation.
LSFM	Selective Plane Illumination Microscopy
hasNarrowSynonym	Digitally scanned Laser Light-sheet Microscopy SPIM
hasRelatedSynonym	Lattice Light-sheet Microscopy Dual-View inverted SPIM Spherical aberrations assisted Extended Depth-of-field Lightsheet Microscopy Bessel Beam Lightsheet Microscopy single objective Selective Plane Illumination Microscopy Hardware implementations: multidirectional SPIM LLSM Multiview Selective Plane Illumination Microscopy MuVSPIM inverted SPIM soSPIM dSPIM COLM SPED Clarity Optimized Light-sheet Microscopy mSPIM iSPIM
seeAlso	https://en.wikipedia.org/wiki/Light-sheet_fluorescence_microscopy
subClassOf	Area-capturing microscopy Point-scanning microscopy

Preferred Name	Filament tracing
Definition	Filament tracing operations are image analysis operations in which there is an image of a filamentous structure (it may be a tree-like structure, a filament network or a agglomeration of single 'stick-like' filaments) as input and outputs data that represent the filament, most commonly a skeleton representation of the filaments and their diameters or surfaces.
hasExactSynonym	Tubular structure extraction
hasNarrowSynonym	Biofilament tracing
hasRelatedSynonym	Curvilinear structure reconstruction Curvilinear structure detection
Related term	Neuron reconstruction
seeAlso	Neuron image analysis
subClassOf	Image segmentation

HIERARCHIES OF EDAM-bioimaging

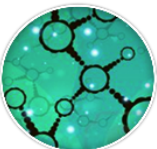


What? EDAM-bioimaging is an extension of the EDAM ontology, dedicated to bioimage analysis, bioimage informatics, and bioimaging.

Why? EDAM-bioimaging enables interoperable descriptions of software, publications, data, and workflows, fostering reliable and transparent science.

How? EDAM-bioimaging is developed in a community spirit, in a welcoming collaboration between numerous bioimaging experts and ontology developers.

How can I contribute? We need your expertise! **You can help by reviewing parts of EDAM-bioimaging, posting comments with suggestions, requirements, or needs for clarification, or participating in a Taggathon or another hackathon.** Please see <https://github.com/edamontology/edam-bioimaging#contributing>.



FRANCE-BIOIMAGING



MIFA: Metadata, Incentives, Formats, and Accessibility guidelines

Reduce number of FORMATS

NGFF/OME-Zarr

GeoJSON

EMDB-SFF

COCO



Reach out to journals

Data browsing

Link to community tools

API

Metadata search

Allow versioning



Improve data ACCESSIBILITY

Standardise METADATA

Authors
Annotation overview
Annotation type
Annotation method
Confidence level
Annotation criteria
Annotation coverage
Source image association
Transformations
Spatial information
Creation time



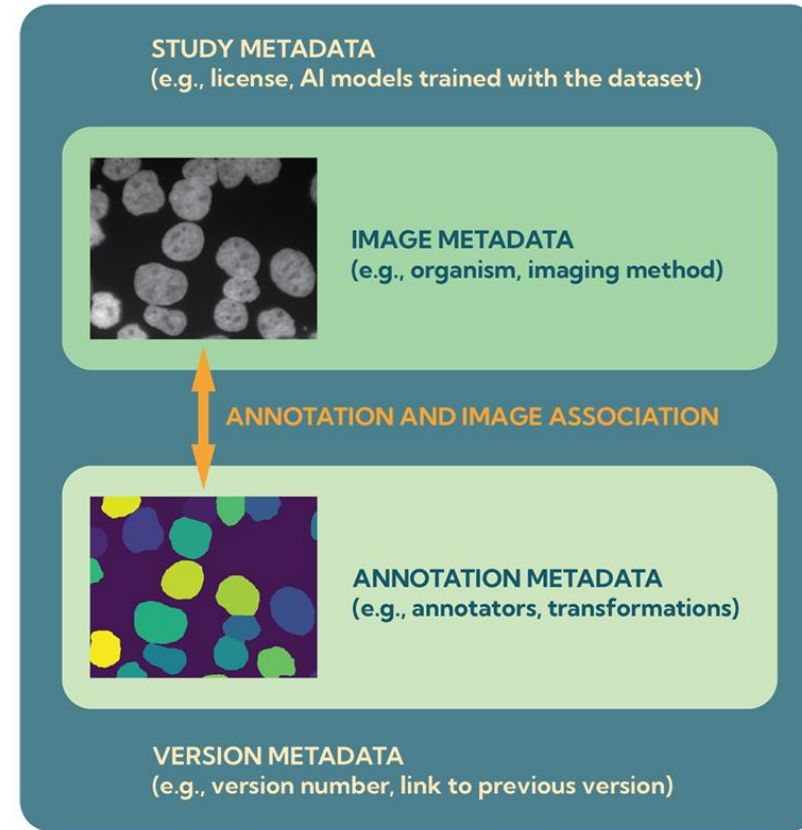
Clear contributor credit

Organise annotation events

Acknowledge exceptional contributors

Preserve credit in different dataset versions

Data production INCENTIVES



Quantitative Biology > Other Quantitative Biology

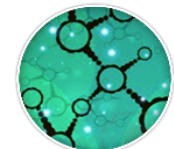
[Submitted on 17 Nov 2023 (v1), last revised 22 Nov 2023 (this version, v2)]

MIFA: Metadata, Incentives, Formats, and Accessibility guidelines to improve the reuse of AI datasets for bioimage analysis

Teresa Zulueta-Coarasa, Florian Jug, Aastha Mathur, Josh Moore, Arrate Muñoz-Barrutia, Liviu Anita, Kola Babalola, Pete Bankhead, Perrine Gilloteaux, Nodar Gogoberidze, Martin Jones, Gerard J. Kleywegt, Paul Korir, Anna Kreshuk, Aybüke Küpcü Yıldız, Luca Marconato, Kedar Narayan, Nils Norlin, Bugra Oezdemir, Jessica Riesterer, Norman Rzepka, Ugis Sarkans, Beatriz Serrano, Christian Tischler, Virginie Uhlmann, Vladimir Ulman, Matthew Hartley

Artificial Intelligence methods are powerful tools for biological image analysis and processing. High-quality annotated images are key to training and developing new methods, but access to such data is often hindered by the lack of standards for sharing datasets. We brought together community experts in a workshop to develop guidelines to improve the reuse of bioimages and annotations for AI applications. These include standards on data formats, metadata, data presentation and sharing, and incentives to generate new datasets. We are positive that the MIFA (Metadata, Incentives, Formats, and Accessibility) recommendations will accelerate the development of AI tools for bioimage analysis by facilitating access to high quality training data.

Comments: 16 pages, 3 figures
Subjects: [Other Quantitative Biology \(q-bio.OT\)](#), [Image and Video Processing \(eess.IV\)](#)
Cite as: [arXiv:2311.10443 \[q-bio.OT\]](#)
(or [arXiv:2311.10443v2 \[q-bio.OT\]](#) for this version)
<https://doi.org/10.48550/arXiv.2311.10443>

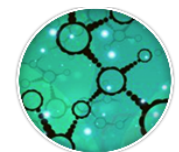


FRANCE-BIOIMAGING



Take Home Messages

- The list of metadata has to be identified from the beginning of a project (Data Management Plans)
- Plan and spend some times curating your data according to where you will deposit them (DMP)
- Try to fit a common effort with no reinvention of the wheel: rather contribute to make existing metadata framework to evolve
- Alignment of ontologies (Mapping) is a key for interoperability and make your initial choice less dramatic





EURO-BIOIMAGING

GLOBAL BIOIMAGING
growing collaboration

Thanks for your attention!

Ome Team
Josh Moore
Jason Swedlow
Kevin Eliceiri

Elixir /IFB Team
Jon Ison
Matus Kalas
Hervé Ménager
Alban Gaignard

FBI data team
Emmanuel Faure
Guillaume Gay
Marc Mongy
Jean-François Guillaume
Guillaume Maucort
Raphaël Braud-Mussi
Thierry Pecot
Dorian Kauffmann
Théo Barnouin

BiolmageArchive Team
Matthew Hartley

Neubias taggers
<https://biii.eu/show-taggers>

FRANCE-BIOIMAGING



A large grid of small cards, each containing a person's name, a small portrait, and their affiliation. The cards are arranged in a grid-like pattern, with some larger cards for team names and contact information. The grid covers most of the page, with text blocks and logos overlaid on it.

