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

data in proprietary format

contains the representation of an image in a technical way

contains

access

Bio-Formats

Proprietary file formats translator: opens proprietary microscopy image data and maps/associates metadata to the *OME Data Model* (real time conversion) and offers different output formats.

access

Technical Metadata

Description of hardware (microscope, other devices), software and settings (default and user-defined).

Experimental Metadata

Contextual information, like organism under consideration.

Sample Preparation Metadata

Expressions to describe the properties of your sample (composition) and its preparation (protocols, devices, reagents, ...).

tool can be used to edit/ gather this metadata

tool can be used to edit/ gather this metadata

tool reads out the metadata automatically AND can be used for completion and correction

tool reads out the metadata automatically OR can be used to gather this metadata

basis model

possible conversion formats

specific standard

Cloud-able formats

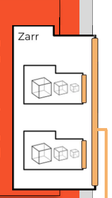
OME-NGFF

OME's specification for next generation file formats (NGFF) for storing Bioimage data in the cloud. Zarr is such a format. (1)

OME-Zarr

Adapted implementation of cloud-able data format **Zarr** to store and stream image data and metadata based on OME Data Model.

Zarr: A format for the storage of chunked, compressed, N-dimensional arrays. Optimal for data storage in the cloud due to the possibility of parallelised, chunk-based data access.



written in concrete format

equivalent metadata

Data Model Microscopy Data

OME Data Model

OME's specification for storing data on biological imaging. The model includes image parameters as well as extensive metadata.

4DN-BINA-OME (NBO) Data Model

Community-driven (4D Nucleome Imaging Working Group, BINA Quality Control and Data Management Working Group and QUAREP-LiMi) extension and further development of OME Data Model. (2)

basis model

Tools to edit/gather Metadata (Examples)

Micro-Meta-App

Visualization of microscope components; input forms.

Output: JSON

MDEmic:

OMERO integrated; input forms; assists in using *ontologies*.

Output: *OME-XML*, CSV, TSV, *ISA-tab*

Swate

Is a Workflow Annotation Tool for Excel (6); assists in using *ontologies*.

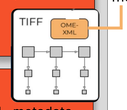
Output: XLS, XLSX

files with metadata description

Classical Formats

OME-XML

Vendor-neutral file format for bioimage data, with an emphasis on metadata using OME Data Model. It can be used as a data file format in its own right, or as a way of encoding metadata within a TIFF or BigTIFF file (for which purpose there is the OME-TIFF specification). (17)



OME-Tiff

A specification of how to embed OME-XML metadata within a TIFF or BigTIFF image file.

possible Linked Data format

JSON-LD

„JavaScript Object Notation for Linked Data". Properties can be URIs or blank nodes. JSON-LD is a valid JSON and therefore compatible with applications that read the JSON data format.

written in concrete format → RDF serialization

formal semantic of

querying

SPARQL

SQL-like language for querying RDF graph sets. It is also a simple protocol for issuing queries and receiving results over HTTP.

visualized in

RDF (Resource Description Framework)

A standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed (4).
Resources and relationships are named with *URIs*. A statement is represented by a *triple*.

Labeled Graph

Mental model of RDF, e.g., as a visual representation.

is a core concept of

based on

Triple

A statement, consisting of two resources (Subject,Object) and a relationship between them (predicate). (7)

Subject → predicate (or verb) → Object

RDF Triplestore

Type of graph database that stores data as a network of objects and uses inference to uncover new information out of existing relations.

based on

Knowledge Graph

Interlinked descriptions of entities – objects, events, situations or abstract concepts – while also encoding the semantics underlying the used terminology.

is a

is a

is a core concept of

Ontology

A conceptual definition how to represent the knowledge in a discipline with specific terms (controlled vocabularies) and the relationships between the terms.
In addition, an ontology defines the links to equivalent terms (synonyms) in another discipline. Ontologies are specified so that machines can read and interpret this knowledge.

Controlled Vocabulary

Predefined list of terms, but without context information or relationships.

OWL

Web Ontology Language (OWL) is a *Semantic Web* language designed to represent rich and complex knowledge about things, groups of things, and relations between things (3); machine-readable.

based technically on

Smart Data

Application of reasoning and inference for intelligent automated evaluation.
Required: RDF, Ontology

Linked Data

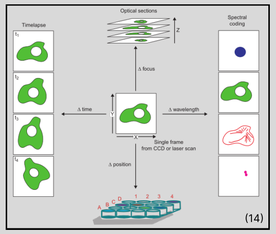
Collection of interrelated machine-readable datasets on the Web. (5)
Required: RDF

based on

URI

Uniform Resource Identifier, a standard format for identifiers on the Internet.
→ is a global identifier
Synonyms: URL, web address, IRI

Physical Arrangement:

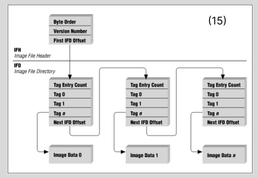


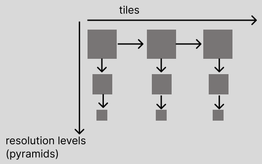
Data are stored as linear sequences of 2D tiles (arrays of numbers).

XYCZ: z1 z1 z2 z2 z3 z3

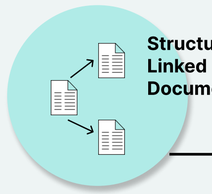
XYZC: z1 z2 z3 z1 z2 z3

Local Organization of a TIFF-File





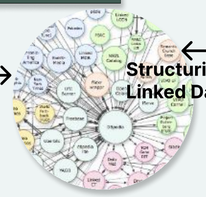
Internet of Things: Data on the internet is connected by linking files (documents, etc.) between locations on the internet.



Structuring as Linked Documents

Semantic Web: Based on content linking instead of linked files; structures data on the internet semantically (context based), identifies resources, and shows the relationships between resources which are formulated as *Linked Data*.

technical shift



Structuring as Linked Data

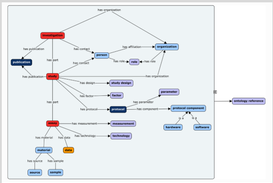
ISA Framework

Built around the 'Investigation' (the project context), 'Study' (a unit of research) and 'Assay' (analytical measurement) data model and serializations (tabular, JSON and *RDF*), the ISA framework helps you to provide rich description of the experimental metadata (i.e. sample characteristics, technology and measurement types, sample-to-data relationships) so that the resulting data and discoveries are reproducible and reusable. (8)

ISA Model (Data Model)

Consists of three core entities to capture experimental metadata including ontology reference:
Investigation: information that are needed to understand the overall goals and means used in an experiment.
Study: information on the subject under study, its characteristics and any treatments applied; contextualising information for one or more assay.
Assay: represent a test performed either on material taken from a subject or on a whole initial subject, producing qualitative or quantitative measurements.

Ontology reference: Combination of Ontology Annotation and Ontology Source.



serialization

ISA-tab

Files are tab separated value (tsv) files, with specific labeled column structures specified underneath.

uses techniques of semantic web, e.g., for serialization

linkedisa isatordf

FDO Framework

A model to represent objects in a digital environment. (10)

FDO

(FAIR Digital Objects) are digital objects (DO) defined in the context of the FAIR principles: "represent data, software or other research resources" and "must be accompanied by persistent identifiers, metadata and contextual documentation to enable discovery, citation and reuse". (9)

uses techniques of semantic web, e.g., for serialization

OME Consortium

A consortium of universities, research labs, industry and developers producing open-source software and format standards for microscopy data. (16) OME = Open microscopy environment

Rating System Linked Open Data (11,12,13)

1. Put data openly on the web
2. Make it available in a structured format (machine-readable)
3. Publish data in open format which everyone can use:
   - use formats that are made by community and available free of charge
4. Use standards that are published (like RDF and SPARQL):
   - Instruction manual for how to publish data should be openly usable by anyone
   - no costs
5. Link to other people's data

1. Moore, Josh, et al.: *OME-NGFF: scalable format strategies for interoperable bioimaging data.*" bioRxiv, 13 Apr. 2021, p. 2021.03.31.437929, doi.org/10.1101/2021.03.31.437929.
2. https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs
3. https://www.w3.org/OWL/
4. https://www.w3.org/2001/sw/wiki/RDF
5. https://www.w3.org/standards/semanticweb/data
6. https://github.com/nfdi4plants/Swate
7. https://www.w3.org/TR/ld-glossary/
8. https://isa-tools.org/
9. European Commission, Directorate-General for Research and Innovation, *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office, 2018, https://data.europa.eu/doi/10.2777/1524
10. "FAIR Digital Object Framework Documentation." 3 Nov. 2021, fairdigitalobjectframework.org.
11. https://www.w3.org/DesignIssues/LinkedData.html
12. Fitzwilliam Museum, *Linked Open Data: how does it works*, https://youtu.be/0m79yDb4AzE
13. https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png
14. Goldberg, I.G., Allan, C., Burel, JM. et al. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biol 6, R47 (2005). https://doi.org/10.1186/gb-2005-6-5-r47
15. "TIFF: Summary from the Encyclopedia of Graphics File Formats." 17 June 2022, www.fileformat.info/format/tiff/egff.htm.
16. https://www.openmicroscopy.org/
17. OME-XML - Open Microscopy Environment XML | DCC. (2022, July 11). Retrieved from https://www.dcc.ac.uk/resources/metadata-standards/ome-xml-open-microscopy-environment-xml

Data Model: Abstract set of rules for representing information

Serialization Format: Way to encode information so that it can be transferred between machines (parsing)

Software/ Tools

Semantic Web Technology

Data Collection Concept

(funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 462231789)

The content of this poster with the exception of the cited logos is published with Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/).

Cite as: Susanne Kunis and Julia Dohle: "*Structuring of Data and Metadata in Bioimaging: Concepts and technical Solutions in the Context of Linked Data*", Information Infrastructure for BioImage Data (I3D:bio); doi: 10.5281/zenodo.6821815

I3D:bio
www.i3dbio.de