Towards OMERO and ARC interoperability for RDM-compliant bio-image data

Project leads:

Niraj Kandpal & Andrea Schrader (University of Cologne)

Acknowledgement

de.NBI & elixir Germany

NFDI4Bioimage	Erick Martins Ratamero
DataPLANT	Timo Mühlhaus
CECAD	Christoph Möhl
CEPLAS	Heinrich Lukas Weil

Thanks to all supporting us on this journey to the **2nd Biohackathon Germany.**

Our Cologne NFDI4Bioimage team!
(Astrid Schauss, Monica Valencia-Schneider, Peter Zentis, Eik Dahms, Niraj Kandpal, Andrea Schrader)

All our online and on-site participants & supporters:

Mohamed Abdrabbou, Tom Boissonnet, Michele Bortolomeazzi, Jean-Marie Burel, Eik Dahms, Parisa Kakanj, Niraj Kandpal, Cristina Martins Rodrigues, Josh Moore, Kevin Schneider, Andrea Schrader, Torsten Stöter, Jens Wendt, Peter Zentis

Teaching & Documentation

Teaching

Aim: Getting into ARC & OMERO (newbie)

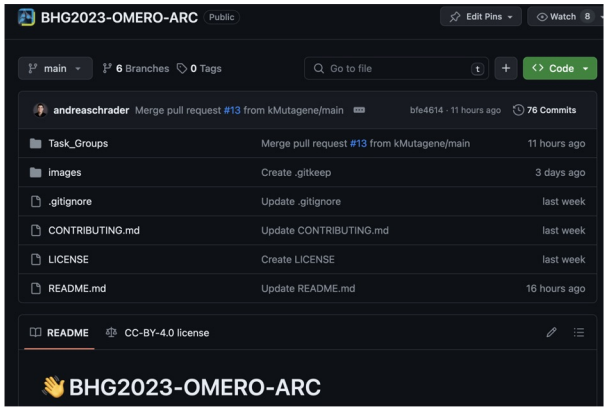
ARC - Annotated Research Context, DataPLANT

- Where to start? - Interview style
  ➡ Available material is overwhelming.
- Guidance needed.
- DataPLANT ARCitect: almost intuitively.
- Newbie issues collected & written.

➔ Insights for teaching
➔ Created an ARC for imaging + some metadata
➔ OMERO, OMERO-ARC

BHG2023 - Documentation

in progress, tbc
@NFDI4Bioimage
@GitHUB:

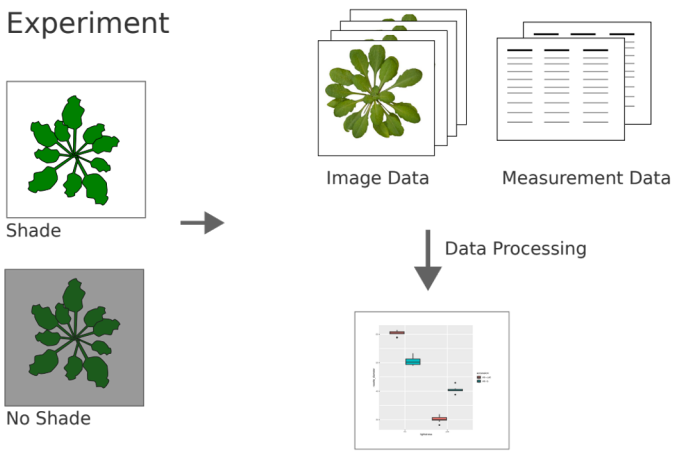


https://github.com/NFDI4BIOIMAGE/BHG2023-OMERO-ARC

Use Case Minimalistic ARC/OMERO

1. Minimalistic plant phenotyping dataset & annotated ARC

➡ **for teaching and testing tools**



Phenotypic Dataset
- Images
- Measured Data
- Scripts + Results

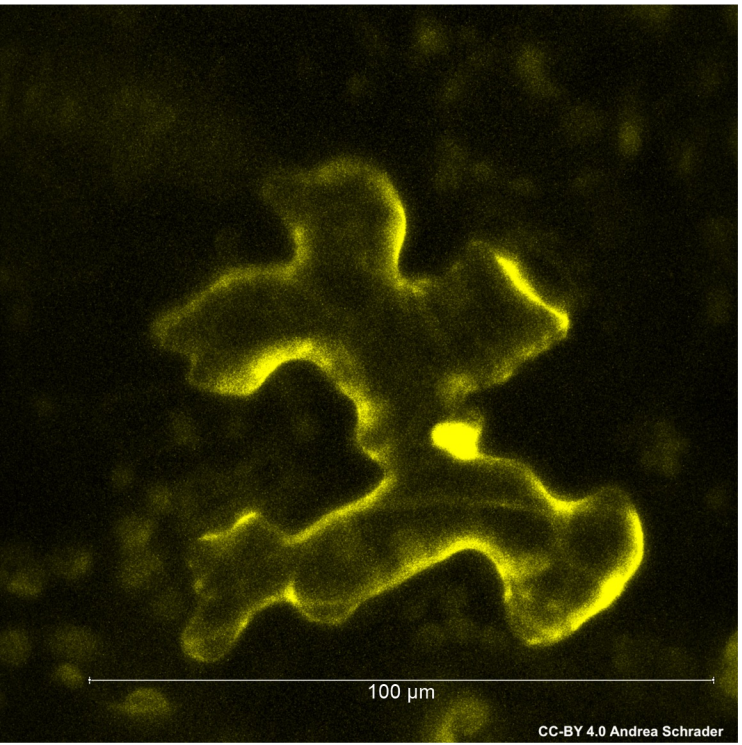
Teaching
- ARC creation
- OMERO/ARC workflows

Development
- OMERO/ARC workflow testing/validation

Use Case: teaching & development

2. First imaging ARC: CLSM z-stacks of leaf epidermal cells

https://git.nfdi4plants.org/natural-variation-and-evolution/microscopy_collection/map-by-seq_clsm-stacks



First imaging ARC to get started

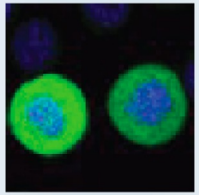
Vendor specific image container vs. extracted open formats in ARC:

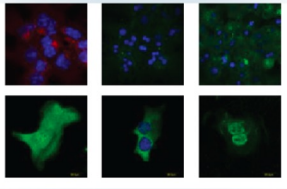
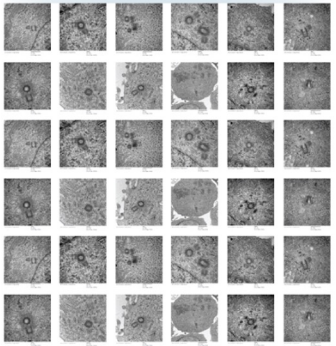
- Leica LIF files: many images
- Problem for metadata annotation per file
- ARC specs: add open formats or workflows for conversion

MIPs (maximum intensity projections) in vendor software > important tech. metadata can be lost

- Use OMERO
  - ➔ **OMERO-ARC roundtrip!**
  - ➔ **File to image mapping required**

File-to-image mapping



Images

Import

Files

DeltaVision Fileset

HeLaCell.dv
HeLaCell.dv.log

DICOM Fileset

ankle.dcm

Aperio SVS Fileset

HeLaCell.svs

InCell 2000 Fileset

INMAC384-DAPI-CM-eGFP_59223_1.xdce
A - 1(fld 1 wv Cy5 - Cy5).tif
A - 1(fld 1 wv DAPI - DAPI).tif
A - 1(fld 1 wv FITC - FITC).tif
A - 2(fld 1 wv Cy5 - Cy5).tif
A - 2(fld 1 wv DAPI - DAPI).tif
A - 2(fld 1 wv FITC - FITC).tif
A - 3(fld 1 wv Cy5 - Cy5).tif
A - 3(fld 1 wv DAPI - DAPI).tif
A - 3(fld 1 wv FITC - FITC).tif
A - 4(fld 1 wv Cy5 - Cy5).tif
…

https://omero.readthedocs.io/en/stable/developers/ImportFS.html

More Use Cases

3. **Multimodal imaging ARC: LightSheet + MRI imaging of rat brain**
   - Case: N files ~ 1 image
   - 4 LS stacks of 4k z-slices, 2 channel => 32k tiff files, each 8MB
   - 4 MRI stacks of 25 z-slices
   - ISA annotation with >32k rows
   - Total 250GB ARC

4. **Spatial Transcriptomics ARC (Medulloblastoma) - multimodal data**
   - ARC with a reproducible cwl image processing workflow
   - https://git.nfdi4plants.org/michele.bortolomeazzi/mben_resolve

5. **Transmission Electron microscopy ARC: HeLa Cell mitochondria - annotation focus**
   - detailed description of sample preparation (WIP)
     - ➢ 8 Swate tables -> for templates creation
     - ➢ some user specific terms, due to lack of fitting ontology terms
     - ➢ traceable provenance of images:
       cell culture sample > epon block > section > image
   - defined an essential set of required image acquisition parameters

RDM - compliant metadata

REMBI = **RE**commended **M**etadata for **B**iological **I**mages (https://rdcu.be/dtAOL)

Aim: Enable REMBI-compliant metadata in imaging ARCs e.g. for submission to **repositories**.

Mapping based on REMBI implementation at BioImage Archive
https://www.ebi.ac.uk/bioimage-archive/

Example for a non-intuitive solution (REMBI-extractable, SWATE annotation)

Factor[genotype]	…	Factor[age]	…	Parameter[REMBI - intrinsic variables]
wild type	…	7	…	Factor[age];Factor[genotype]
my_mutant	…	88	…	Factor[age];Factor[genotype]

➔ **Mapping all mandatory attributes to ARC accomplished**
➔ Added a REMBI table for **Ca2+ imaging**

Case study: IDR templates to ARC

IDR (**Image Data Resource**) templates files

- https://github.com/IDR/idr0000-lastname-example/archive/master.zip
- Examples of completed templates of other studies can be found at https://github.com/IDR/idr-metadata/

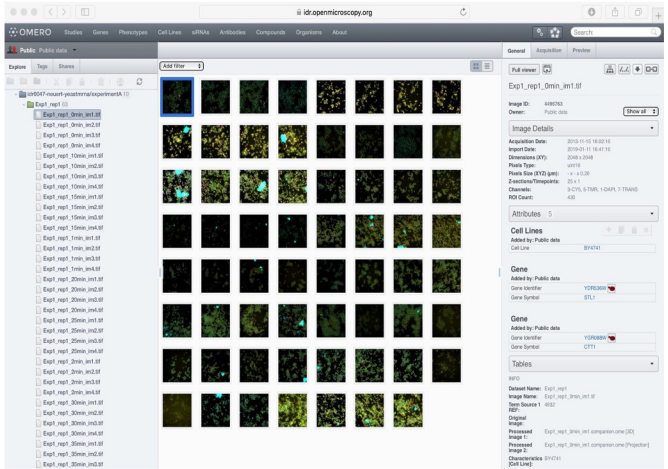
Convert to ARC

ARC

Import to IDR

Study metadata

Assay data & metadata



OMERO

Adapted from: https://www.nfdi4plants.de/nfdi4plants.knowledgebase/docs/guides/isa_FileTypes.html

ARC to OMERO transfer

(*)

ARC

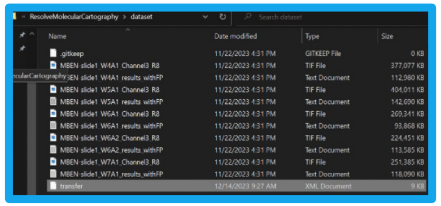
Prototype tool for transfer of image data and metadata from ARC to OMERO

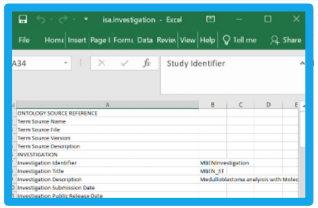
OMERO

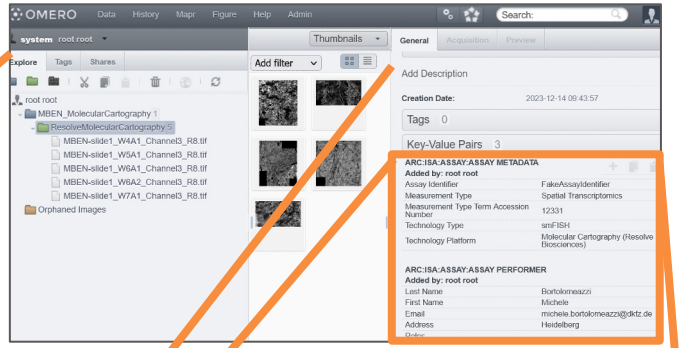
(**)

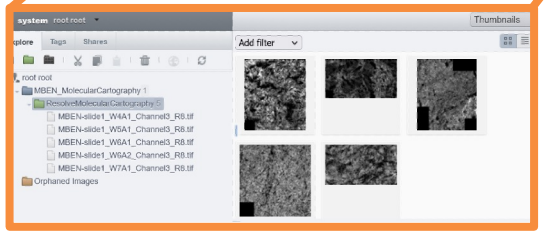
Study metadata

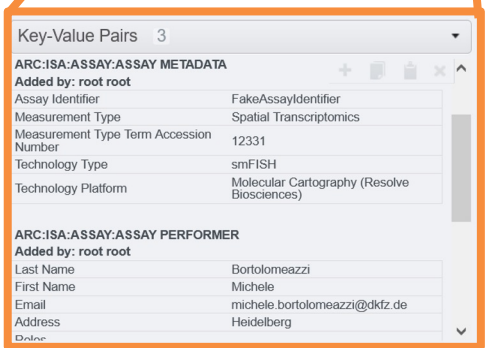
Assay data & metadata











(*) https://www.fz-juelich.de/de/ibg/ibg-4/aktuelles/meldungen/arc-overview.jpg | (**) modified from (TA3) - DataPLANT

OMERO to ARC transfer

(*) ARC

omero-arc plugin

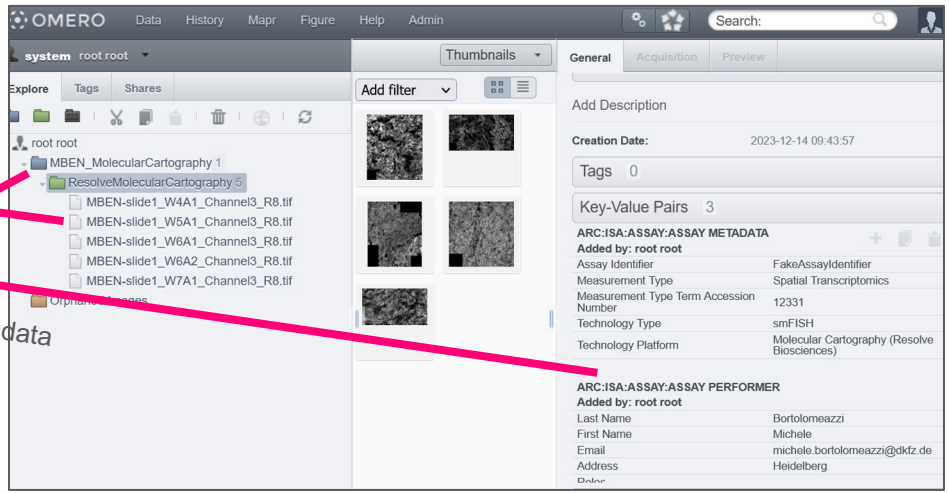
OMERO

```
MBEN_MolecularCartography/
├── assays
│   └── FakeAssayIdentifier
│       ├── dataset
│       │   ├── MBEN-slide1_W4A1_Channel3_R8.tif
│       │   ├── MBEN-slide1_W5A1_Channel3_R8.tif
│       │   ├── MBEN-slide1_W6A1_Channel3_R8.tif
│       │   ├── MBEN-slide1_W6A2_Channel3_R8.tif
│       │   └── MBEN-slide1_W7A1_Channel3_R8.tif
│       ├── isa.assay.xlsx
│       ├── protocols
│       │   ├── ImageID687_metadata.json
│       │   ├── ImageID688_metadata.json
│       │   ├── ImageID689_metadata.json
│       │   ├── ImageID690_metadata.json
│       │   └── ImageID691_metadata.json
│       └── README.md
├── isa.investigation.xlsx
├── runs
├── studies
│   └── mben_molecularcartography
│       ├── isa.study.xlsx
│       ├── protocols
│       ├── README.md
│       └── resources
└── workflows
```

Image Data

Assay Metadata

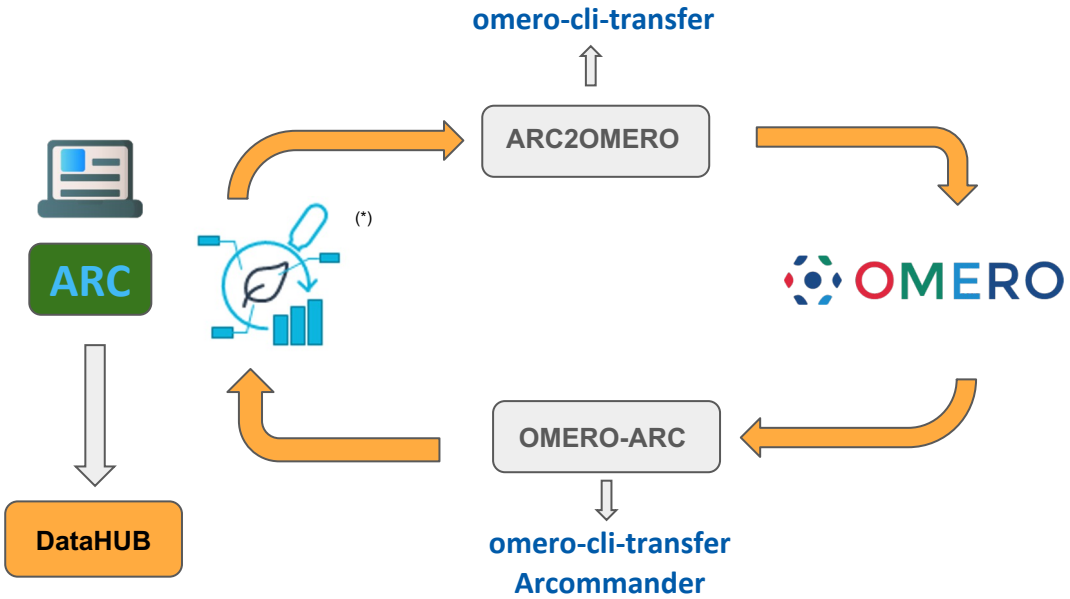
Study



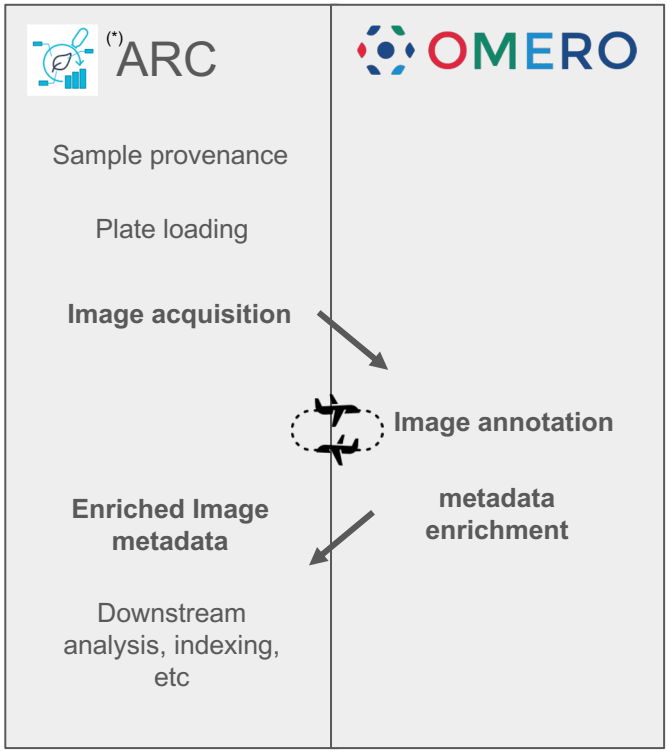
omero-arc plugin → omero-cli-transfer

(*) https://www.fz-juelich.de/de/ibg/ibg-4/aktuelles/meldungen/arc-overview.jpg

Workflow - Round Trip



Metadata enrichment process (Screening example):



Completes the full circle

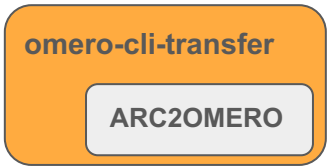
(*) https://www.fz-juelich.de/de/ibg/ibg-4/aktuelles/meldungen/arc-overview.jpg

Image to file Mapping

Future Work

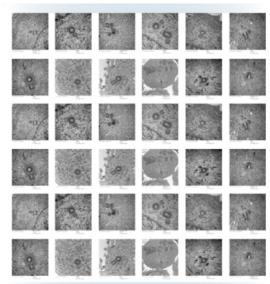
1) ARC2OMERO Prototype to a plugin

Refactor the ARC2OMERO prototype as an omero-cli-transfer plugin.

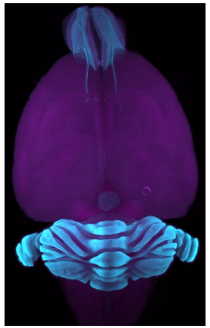


2) Image to a File Mapping

Implement the image to file mapping for the metadata enrichment workflow



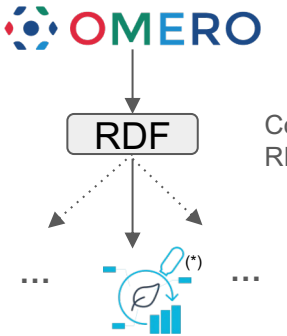
```
INMAC384-DAPI-CM-eGFP_59223_1.xdce
A - 1(fld 1 wv Cy5 - Cy5).tif
A - 1(fld 1 wv DAPI - DAPI).tif
A - 1(fld 1 wv FITC - FITC).tif
A - 2(fld 1 wv Cy5 - Cy5).tif
A - 2(fld 1 wv DAPI - DAPI).tif
A - 2(fld 1 wv FITC - FITC).tif
A - 3(fld 1 wv Cy5 - Cy5).tif
A - 3(fld 1 wv DAPI - DAPI).tif
A - 3(fld 1 wv FITC - FITC).tif
A - 4(fld 1 wv Cy5 - Cy5).tif
…
```



3) Complete use case implementations

- Multimodal Imaging data
- REMBI to ARC
- Teaching material
- …

4) OMERO to RDF



Convert the OMERO2ARC workflow to use RDF using the omero-rdf exporter tool

(*) https://www.fz-juelich.de/de/ibg/ibg-4/aktuelles/meldungen/arc-overview.jpg

Thank You