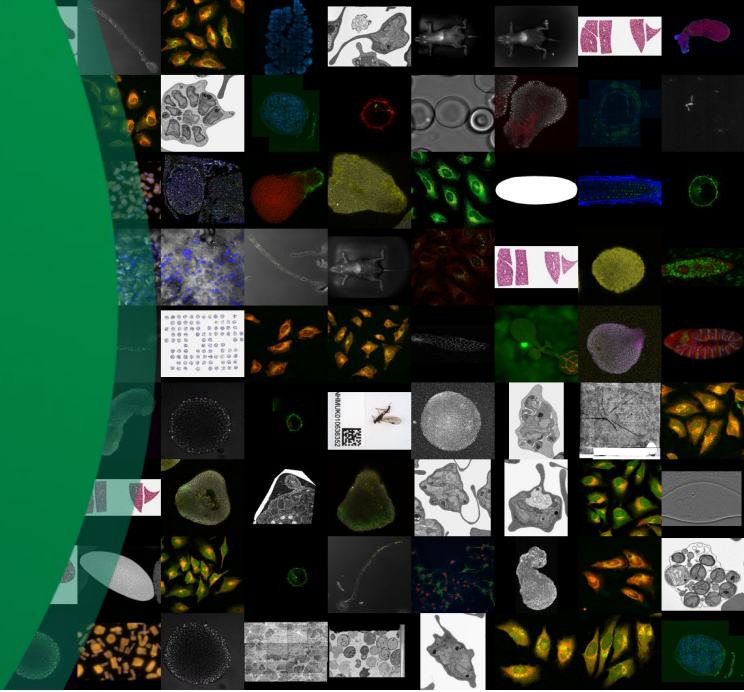
Image data interoperability

with OME-NGFF, RO-Crate and linked data



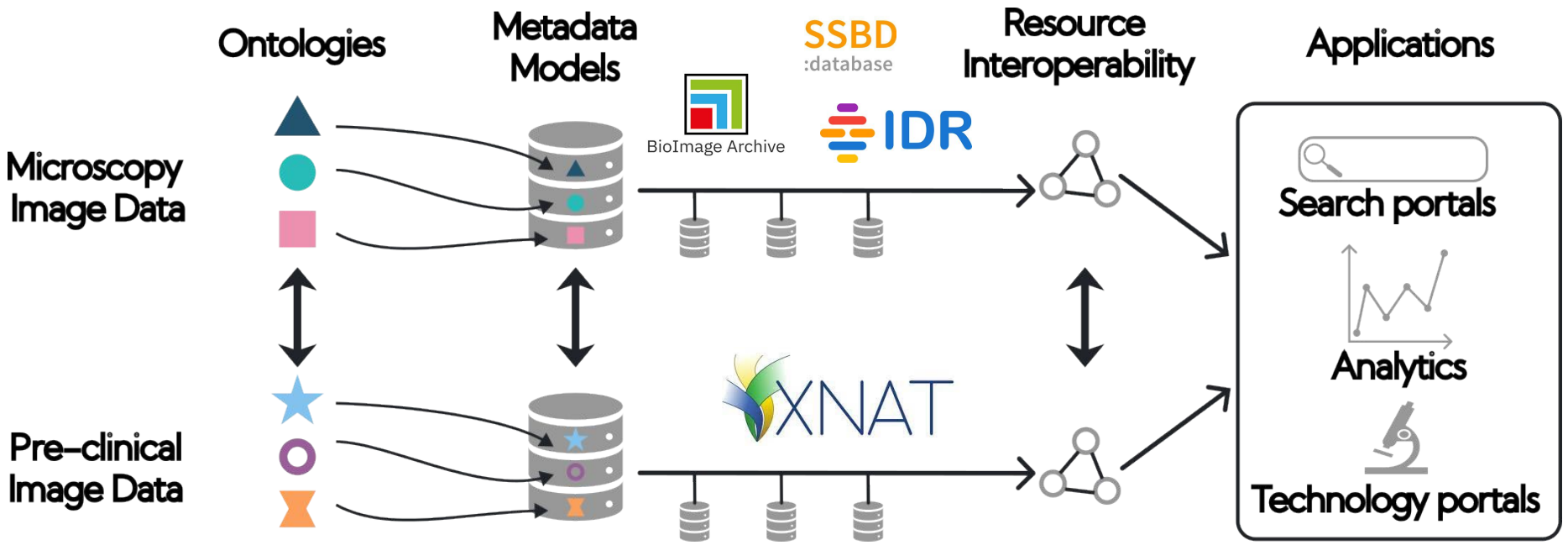
Matthew Hartley

Team Leader, BioImage Archive / EMPIAR

FoundingGIDE Community Event, 2024

EMBL-EBI

FoundingGIDE, technical stream



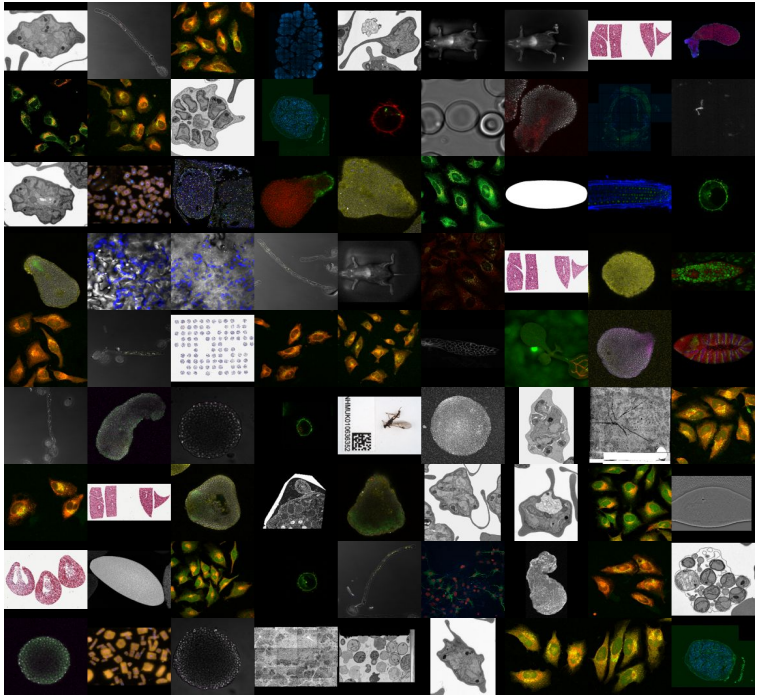
What is FoundingGIDE bioimaging attempting?

Can we agree on…

- …enough shared information…
- …between three repositories…
- …to enable some interoperability…
- …proved through search?

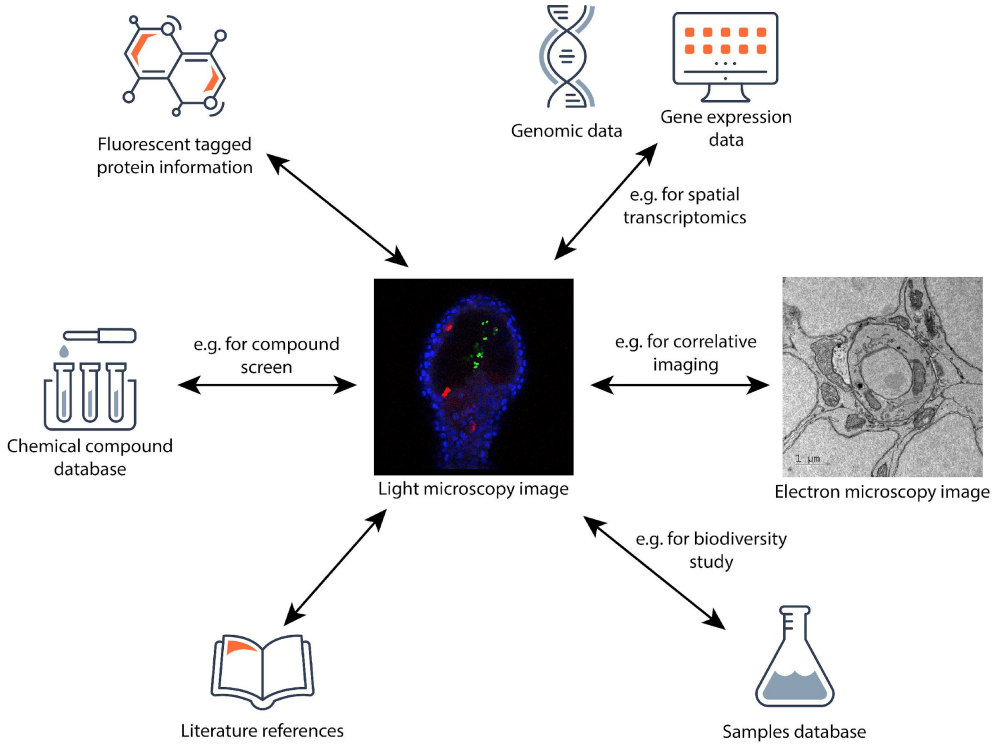
EMBL-EBI

Managing image data is very complicated! So…



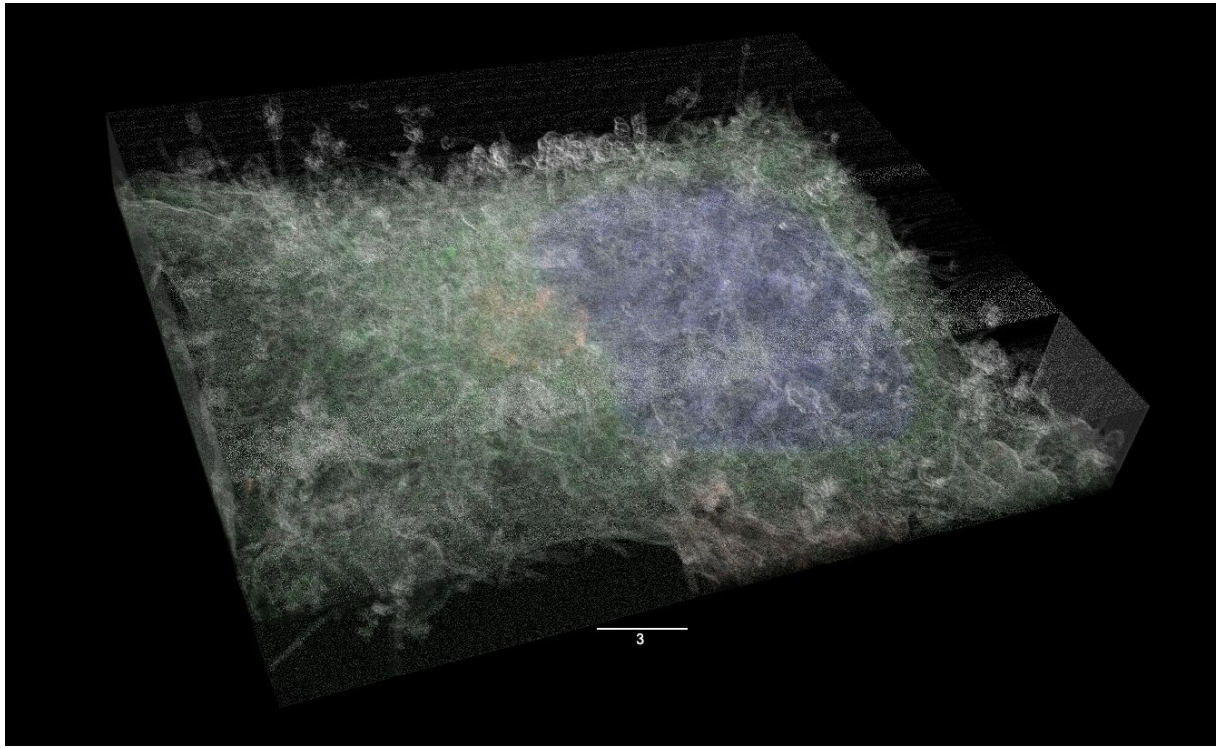
EMBL-EBI

Let’s make more complicated!



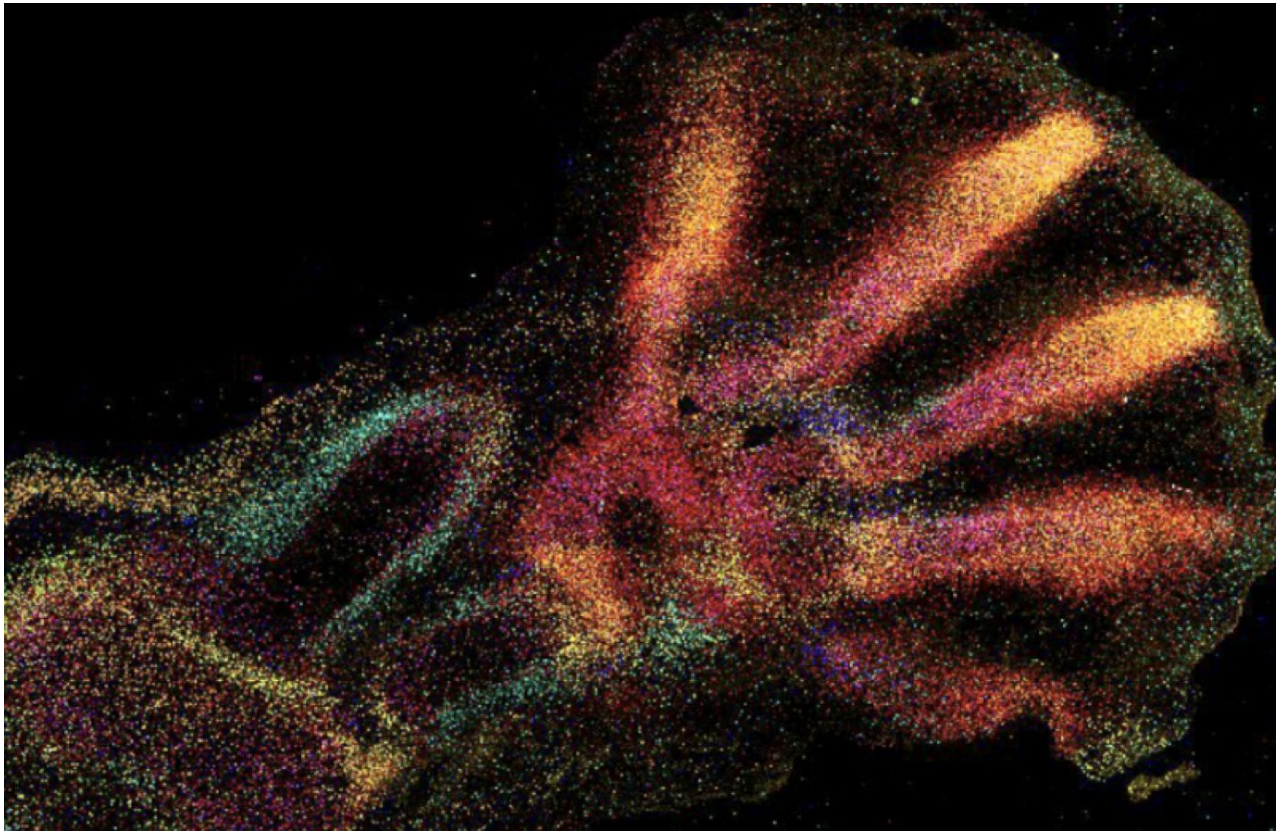
How to we link imaging data with other biomolecular data and beyond?

Data integration – image + image



EMBL-EBI

Data integration – image + other biological



ArrayExpress

+

BioStudies.

+

ENA
European Nucleotide Archive

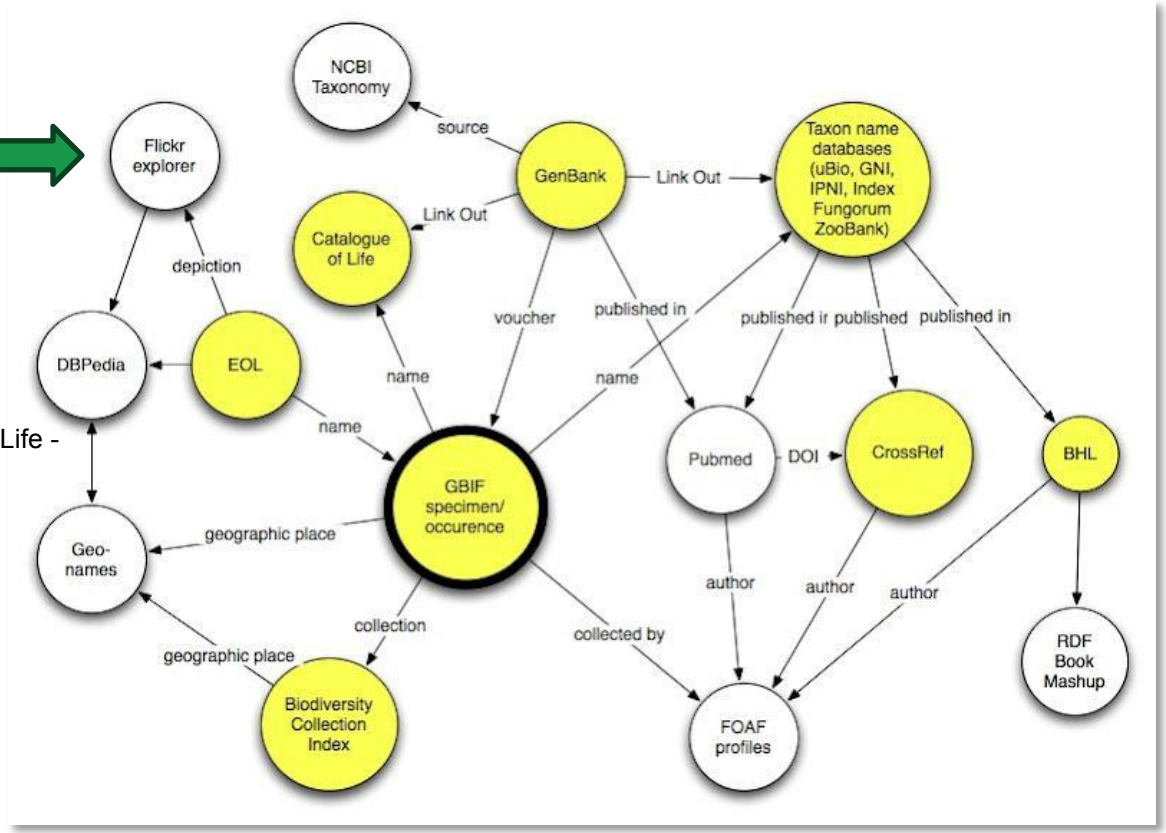
+

BioImage Archive

Sometimes images are at the periphery

Images here

From **S-BIAD588**, Darwin Tree of Life - NHM samples image catalogue



https://iphylo.blogspot.com/2009/08/gbif-and-linked-data.html

EMBL-EBI

What are we trying to do (long term)?

"Show me images of how mitochondria in marine organisms respond to differing climate conditions"

- Need to understand the terms
- Need to link across domains
- One repository will not have all of the information!

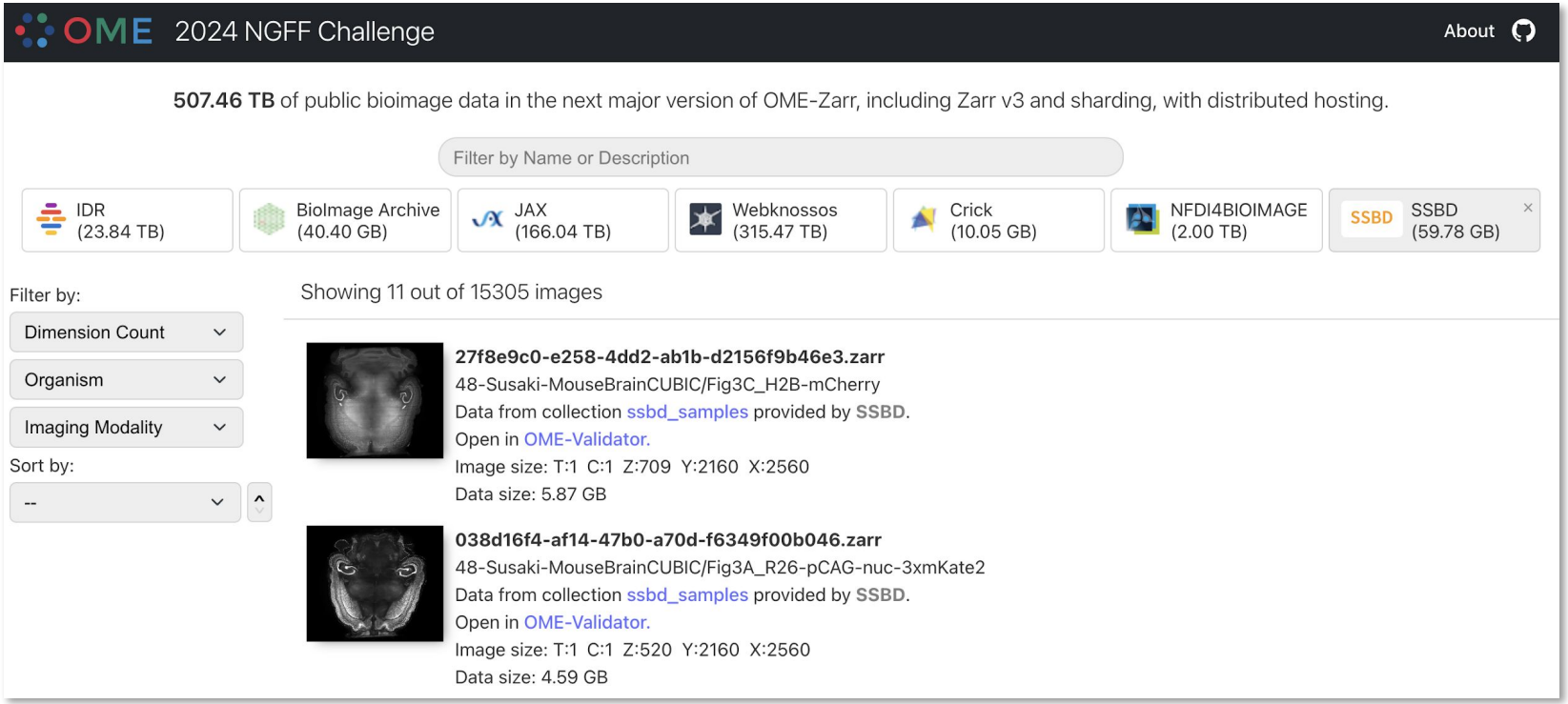
(one of) The problem (s)

- BioImaging is many domains together
- Broad / deep
- Local institutions, communities always have specific needs
- Work on some degree of interoperability

The solution (sort of):

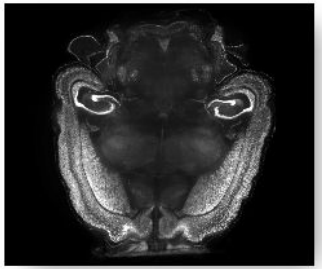
Step 1: Do something useful
Step 2: Make it better

OME 2024 NGFF challenge



EMBL-EBI

How does this work?



038d16f4-af14-47b0-a70d-f6349f00b046.zarr

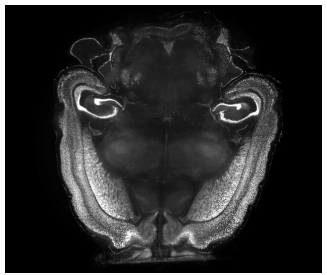
48-Susaki-MouseBrainCUBIC/Fig3A_R26-pCAG-nuc-3xmKate2

Data from collection ssbd_samples provided by **SSBD**.

Open in OME-Validator.

Image size: T:1 C:1 Z:520 Y:2160 X:2560

Data size: 4.59 GB



RO-Crate metadata

License: ✓ https://creativecommons.org/licenses/by/4.0/

Name: ✓ 48-Susaki-MouseBrainCUBIC/Fig3A_R26-pCAG-nuc-3xmKate2

Description: ✓ Light-sheet fluorescence microscopy (LSFM) images of R26-pCAG-nuc-3 x mKate2 mouse brain

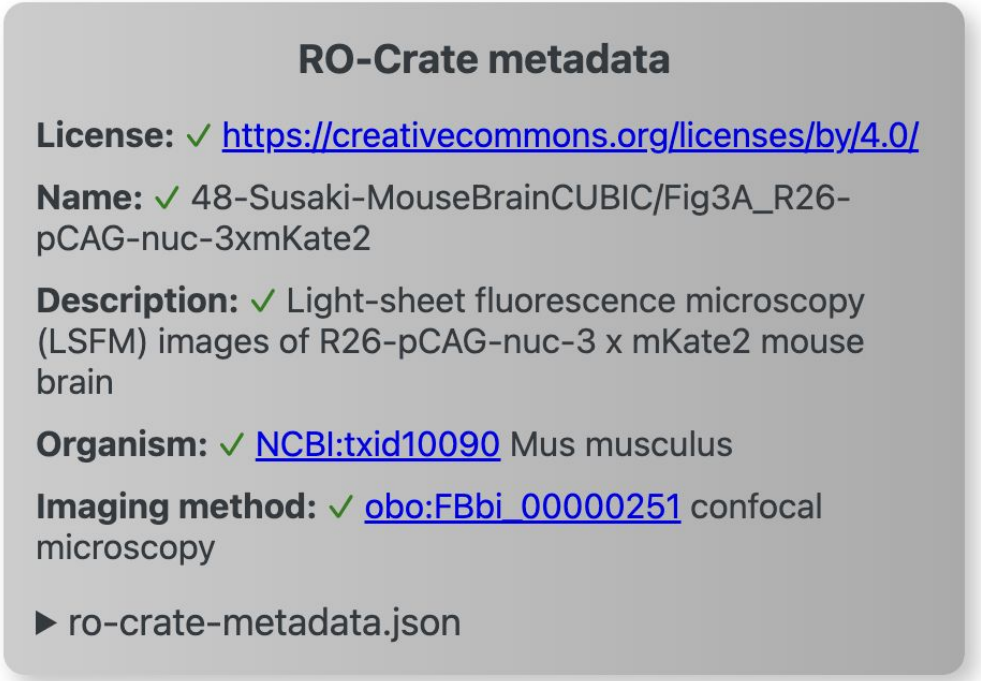
Organism: ✓ NCBI:txid10090 Mus musculus

Imaging method: ✓ obo:FBbi_00000251 confocal microscopy

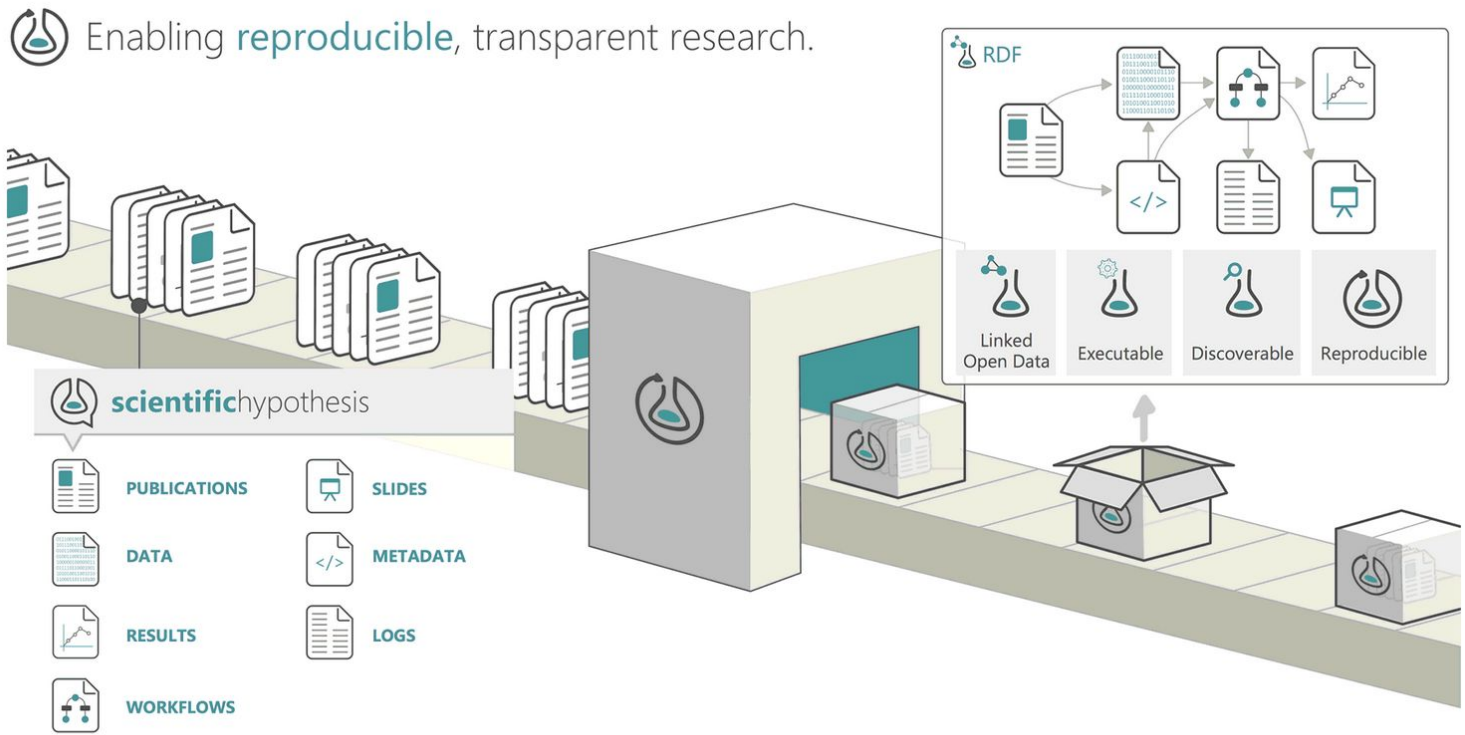
▶ ro-crate-metadata.json

EMBL-EBI

What's in the box?



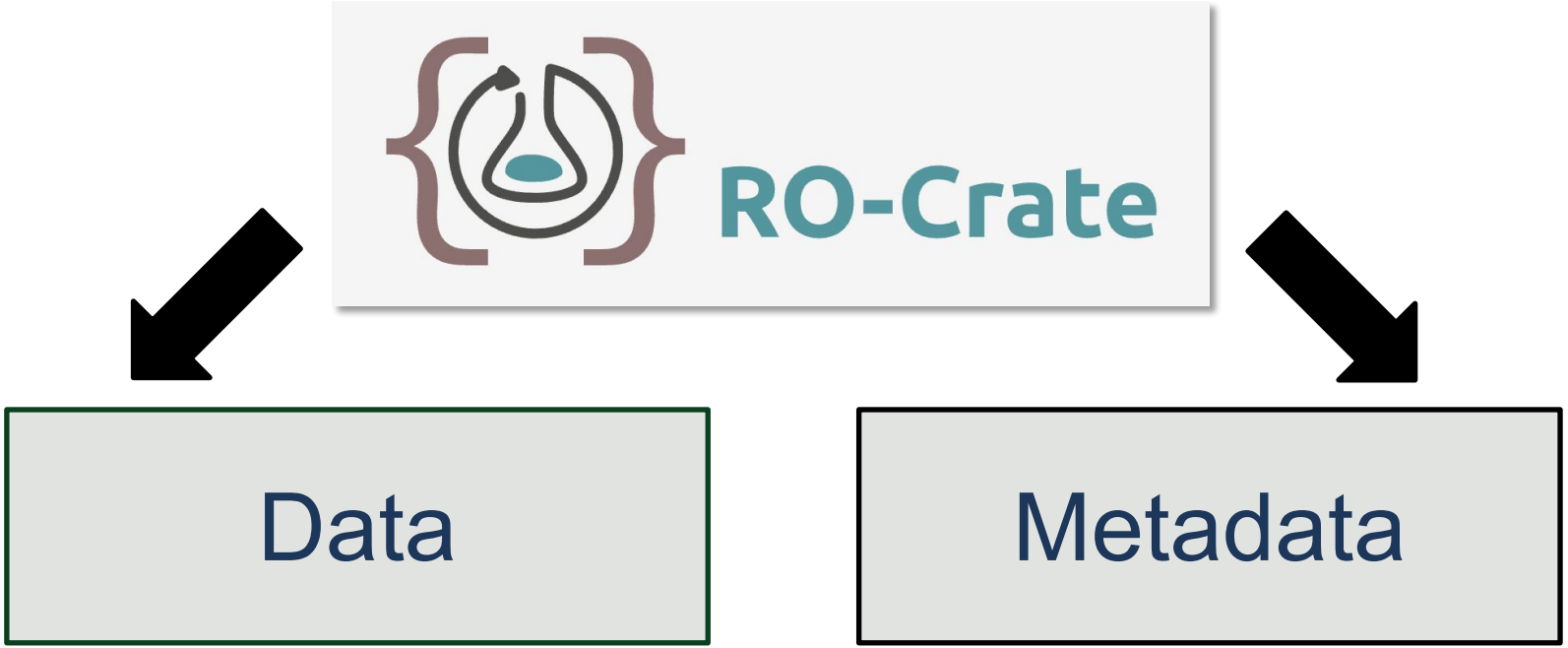
What is RO-Crate?



EMBL-EBI

What is RO-Crate?

- A Research Object, which combines both data and metadata



Metadata – JSON-LD

JSON

What is the stuff?

Context

How are we describing it?

Structuring of Data and Metadata in Bioimaging: Concepts and technical Solutions in the Context of Linked Data

- Susanne Kunis & Julia Dohle

- 10.5281/zenodo.7018750

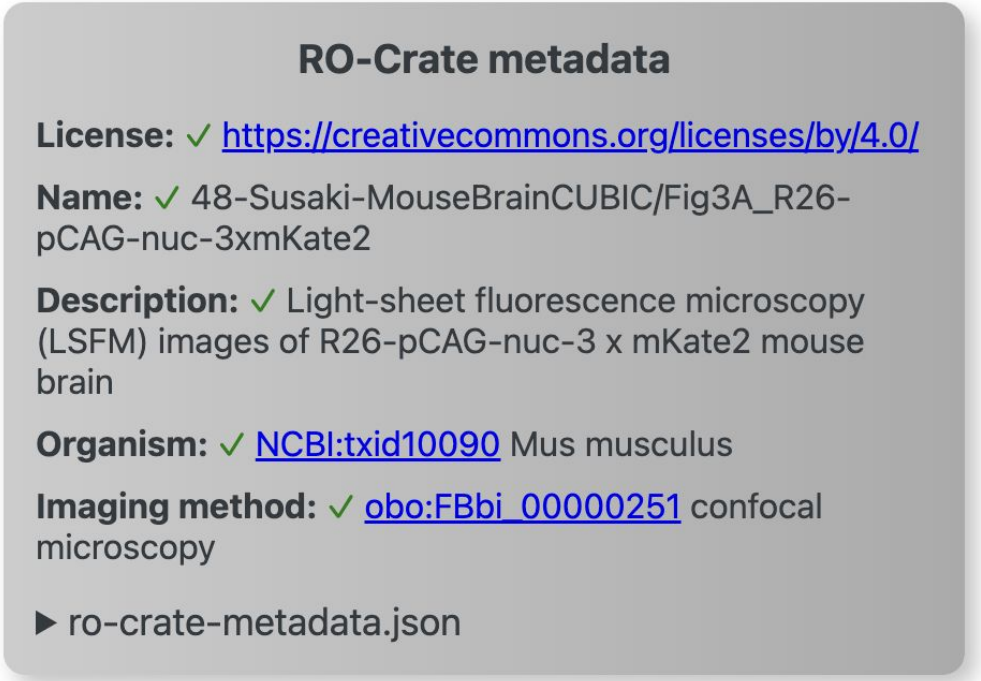
EMBL-EBI

The promise of linked data

The downsides

- More complex
- Tends to be harder for developers to work with

What's in the box?



ro-crate-metadata.json


```
{
  "@context": [
    "https://w3id.org/ro/crate/1.1/context",
    {
      "organism_classification": "https://schema.org/taxonomicRange",
      "BioChemEntity": "https://schema.org/BioChemEntity",
      "channel": "https://www.openmicroscopy.org/Schemas/Documentation/Generated/OME-2016-06/ome_xsd.html#Channel",
      "obo": "http://purl.obolibrary.org/obo/",
      "FBcv": "http://ontobee.org/ontology/FBcv/",
      "acquisiton_method": {
        "@reverse": "https://schema.org/result",
        "@type": "@id"
      },
      "biological_entity": "https://schema.org/about",
      "biosample": "http://purl.obolibrary.org/obo/OBI_0002648",
      "preparation_method": "https://www.wikidata.org/wiki/Property:P1537",
      "specimen": "http://purl.obolibrary.org/obo/HSO_0000308"
    }
  ],
  "@graph": [
    {
      "@id": "./",
      "@type": "Dataset",
      "name": "Arabidopsis phloem pole",
      "description": "Five-day-old Col-0 Arabidopsis thaliana root, phloem pole unloading zone (339-381 um from the root tip)",
      "license": "https://creativecommons.org/licenses/by/4.0/",
      "resultOf": {
        "@id": "#a704c99d-1cac-4d2a-86c3-13d53841cc58"
      }
    },
    {
      "@id": "ro-crate-metadata.json",
      "@type": "CreativeWork",
      "conformsTo": {
        "@id": "https://w3id.org/ro/crate/1.1"
      },
      "about": {
        "@id": "./"
      }
    },
    {
      "@id": "#f377f8fb-3ee7-430e-8d63-7948027c1d7c",
      "@type": "biosample",
      "organism_classification": {
        "@id": "NCBI:txid3701"
      }
    },
    {
      "@id": "#16762b2e-ee4c-4880-897b-6102760ba5a9",
      "@type": "specimen",
      "biosample": {
        "@id": "#f377f8fb-3ee7-430e-8d63-7948027c1d7c"
      }
    },
    {
      "@id": "#a704c99d-1cac-4d2a-86c3-13d53841cc58",
      "@type": "image_acquisition",
      "fbbi_id": {
        "@id": "obo:FBbi_00000585"
      },
      "specimen": {
        "@id": "#16762b2e-ee4c-4880-897b-6102760ba5a9"
      }
    }
```


EMBL-EBI

“Minimal” image metadata

What is BioImaging?

Bio

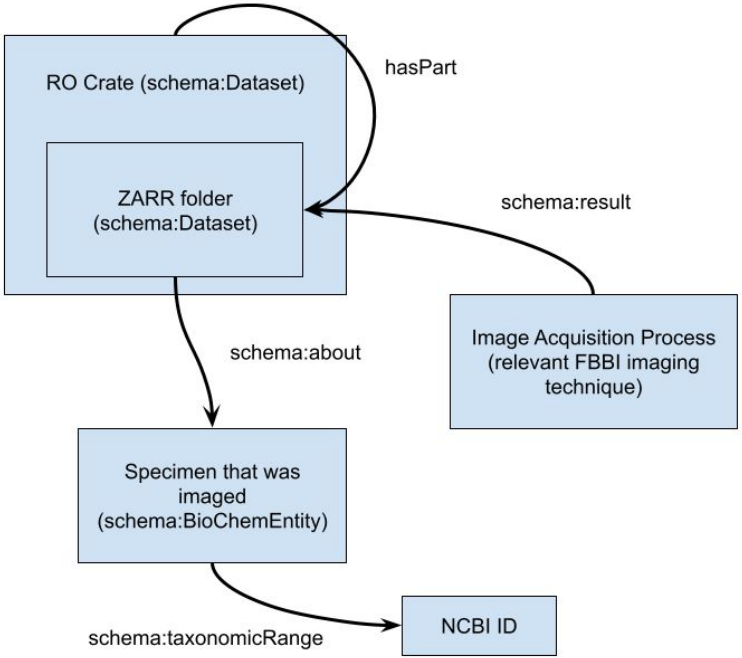
imaging

EMBL-EBI

What does the context look like?

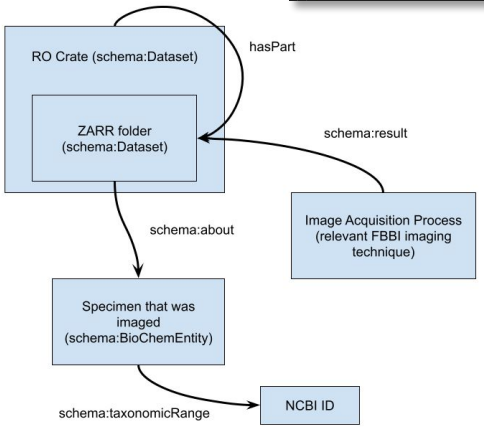
```
"@context": [
  "https://w3id.org/ro/crate/1.1/context",
  {
    "organism_classification": "https://schema.org/taxonomicRange",
    "BioChemEntity": "https://schema.org/BioChemEntity",
    "channel": "https://www.openmicroscopy.org/Schemas/Documentation/Generated/OME-2016-06/ome_xsd.html#Channel",
    "obo": "http://purl.obolibrary.org/obo/",
    "FBcv": "http://ontobee.org/ontology/FBcv/",
    "acquisiton_method": {
      "@reverse": "https://schema.org/result",
      "@type": "@id"
    },
    "biological_entity": "https://schema.org/about",
    "biosample": "http://purl.obolibrary.org/obo/OBI_0002648",
    "preparation_method": "https://www.wikidata.org/wiki/Property:P1537",
    "specimen": "http://purl.obolibrary.org/obo/HSO_0000308"
  }
],
```

Linked together



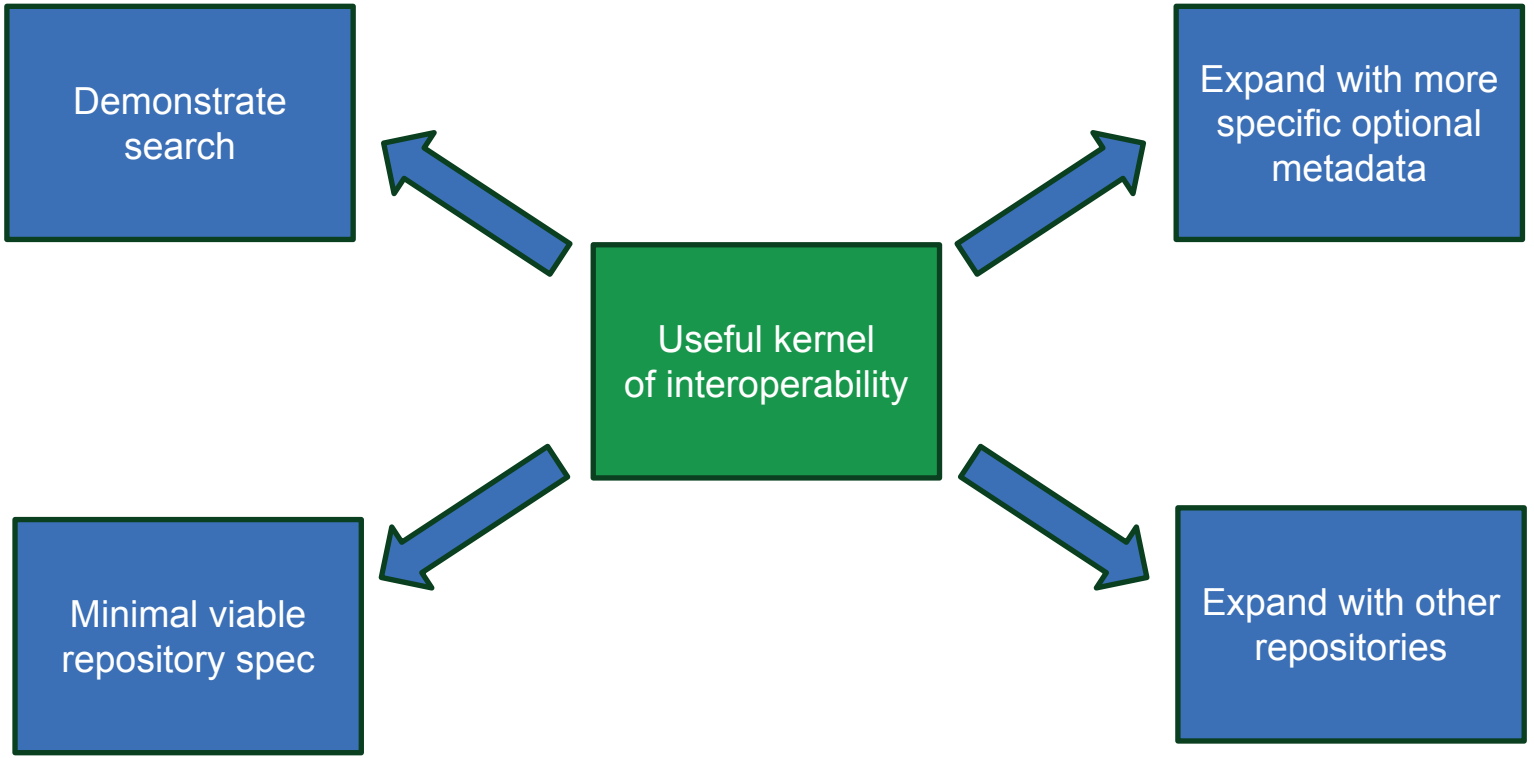
What do we have? Why is it useful?

- Functional, minimal proof of concept of different sources of imaging data coming together
- Allows us to expand and try new things, while ensuring that we continually test what is useful
- Spurs development of tooling
- Forces decisions about ontologies / controlled vocabulary
- Highlights where the missing parts are



EBI

Where is it going?



Thanks, part I…

UK Research and Innovation

EMBL

Liviu Anita

Kola Babalola

Teresa Zulueta

Aybuke Yoldas

François Sherwood

Craig Russell

Ugis Sarkans

UKRI Biotechnology and Biological Sciences Research Council

wellcome

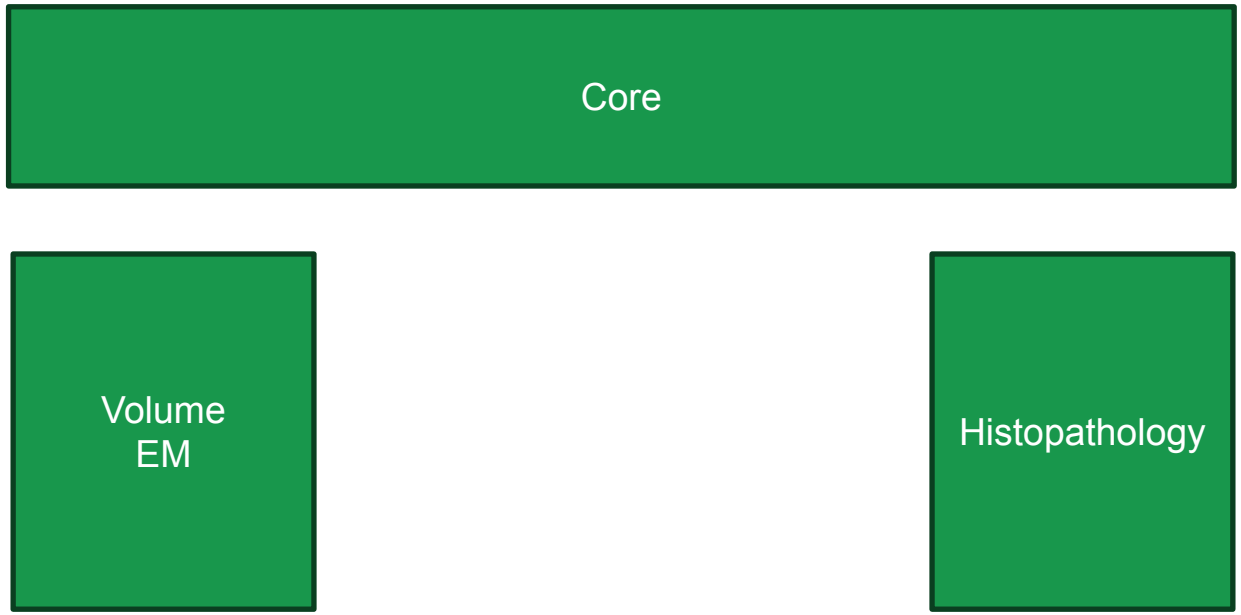
UKRI Medical Research Council

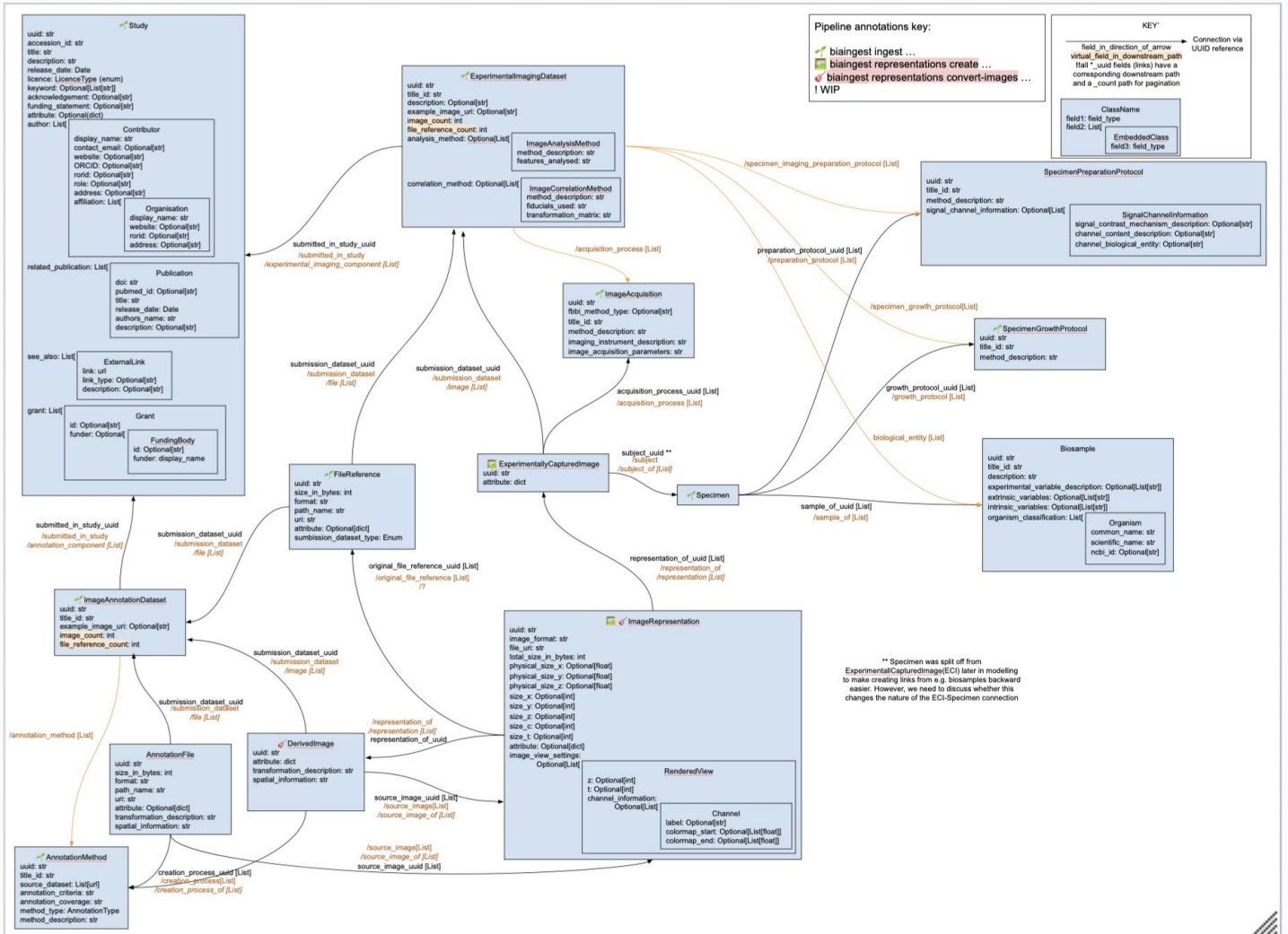
GLOBAL BIOIMAGING growing collaboration

EURO BIOIMAGING

Projects have been funded by the European Union's Horizon 2020 research and innovation programme

Next steps: pilots for subdomains?




Pipeline annotations key:
biaingest ingest …
biaingest representations create …
biaingest representations convert-images …
! WIP
KEY
field_in_direction_of_arrow
virtual_field_in_downstream_path
If all *_uuid fields (links) have a corresponding downstream path and a _count path for pagination
Connection via UUID reference
ClassName
field1: field_type
field2: List[
EmbeddedClass
field3: field_type
Study
uuid: str
accession_id: str
title: str
description: str
release_date: Date
licence: LicenceType (enum)
keyword: Optional[List[str]]
acknowledgement: Optional[str]
funding_statement: Optional[str]
attribute: Optional[dict]
author: List[
Contributor
display_name: str
contact_email: Optional[str]
website: Optional[str]
ORCID: Optional[str]
rorid: Optional[str]
role: Optional[str]
address: Optional[str]
affiliation: List[
Organisation
display_name: str
website: Optional[str]
rorid: Optional[str]
address: Optional[str]
related_publication: List[
Publication
doi: str
pubmed_id: Optional[str]
title: str
release_date: Date
authors_name: str
description: Optional[str]
see_also: List[
ExternalLink
link: url
link_type: Optional[str]
description: Optional[str]
grant: List[
Grant
id: Optional[str]
funder: Optional[
FundingBody
id: Optional[str]
funder: display_name
ExperimentalImagingDataset
uuid: str
title_id: str
description: Optional[str]
example_image_uri: Optional[str]
image_count: int
file_reference_count: int
analysis_method: Optional[List[
ImageAnalysisMethod
method_description: str
features_analysed: str
correlation_method: Optional[List[
ImageCorrelationMethod
method_description: str
fiducials_used: str
transformation_matrix: str
/specimen_imaging_preparation_protocol [List]
SpecimenPreparationProtocol
uuid: str
title_id: str
method_description: str
signal_channel_information: Optional[List[
SignalChannelInformation
signal_contrast_mechanism_description: Optional[str]
channel_content_description: Optional[str]
channel_biological_entity: Optional[str]
submitted_in_study_uuid
/submitted_in_study
/experimental_imaging_component [List]
/acquisition_process [List]
preparation_protocol_uuid [List]
/preparation_protocol [List]
ImageAcquisition
uuid: str
fbbi_method_type: Optional[str]
title_id: str
method_description: str
imaging_instrument_description: str
image_acquisition_parameters: str
/specimen_growth_protocol[List]
SpecimenGrowthProtocol
uuid: str
title_id: str
method_description: str
submission_dataset_uuid
/submission_dataset
/file [List]
submission_dataset_uuid
/submission_dataset
/image [List]
acquisition_process_uuid [List]
/acquisition_process [List]
growth_protocol_uuid [List]
/growth_protocol [List]
biological_entity [List]
FileReference
uuid: str
size_in_bytes: int
format: str
path_name: str
uri: str
attribute: Optional[dict]
sumbission_dataset_type: Enum
ExperimentallyCapturedImage
uuid: str
attribute: dict
subject_uuid **
/subject
/subject_of [List]
Specimen
Biosample
uuid: str
title_id: str
description: str
experimental_variable_description: Optional[List[str]]
extrinsic_variables: Optional[List[str]]
intrinsic_variables: Optional[List[str]]
organism_classification: List[
Organism
common_name: str
scientific_name: str
ncbi_id: Optional[str]
sample_of_uuid [List]
/sample_of [List]
submitted_in_study_uuid
/submitted_in_study
/annotation_component [List]
submission_dataset_uuid
/submission_dataset
/file [List]
original_file_reference_uuid [List]
/original_file_reference [List]
/?
representation_of_uuid [List]
/representation_of
/representation [List]
ImageAnnotationDataset
uuid: str
title_id: str
example_image_uri: Optional[str]
image_count: int
file_reference_count: int
ImageRepresentation
uuid: str
image_format: str
file_uri: str
total_size_in_bytes: int
physical_size_x: Optional[float]
physical_size_y: Optional[float]
physical_size_z: Optional[float]
size_x: Optional[int]
size_y: Optional[int]
size_z: Optional[int]
size_c: Optional[int]
size_t: Optional[int]
attribute: Optional[dict]
image_view_settings: Optional[List[
RenderedView
z: Optional[int]
t: Optional[int]
channel_information: Optional[List[
Channel
label: Optional[str]
colormap_start: Optional[List[float]]
colormap_end: Optional[List[float]]
** Specimen was split off from ExperimentallCapturedImage(ECI) later in modelling to make creating links from e.g. biosamples backward easier. However, we need to discuss whether this changes the nature of the ECI-Specimen connection
/annotation_method [List]
AnnotationFile
uuid: str
size_in_bytes: int
format: str
path_name: str
uri: str
attribute: Optional[dict]
transformation_description: str
spatial_information: str
DerivedImage
uuid: str
attribute: dict
transformation_description: str
spatial_information: str
/representation_of
/representation [List]
representation_of_uuid
source_image_uuid [List]
/source_image[List]
/source_image_of [List]
/source_image[List]
/source_image_of [List]
source_image_uuid [List]
AnnotationMethod
uuid: str
title_id: str
source_dataset: List[uri]
annotation_criteria: str
annotation_coverage: str
method_type: AnnotationType
method_description: str
creation_process_uuid [List]
/creation_process[List]
/creation_process_of [List]

Breakout 2: Harmonizing Metadata Models: Bridging Efforts and Overcoming Challenges

- **Needs** – what do we all need?
- **Challenges –** what are the blockers?
- **What's possible –** what **might** we be able to do?
- **Next steps –** what **should** we try to do next?