Combining the BIDS and ARC Directory Structures for Multimodal Research Data Organization

Torsten Stöter[1], Tobias Gottschall[1] (orcid: 0000-0003-3001-1491), Andrea Schrader[2] (orcid: 0000-0002-3879-7057), Peter Zentis[3] (orcid: 0000-0002-6999-132X), Monica Valencia-Schneider[4] (orcid: 0000-0003-3430-2683), Niraj Kandpal[3,4] (orcid: 0009-0007-5101-4786), Werner Zuschratter[1] (orcid: 0000-0002-9845-6393), Astrid Schauss[3] (orcid: 0000-0002-6658-2192), Timo Dickscheid[5] (orcid: 0000-0002-9051-3701), Timo Mühlhaus[6] (orcid: 0000-0003-3925-6778), Dirk von Suchodoletz[7] (orcid: 0000-0002-4382-5104)

[1] Combinatorial NeuroImaging Core Facility, Leibniz Institute for Neurobiology, Magdeburg, Germany; [2] Data Science and Management, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Cologne, Germany; [3] Imaging Facility, Cellular Stress Responses in Aging-Associated Diseases (CECAD), Cologne, University of Cologne, Germany; [4] Regional Computing Centre (RRZK), University of Cologne, Cologne, Germany; [5] Big Data Analytics Working Group, Forschungszentrum Jülich, Germany; [6] Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Germany; [7] Albert-Ludwigs-Universität Freiburg, Germany

BIDS microscopy template

```
dataset
├── sub-<label>
│   └── ses-<label>
│       └── micr
│           ├── sub-<label>_ses-<label>_sample-<label>_stain-<label>_chunk-<index>_<suffix>.<extension>
│           └── sub-<label>_ses-<label>_sample-<label>_stain-<label>_chunk-<index>_<suffix>.json
├── dataset_description.json
├── participants.json
├── participants.tsv
├── samples.json
└── samples.tsv
```

```
sample_id       participant_id   sample_type
sample-<label>  sub-<label>      tissue
```

```
{ "PixelSize": [0.07, 0.07],
  "PixelSizeUnits": "um",
  "BodyPart": "CSPINE",
  "SampleEnvironment": "ex vivo" }
```

```
participant_id  species              strain                 strain_rrid
sub-<label>     rattus norvegicus    SPRD Sprague-Dawley    RRID:RGD_1566457
```

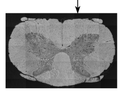

FIGURE 1 strict and clearly defined file format (png, tiff, […], nii, ome.zarr), directory structure (e.g. sub-, code, derivatives), naming rules (e.g. prefix: “sub-200928“; suffix: “_FLUO”; microscopy: “micr” directory) and metadata (tsv & json files), little flexibility. BIDS standard version 1.8.0 [5] for microscopy data [6] files

Interdisciplinary collaboration and integration of large and diverse datasets are becoming increasingly important. Answering complex research questions requires combining and analysing multimodal datasets. Research data management follows the FAIR principles [1] making data findable, accessible, interoperable, and reusable. However, there are challenges in capturing the entire research cycle and contextualizing data according, not only for the DataPLANT [2] and NFDI4BIOIMAGE [3] communities. To address these challenges, DataPLANT developed a data structure called Annotated Research Context (ARC) [4]. The Brain Imaging Data Structure (BIDS) [5] originated from the neuroimaging community extended for microscopic image data [6]. Both concepts provide standardised and file system based data storage structures for organising and sharing research data accompanied with metadata. We exemplarily compare the ARC and BIDS designs and propose structural and metadata mapping.

ARC standard structure

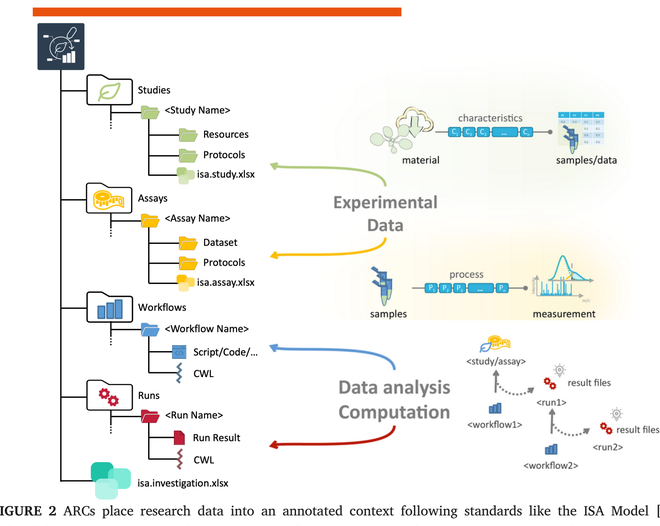

FIGURE 2 ARCs place research data into an annotated context following standards like the ISA Model [7] with a defined directory structure (studies, assays, workflows, runs) & mandatory investigation / study / assay metadata files. ARC Commander [8] and SWATE [9] support with Git / GitLab commands and (ontology-based) annotation.

Electronic labbook

Clearing protocol with TO-PRO-1 and tert-butanol
- transcradial perfusion with 1x PBS (phosphate-buffered saline) (pH 7.4) as pre-rinse with 4% PFA (perfluoroalkoxy) (pH 7.4) as fixative
- 4h post-fixation in 4% PFA at 4°C
- storage of the tissue in 1x PBS until the start of clearing
- wash in 1x PBS --> 3x 30min

Permeabilization solution (2 d)
- (1.5% Goat serum,
- 0.5% Triton X-100,
- 0.5 mM methyl-betacyclodextrin,
- 0.2% trans-1-acetyl-4-hydroxy-proline,
- 0.05% sodium acide in 1x PBS)
at 37°C, slow rotation (in 20 ml rolled rim tubes)

Incubation solution (5 d)
- (= permeabilization solution plus TO-PRO-1, 1:1000)
at 37°C, slow rotation, dark!

Washing in wash solution (3x 2 h)
- (1.5% Goat serum,
- 0.5% Trition-X-100,
- 0.05% Sodium acide in 1x PBS)
at 37°C, slow rotation, dark

OME-XML Metadata (from ome.tif File)

```
ImageDescription: <?xml version="1.0" encoding="UTF-8"?><OME xmlns="http://
www.openmicroscopy.org/Schemas/OME/2008-02" xmlns:ca="http://www.
openmicroscopy.org/Schemas/CA/2008-02" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance" xsi:schemaLocation="http://www.openmicroscopy.org/
Schemas/OME/2008-02 http://www.openmicroscopy.org/Schemas/OME/2008-02/ome.
xsd" UUID="urn:uuid:251F06D7-A0F4-471C-84E5-A908DE0AFE37">
<Experimenter ID="Experimenter:Tobias_Gottschall">
<FirstName>Tobias</FirstName>
<LastName>Gottschall</LastName>
<Email>tobias.gottschall                    </Email>
<Institution>LIN</Institution>
</Experimenter>
<Group ID="urn:lsid:imspector.group.id:Group:SP-ELMI" Name="SP-ELMI"/>
<Image ID="Image:100DD12B-DD11-4E7C-AACD-4C14334C7B08"
Name="13-05-56_Standard_gerbil_5418_TOPRO-3_AF_4x_2LS_2mue_nDF_4x5_20percen
t_150ms_UltraII[00 x 00]_C00_xyz-Table Z0000.ome.tif"
DefaultPixels="Pixels:100DD12B-DD11-4E7C-AACD-4C14334C7B08">
<CreationDate>2020-12-12T12:33:26</CreationDate>
<Description>not_specified</Description>
<Pixels ID="Pixels:100DD12B-DD11-4E7C-AACD-4C14334C7B08"
DimensionOrder="XYZCT" PixelType="uint16" BigEndian="false" SizeX="2048"
SizeY="2048" SizeZ="3585" SizeT="1" SizeC="2" PhysicalSizeX="1.625000"
PhysicalSizeY="1.625000" PhysicalSizeZ="2.000000">
<TiffData FirstZ="0" FirstC="0">
<UUID
FileName="13-05-56_Standard_gerbil_5418_TOPRO-3_AF_4x_2LS_2mue_nDF_4x5_20pe
rcent_150ms_UltraII[00 x 00]_C00_xyz-Table Z0000.ome.
tif">urn:uuid:251F06D7-A0F4-471C-84E5-A908DE0AFE37</UUID>
</TiffData>
..
<Filter0 ExcitationWL="500" EmissionWL="535"/>
<Filter1 ExcitationWL="642" EmissionWL="661"/>
</OME>
```

BIDS microscopy full example

```
.
├── atlas_transform_2d-to-3d
│   ├── dataset_description.json
│   ├── README
│   ├── samples.tsv
│   ├── sourcedata
│   │   ├── protocol
│   │   │   ├── clearing_protocol_SD832_full.json
│   │   │   └── Miltenyi_UltraMicroscopeII_manuel.pdf
│   │   ├── sub-200928
│   │   │   └── micr
│   │   │       ├── 200928_[00 x 00]_C00.ome.tif
│   │   │       ├── 200928_[00 x 00]_C01.ome.tif
│   │   │       └── ..
│   │   └── sub-THG5418
│   │       └── micr
│   │           ├── THG5418_[00 x 00]_C00.ome.tif
│   │           ├── THG5418_[00 x 00]_C01.ome.tif
│   │           └── ..
│   ├── sub-200928
│   │   └── micr
│   │       ├── sub-200928_sample-tissue_FLUO.json
│   │       └── sub-200928_sample-tissue_FLUO.ome.zarr
│   ├── sub-THG5418
│   │   └── micr
│   │       ├── sub-THG5418_sample-tissue_FLUO.json
│   │       └── sub-THG5418_sample-tissue_FLUO.ome.zarr
│   ├── derivatives
│   │   ├── 3D-gerbiLIN-atlas_alpha
│   │   │   ├── 3D-gerbiLIN-atlas_alpha_template.nii
│   │   │   ├── 3D-gerbiLIN-atlas_alpha_annotations.nii
│   │   │   └── 3D-gerbiLIN-atlas_alpha_labels.csv
│   │   ├── downsampled
│   │   │   ├── sub-200928
│   │   │   │   └── micr
│   │   │   │       ├── sub-200928_sample-tissue_res-17x_FLUO.json
│   │   │   │       └── sub-200928_sample-tissue_res-17x_FLUO.ome.zarr
│   │   │   └── sub-THG5418
│   │   │       └── micr
│   │   │           ├── sub-THG5418_sample-tissue_res-17x_FLUO.json
│   │   │           └── sub-THG5418_sample-tissue_res-17x_FLUO.ome.zarr
│   │   ├── metricCalc
│   │   │   ├── 201205_LSstandardMouse
│   │   │   ├── 20201005_072702_Rat
│   │   │   └── 5679_LSgerbil
│   │   └── publication
│   │       ├── Gottschall_2022_MSC_gerbil-atlas_2d-to-3d.pdf
│   │       └── gerbil-3d-atlas-development_2021_pre-registration.pdf
│   └── code
│       ├── example
│       │   └── reworked_gerbil_itk_snap_label.label
│       ├── imgs
│       │   └── imageJ_background_grey-value.png
│       ├── README.md
│       ├── run_main.py
│       └── src
│           └── 2D_multi-registration_transformation_v.py
```

FIGURE 3 (bottom-left) export of the OME-XML metadata from an ome.tif file automatically created by the control software of BioTec UltraMicroscope II; (right) manually created dataset as complete as possible of the development of a 2D to 3D Mongolian gerbil brain atlas using light sheet microscopy [7]

ARC conversion of BIDS example

```
.
├── atlas_transform_2d-to-3d
│   ├── isa.investigation.xlsx
│   ├── assays
│   │   ├── atlas_transform_2d-to-3dDownsampled
│   │   │   ├── dataset
│   │   │   │   ├── sub-200928
│   │   │   │   │   └── micr
│   │   │   │   │       ├── sub-200928_sample-tissue_res-17x_FLUO.json
│   │   │   │   │       ├── sub-200928_sample-tissue_res-17x_FLUO.ome.zarr
│   │   │   │   │       └── ..
│   │   │   │   └── sub-THG5418
│   │   │   │       └── micr
│   │   │   │           ├── sub-THG5418_sample-tissue_res-17x_FLUO.json
│   │   │   │           ├── sub-THG5418_sample-tissue_res-17x_FLUO.ome.zarr
│   │   │   │           └── ..
│   │   │   ├── isa.assay.xlsx
│   │   │   ├── protocols
│   │   │   │   ├── clearing_protocol_SD832_full.json
│   │   │   │   └── Miltenyi_UltraMicroscopeII_manuel.pdf
│   │   │   └── README.md
│   │   ├── atlas_transform_2d-to-3dRegular
│   │   │   ├── dataset
│   │   │   │   ├── sub-200928
│   │   │   │   │   └── micr
│   │   │   │   │       ├── sub-200928_sample-tissue_FLUO.json
│   │   │   │   │       ├── sub-200928_sample-tissue_FLUO.ome.zarr
│   │   │   │   │       └── ..
│   │   │   │   └── sub-THG5418
│   │   │   │       └── micr
│   │   │   │           ├── sub-THG5418_sample-tissue_FLUO.json
│   │   │   │           ├── sub-THG5418_sample-tissue_FLUO.ome.zarr
│   │   │   │           └── ..
│   │   │   ├── isa.assay.xlsx
│   │   │   ├── protocols
│   │   │   │   ├── clearing_protocol_SD832_full.json
│   │   │   │   └── Miltenyi_UltraMicroscopeII_manuel.pdf
│   │   │   └── README.md
│   │   └── atlas_transform_2d-to-3dSourceData
│   │       ├── dataset
│   │       │   ├── sub-200928
│   │       │   │   └── micr
│   │       │   │       ├── 200928_[00 x 00]_C00.ome.tif
│   │       │   │       ├── 200928_[00 x 00]_C01.ome.tif
│   │       │   │       └── ..
│   │       │   └── sub-THG5418
│   │       │       └── micr
│   │       │           ├── THG5418_[00 x 00]_C00.ome.tif
│   │       │           ├── THG5418_[00 x 00]_C01.ome.tif
│   │       │           └── ..
│   │       ├── isa.assay.xlsx
│   │       ├── protocols
│   │       │   ├── clearing_protocol_SD832_full.json
│   │       │   └── Miltenyi_UltraMicroscopeII_manuel.pdf
│   │       └── README.md
│   ├── runs
│   │   ├── 3D-gerbiLIN-atlas_alpha
│   │   │   ├── 3D-gerbiLIN-atlas_alpha_annotations.nii
│   │   │   ├── 3D-gerbiLIN-atlas_alpha_labels.csv
│   │   │   └── 3D-gerbiLIN-atlas_alpha_template.nii
│   │   └── metricCalc
│   │       ├── 201205_LSstandardMouse
│   │       ├── 20201005_072702_Rat
│   │       └── 5679_LSgerbil
│   ├── studies
│   │   └── atlas_transform_2d-to-3dStudy
│   │       ├── isa.study.xlsx
│   │       ├── protocols
│   │       │   └── clearing_protocol_SD832_full.json
│   │       ├── README.md
│   │       └── resources
│   │           ├── example
│   │           │   └── reworked_gerbil_itk_snap_label.label
│   │           ├── imgs
│   │           │   └── imageJ_background_grey-value.png
│   │           └── publication
│   │               ├── Gottschall_2022_MSC_gerbil-atlas_2d-to-3d.pdf
│   │               └── gerbil-3d-atlas-development_2021_pre-registration.pdf
│   └── workflows
│       ├── README.md
│       ├── run.cwl
│       ├── run.yml
│       └── src
│           ├── 2D_multi-registration_transformation_v.py
│           ├── run_main.py
│           └── stats_testing.py
```

isa.assay.xlsx SWATE spreadsheet

Source Name	Characteristic [sample type]	Parameter [Instrument Model]	Component [Perfusion]	Characteristic [fixative]	Term Source REF (CHEBI:50913
sub-200928	Tissue	Light Microscopy	phosphate buffered saline solution	perfluoroalkyl substance	CHEBI
sub-THG5418	Tissue	Light Microscopy	phosphate buffered saline solution	perfluoroalkyl substance	CHEBI

Term Accession Number (CHEBI:50913	Characteristic [Time]	Characteristic [temperature day]	Characteristic [ratio]
ibrary.org/obo/CHEBI_172397	4.00 hour	Celsius based unit	percentage based unit
ibrary.org/obo/CHEBI_172397	4.00 hour	Celsius based unit	percentage based unit

isa.study.xlsx SWATE spreadsheet

Source Name	Sample Name	Characteristic [Animal]	Characteristic [Age]
sub-200928	sample-tissue	Mouse	14.00 week
sub-THG5418	sample-tissue	Mongolian Gerbil	16.00 week

isa.investigation.xlsx ARC structure metadata

ONTOLOGY SOURCE REFERENCE	
INVESTIGATION	
Investigation Identifier	atlas_transform_2d-to-3dInvestigation
Investigation Title	atlas_transform_2d-to-3d
Investigation Description	
Investigation Submission Date	
Investigation Public Release Date	
INVESTIGATION PUBLICATIONS	
INVESTIGATION CONTACTS	
Investigation Person Last Name	Gottschall
Investigation Person First Name	Tobias
Investigation Person Mid Initials	
Investigation Person Email	tobias.gottschall@lin-magdeburg.de
Investigation Person Phone	
Investigation Person Fax	
Investigation Person Address	
Investigation Person Affiliation	Leibniz Institute for Neurobiology
Investigation Person Roles	
Investigation Person Roles Term Accession Number	
Investigation Person Roles Term Source REF	
STUDY	
Study Identifier	atlas_transform_2d-to-3dStudy
Study Title	
Study Description	
Study Submission Date	
Study Public Release Date	
Study File Name	atlas_transform_2d-to-3dStudy/isa.study.xlsx
STUDY DESIGN DESCRIPTORS	
STUDY PUBLICATIONS	
STUDY FACTORS	
STUDY ASSAYS	
Study Assay Measurement Type	
Study Assay Measurement Type Term Accession Number	
Study Assay Measurement Type Term Source REF	
Study Assay Technology Type	
Study Assay Technology Type Term Accession Number	
Study Assay Technology Type Term Source REF	
Study Assay Technology Platform	
Study Assay File Name	atlas_transform_2d-to-3dAssay/isa.assay.xlsx
STUDY PROTOCOLS	
STUDY CONTACTS	

FIGURE 4 (left) representation of the directory structure of a valid dataset in the ARC standard manually converted from BIDS directory (compare **FIGURE 3** right); (top-right) small section of clearing_protocol_SD832_full.json transfered from electronic labbook and OME-XML metadata to isa.assay.xlsx file using the SWATE tool [8]; section of BIDS samples.tsv and clearing_protocol_SD832_full.json file transfered to isa.study.xlsx

LEGEND FIGURE 3 & 4 (blue background) dataset components related to the creation, transformation, or analysis of the data, such as source code or parameters necessary for the analysis; (light red background) study data itself as well as its descriptive metadata; (yellow background) analysis results, figures created from the data, and components essential for the publication of the results; (triple color background) file content must be cut and reassigned to multiple other files; (part of figures below grey bar) these contents are part of the dataset but not exclusive or mandatory for the BIDS standard; (part of figures below blue bar) valid recommended or mandatory structures of the BIDS standard; (part of figures below red bar) valid recommended or mandatory structures of the ARC standard

The basic conclusion can be that a transformation from ARC to BIDS and vice versa is possible without loss of information. The flexibility of the ARC standard allows for fast and efficient conversion or embedding of a BIDS dataset, in the embedding case losing all relevant advantages of the ARC standard. For a conversion the ISA-XLSX [9] metadata files should be filled with metadata, supported by the ARC Commander [10] and SWATE tool [8]. Converting from ARC to BIDS, means in the simplest case merging information from the XLSX metadata files into mandatory subject specific .json sidecar files. In addition measurement data files need splitting into individual subject data files (for example, converting .fastq files to .csv files to ome.zarr with additional .json metadata files). Also relative path specifications in the source code files must be adjusted. To follow the BIDS standard, conversion is only possible if the used measurement method has already been specified, since integration of additional methods through community based BIDS Extension Proposals is time consuming.

In summary, the benefit of the transformation of one of these formats into the other is open for debate. It is unlikely that a transformation would allow format-specific tools created for one dataset format to be used with a transformed dataset, if they even exist. The biggest benefit could be to be able to make multimodal studies and to allow further use of the data in larger systems such as OMERO [11].

[1] Wilkinson, M. D. et al., 2016, https://doi.org/10.1038/sdata.2016.18
[2] Suchodoletz, D. v. et al., 2021, https://doi.org/10.17192/bfdm.2021.2.8335 (german)
[3] NFDI4BIOIMAGE, 2023, https://nfdi4bioimage.de/
[4] DataPLANT community, 2023, https://doi.org/10.5281/zenodo.8302662
[5] Gorgolewski, K. J. et al., 2016, https://doi.org/10.1038/sdata.2016.44
[6] Bourget, M.-H. et al., 2022, https://doi.org/10.3389/fnins.2022.871228
[7] Gottschall, T., & Budinger, E., Development of an 3D atlas for Mongolian gerbil, (unpublished)
[8] Sansone, S.-A. et al., 2016, https://doi.org/10.5281/zenodo.163640
[9] Weil, H. L. et al., 2022, https://github.com/nfdi4plants/arcCommander
[10] Frey, K. et al., 2022, https://github.com/nfdi4plants/Swate
[11] Allan, C. et al., 2012, https://doi.org/10.1038/nmeth.1896

We acknowledge the support of DataPLANT and NFDI4BIOIMAGE, both funded by the German Research Foundation (DFG) within the framework of the NFDI – project numbers: 442077441 and 501864659; and CEPLAS, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within Germany’s Excellence Strategy – EXC-2048/1 – project number 390686111.