Metadata Matters 2024

a.k.a. "Next Generation Metadata (Framework)"

Josh Moore, Senior RDM Officer

German BioImaging, e.V. / Open Microscopy Environment

Metadata Matters 2010


JCB: Comment

Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[5] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]

Data sharing is important in the biological sciences to prevent duplication of effort, to promote scientific integrity, and to facilitate and disseminate scientific discovery. Sharing requires centralized repositories, and submission to and utility of these resources require common data formats. This is particularly challenging for multidimensional microscopy image data, which are acquired from a variety of platforms with a myriad of proprietary file formats (PFFs). In this paper, we describe an open standard format that we have developed for microscopy image data. We call on the community to use open image data standards and to insist that all imaging platforms support these file formats. This will build the foundation for an open image data repository.

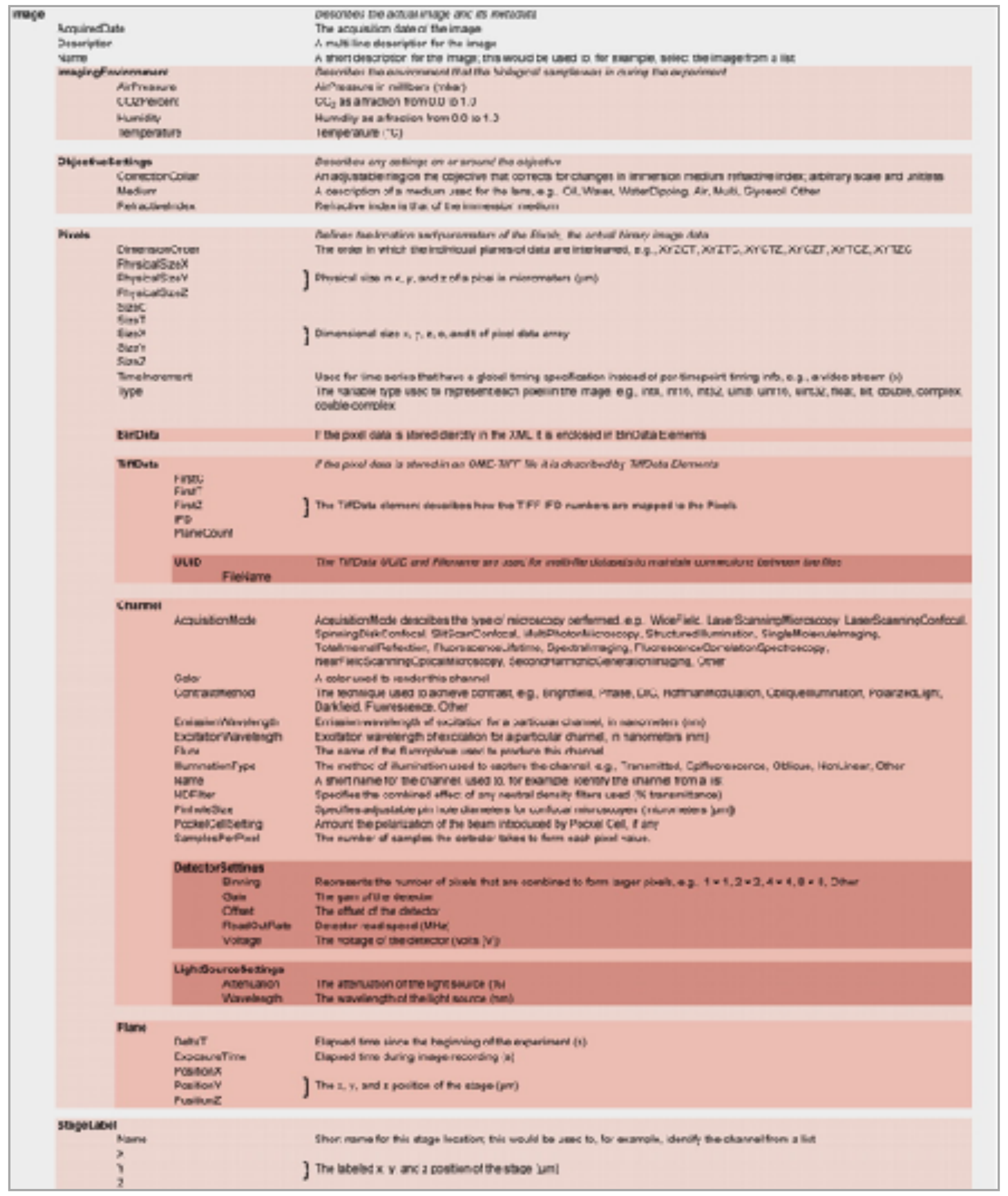
We believe the imaging community can achieve the same success for digital image data. In this paper, we review the current status of online biological image repositories and provide a set of recommendations to drive the use of open standardized data formats in biological microscopy as a prerequisite for creating a global image data repository.

Scientific image data repositories for the life sciences

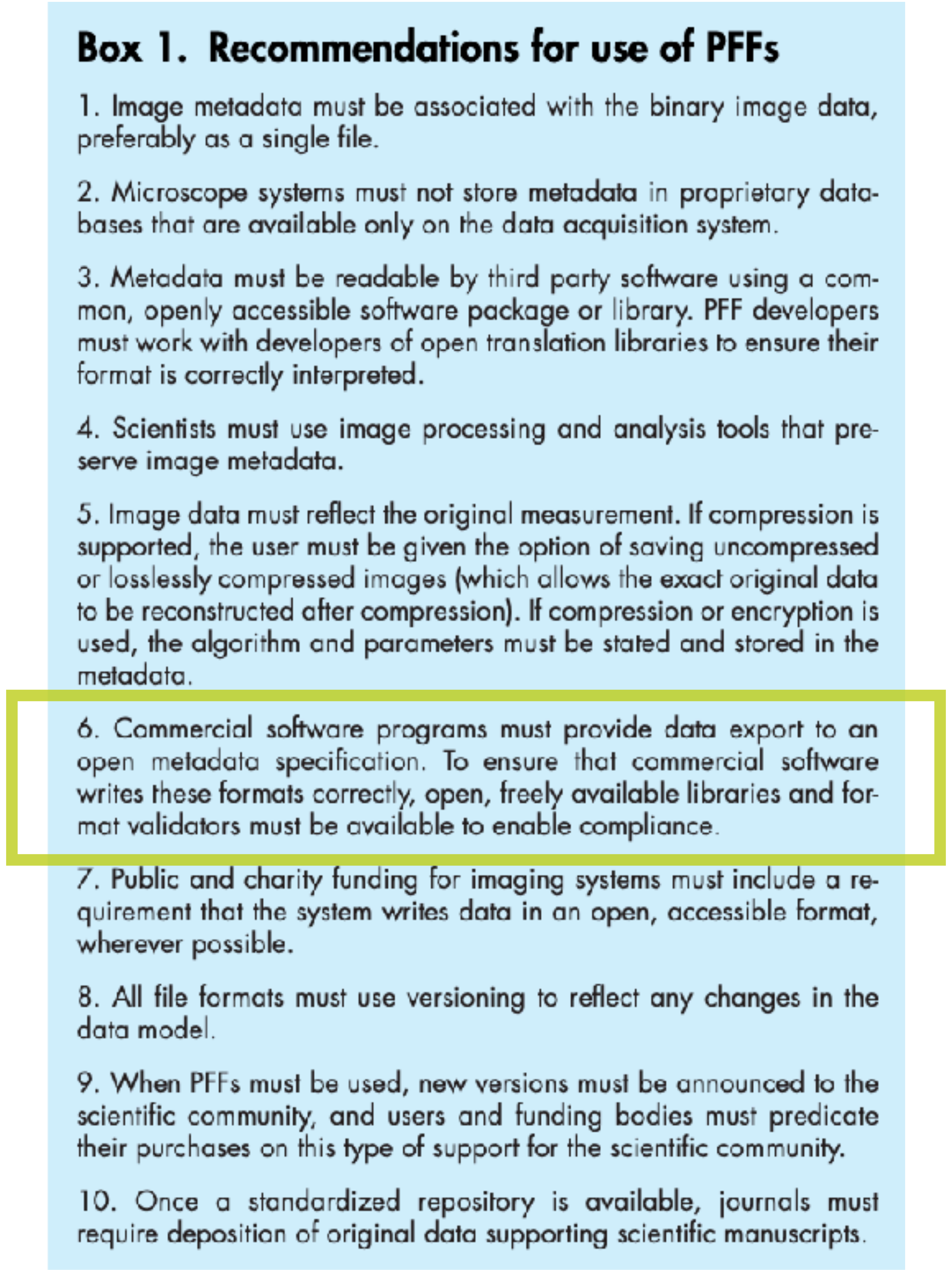
THE JOURNAL OF CELL BIOLOGY

https://doi.org/10.1083/jcb.201004104

(Most cited OME paper)



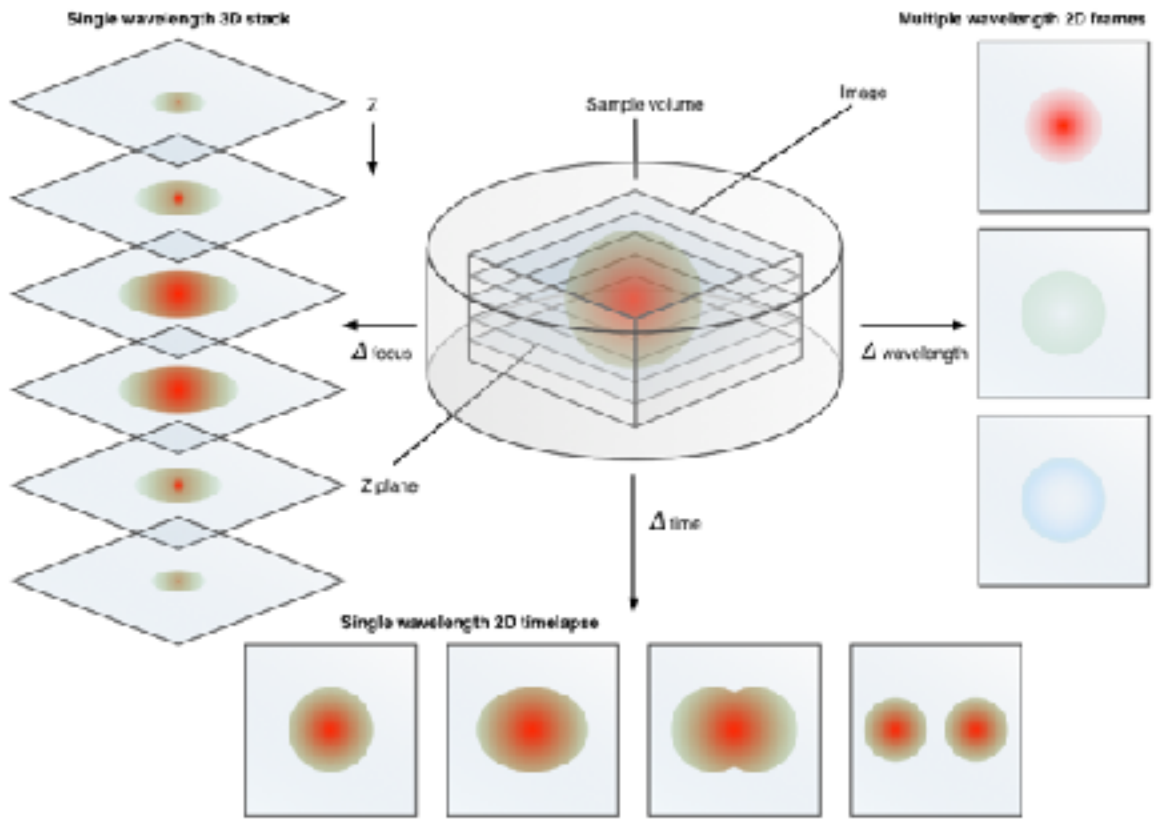
Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).



Box 1. Recommendations for use of PFFs

1. Image metadata must be associated with the binary image data, preferably as a single file.
2. Microscope systems must not store metadata in proprietary databases that are available only on the data acquisition system.
3. Metadata must be readable by third party software using a common, openly accessible software package or library. PFF developers must work with developers of open translation libraries to ensure their format is correctly interpreted.
4. Scientists must use image processing and analysis tools that preserve image metadata.
5. Image data must reflect the original measurement. If compression is supported, the user must be given the option of saving uncompressed or losslessly compressed images (which allows the exact original data to be reconstructed after compression). If compression or encryption is used, the algorithm and parameters must be stated and stored in the metadata.
6. Commercial software programs must provide data export to an open metadata specification. To ensure that commercial software writes these formats correctly, open, freely available libraries and format validators must be available to enable compliance.
7. Public and charity funding for imaging systems must include a requirement that the system writes data in an open, accessible format, wherever possible.
8. All file formats must use versioning to reflect any changes in the data model.
9. When PFFs must be used, new versions must be announced to the scientific community, and users and funding bodies must predicate their purchases on this type of support for the scientific community.
10. Once a standardized repository is available, journals must require deposition of original data supporting scientific manuscripts.

Maintaining a model



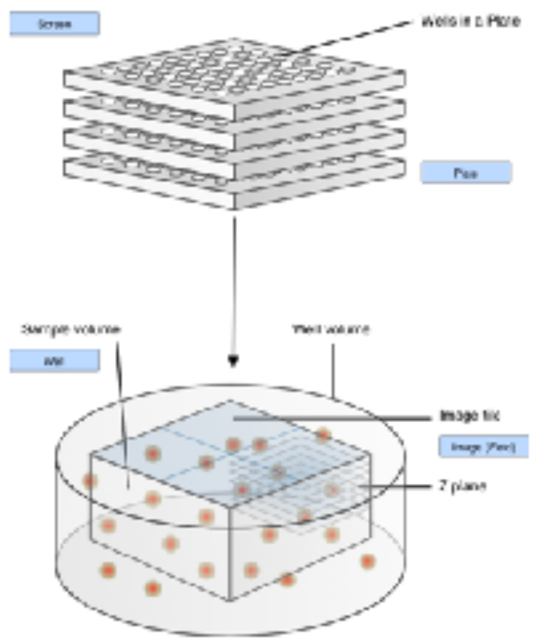
> *The OME Model is a specification for storing data on biological imaging. The model includes image parameters, such as XYZ dimensions and pixels type, as well as extensive metadata on, for example, image acquisition, annotation, and regions of interest (ROIs). This common specification is essential for the exchange of image data between different software packages.*

Swedlow et al. (2003) ***Informatics and quantitative analysis in biological imaging.*** *Science 300(5616), 100-2. Published 4 April 2003*

DOI: 10.1126/science.1082602

https://bit.ly/citing-ome

2008



Last major modeling effort for **high-content screening**

2010

```
▼<OME xmlns="http://www.openmicroscopy.org/Schemas/OME/2016-06"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.openmicroscopy.org/Schemas/OME/2016-
 06 http://www.openmicroscopy.org/Schemas/OME/2016-06/ome.xsd">
 ▼<Image ID="Image:0" Name="6x6x1x8-swatch.tif">
    <AcquisitionDate>2010-02-23T12:51:30</AcquisitionDate>
   ▶<Pixels DimensionOrder="XYCZT" ID="Pixels:0:0"
    PhysicalSizeX="10000.0" PhysicalSizeY="10000.0" Type="uint8"
    SizeC="1" SizeT="1" SizeX="6" SizeY="4" SizeZ="1">...</Pixels>
    <AnnotationRef ID="Annotation:1"/>
  </Image>
 ▼<StructuredAnnotations>
   ▼<CommentAnnotation ID="Annotation:1">
      <Value>Fred</Value>
    </CommentAnnotation>
  </StructuredAnnotations>
</OME>
```

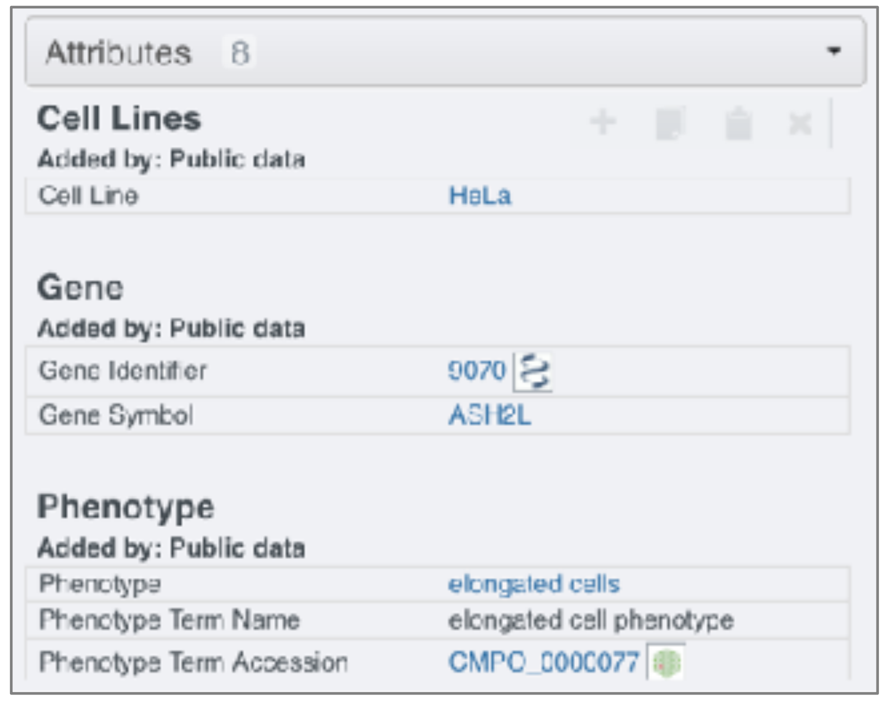
Added flexible **structured annotations** for user-extensions

2013

Study Type	Protein localisation using fluorescence microscopy
Organism	Homo sapiens
Experiment Type	Immunocytochemistry
Imaging Method	Deconvolution widefield fluorescence microscopy
Data Analysis	OMERO.mtools http://help.openmicroscopy.org/workflows-mtools.html
Publication Title	The Ndc80 complex targets Bod1 to human mitotic kinetochores
Publication Authors	Katharina Schleicher, Michael Porter, Sara ten Have, Ramasubramanian Sundaramoorthy, Iain M Porter, Jason R Swedlow
PubMed ID	29142109
PMC ID	tba
Publication DOI	10.1098/rsob.170099
License	Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0/
Data Publisher	University of Dundee
Data DOI	http://dx.doi.org/10.17867/10000109

Add **key-value** (K/V) pair annotation

2016



Cell Lines

Added by: Public data

Cell Line	HeLa

Gene

Added by: Public data

Gene Identifier	9070
Gene Symbol	ASH2L

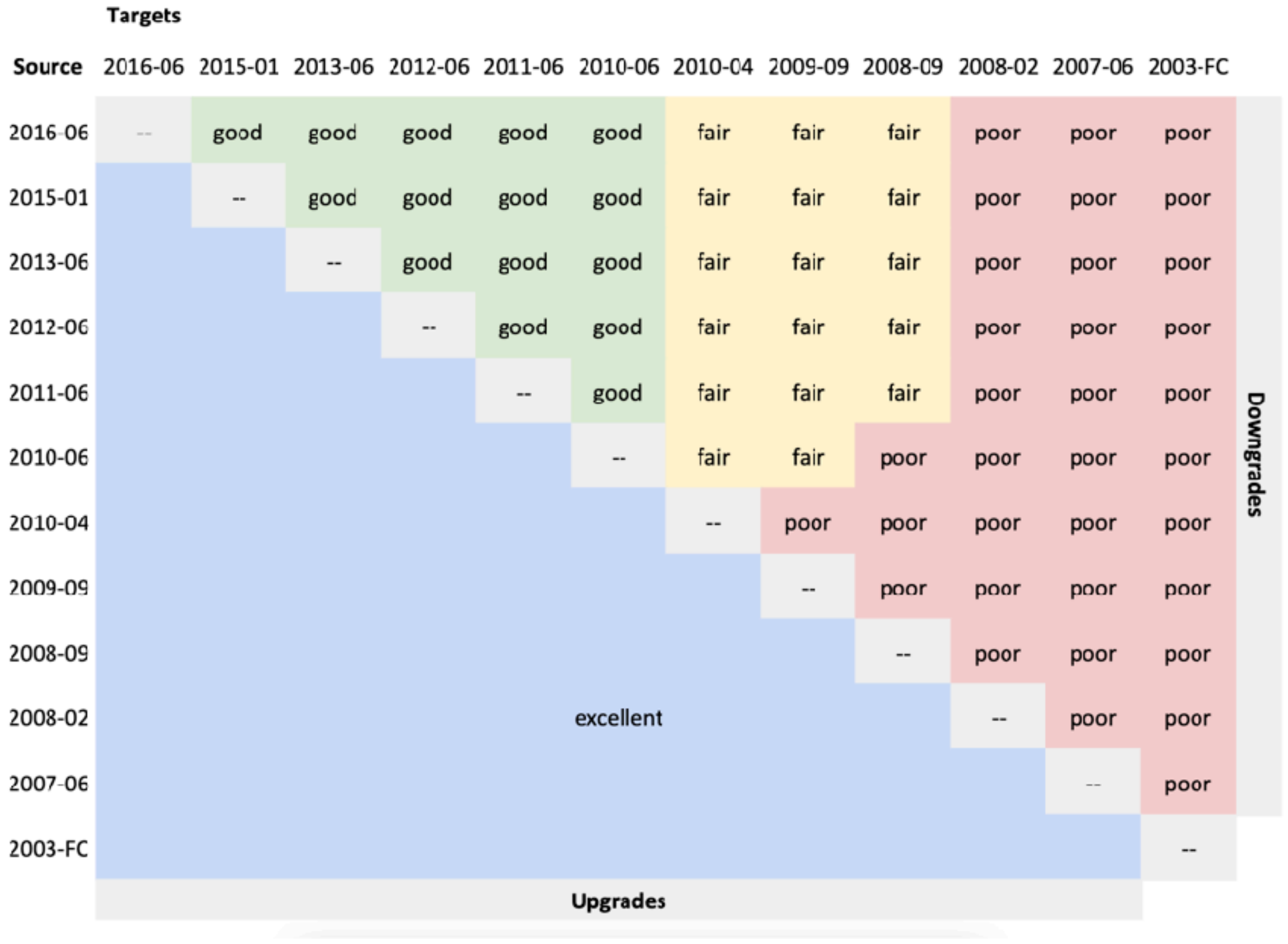
Phenotype

Added by: Public data

Phenotype	elongated cells
Phenotype Term Name	elongated cell phenotype
Phenotype Term Accession	CMPO_0000077

Use k/v for **ontology links**

Maintaining a model

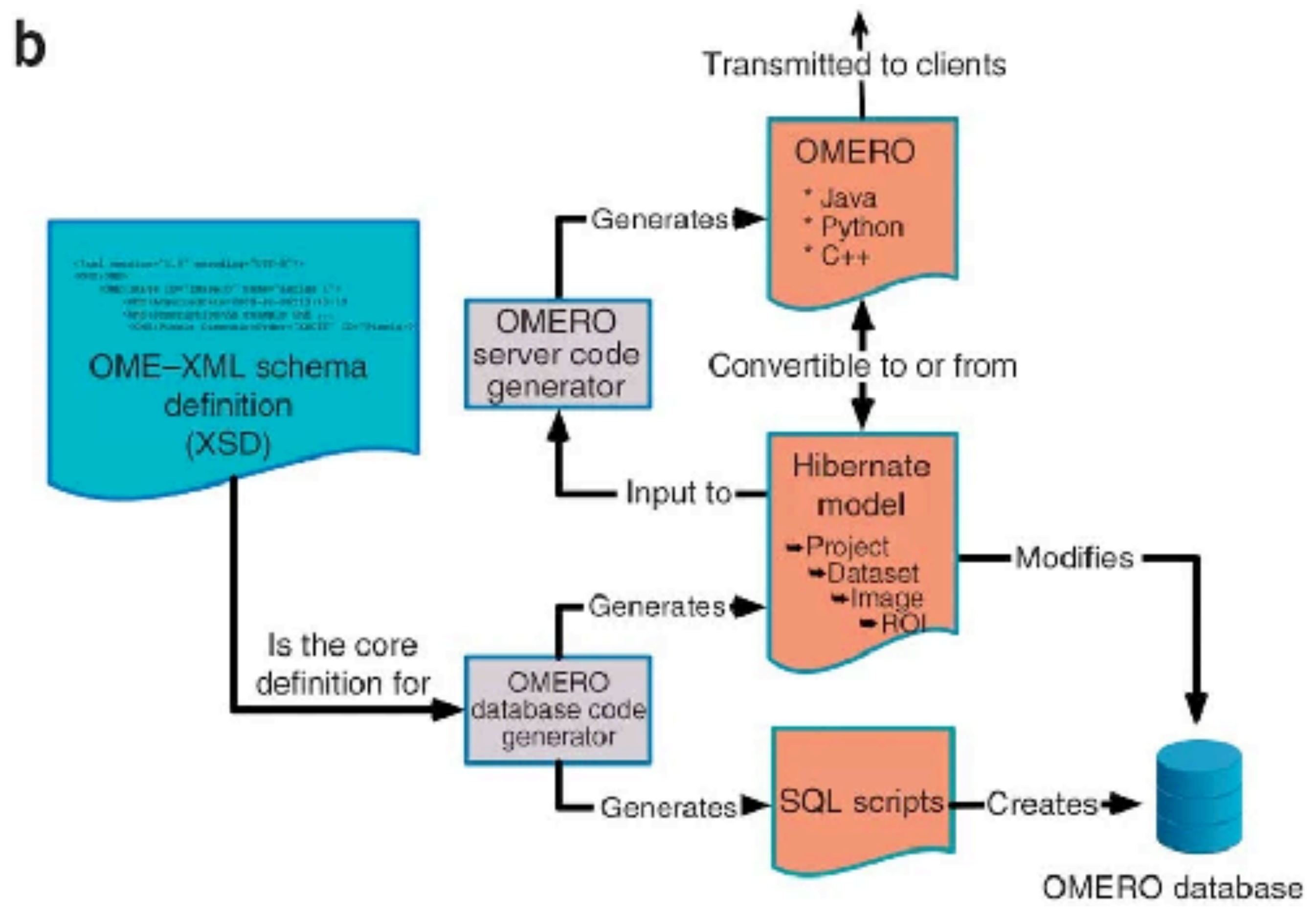


Targets

Source	2016-06	2015-01	2013-06	2012-06	2011-06	2010-06	2010-04	2009-09	2008-09	2008-02	2007-06	2003-FC
2016-06	--	good	good	good	good	good	fair	fair	fair	poor	poor	poor
2015-01		--	good	good	good	good	fair	fair	fair	poor	poor	poor
2013-06			--	good	good	good	fair	fair	fair	poor	poor	poor
2012-06				--	good	good	fair	fair	fair	poor	poor	poor
2011-06					--	good	fair	fair	fair	poor	poor	poor
2010-06						--	fair	fair	poor	poor	poor	poor
2010-04							--	poor	poor	poor	poor	poor
2009-09								--	poor	poor	poor	poor
2008-09									--	poor	poor	poor
2008-02						excellent				--	poor	poor
2007-06											--	poor
2003-FC												--

Downgrades

Upgrades

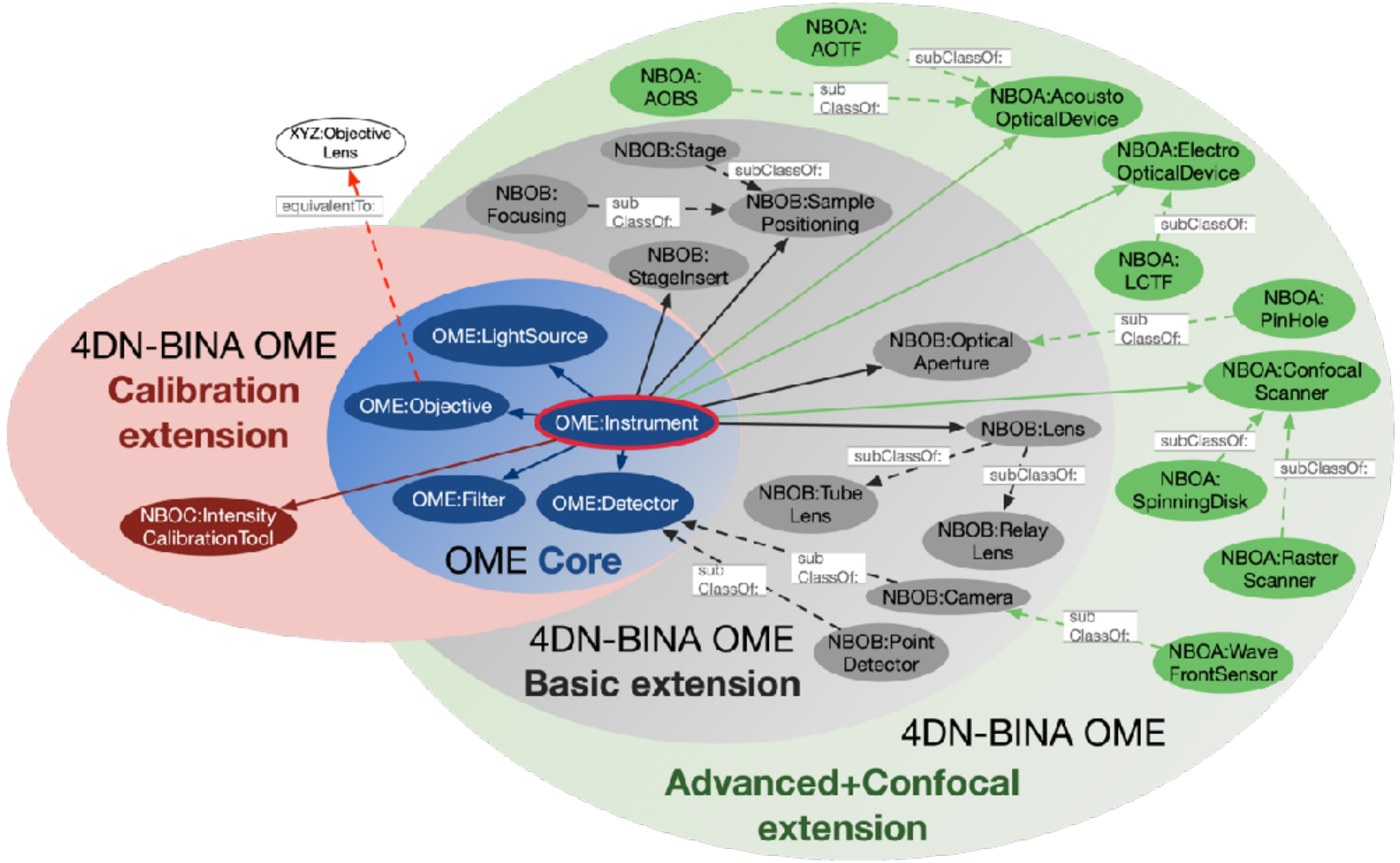


Allan et al. (2012) ***OMERO: flexible, model-driven data management for experimental biology.*** *Nature Methods 9, 245–253. Published 28 February 2012*

DOI: [10.1038/nmeth.1896](https://doi.org/10.1038/nmeth.1896)

https://bit.ly/citing-ome

Extending a model??



4D-Nucleome / BINA / OME / QUAREP-LiMi (2018-2019)

https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs


Caterina
Strambio De Castillia


Damir Sudar

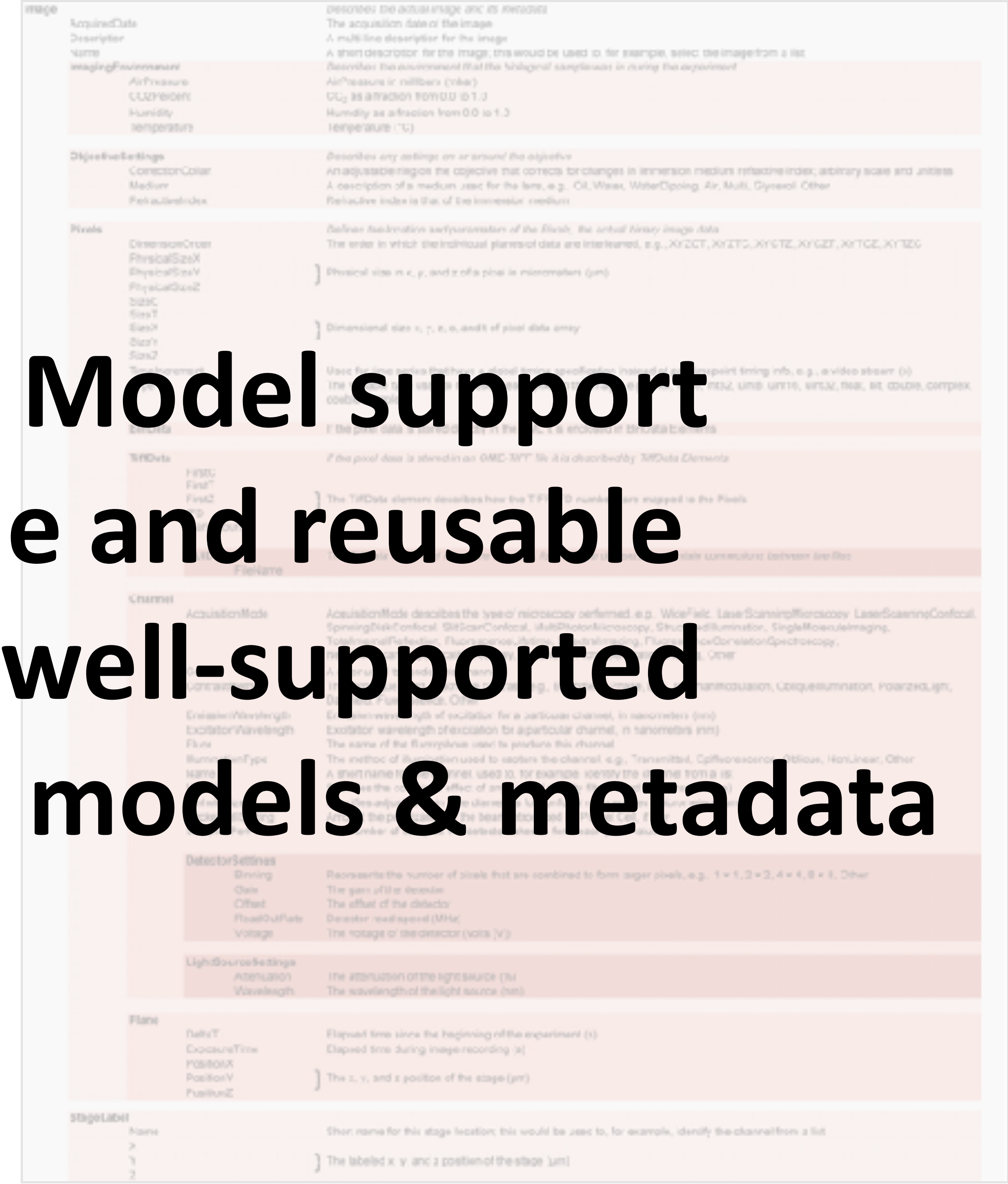
Metadata Matters 2024

- Superset of OME Model support
- Models extensible and reusable
- Base models are well-supported
- Vendors produce models & metadata

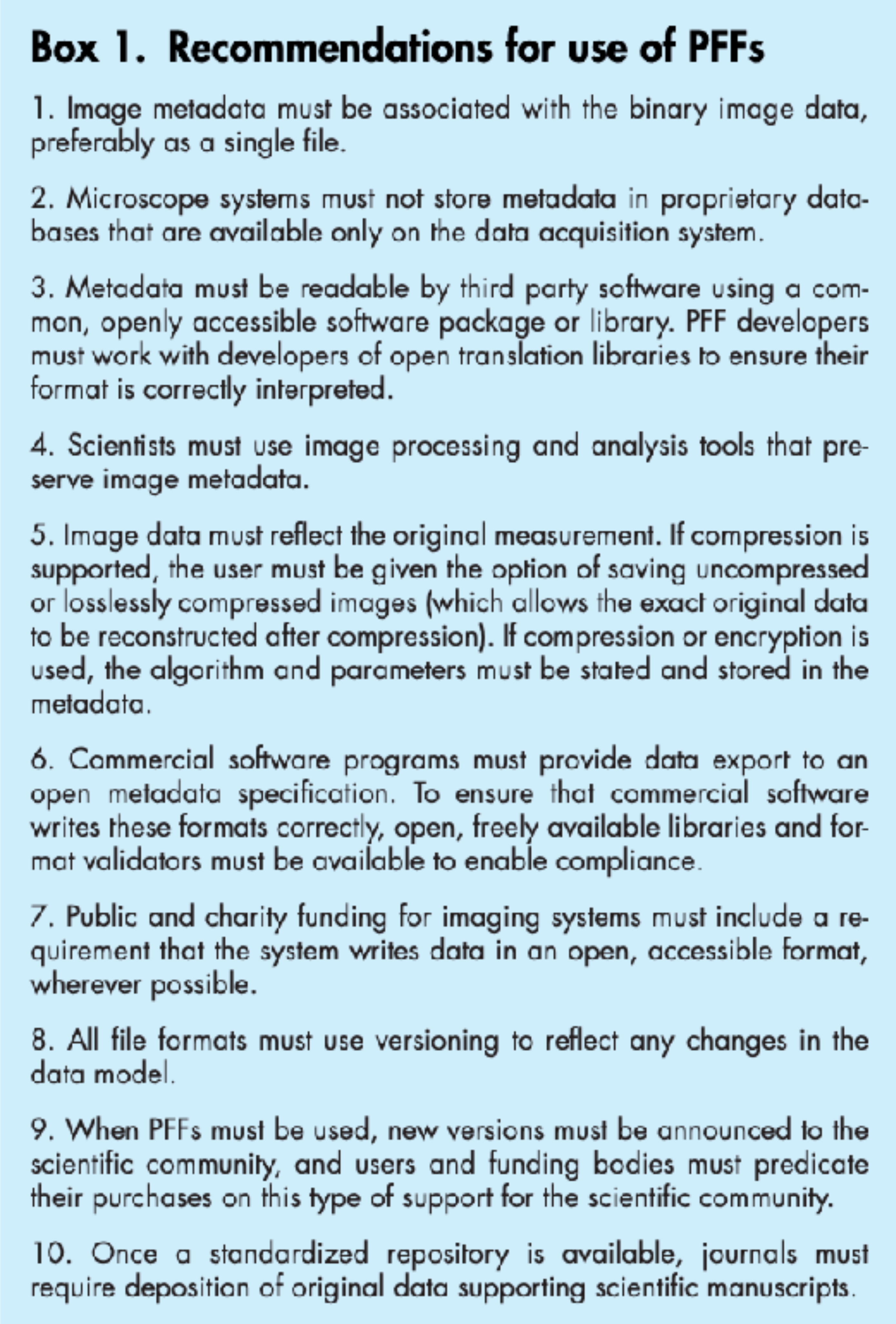
JCB: Comment

Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[5] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]



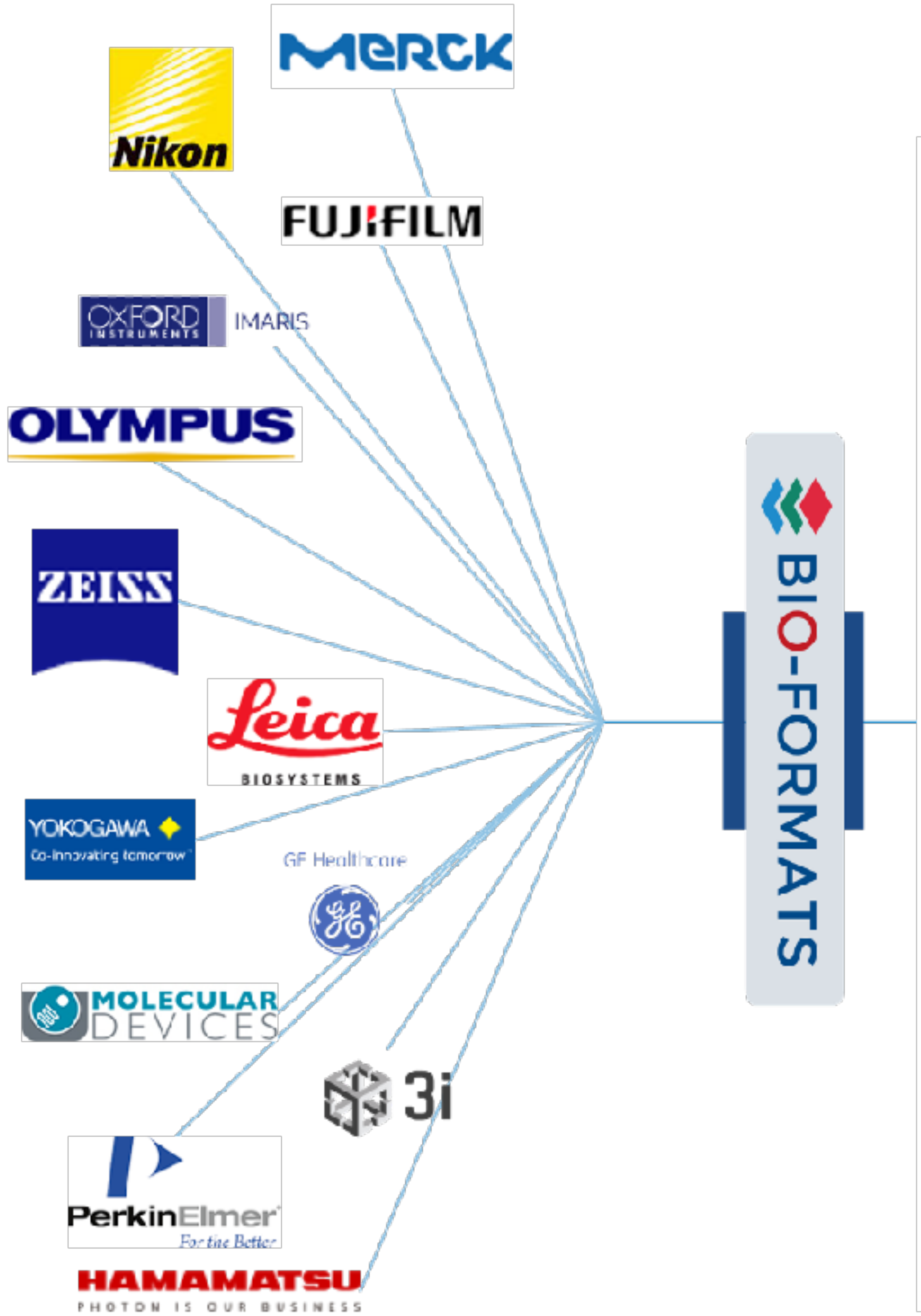
Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).



Box 1. Recommendations for use of PFFs

1. Image metadata must be associated with the binary image data, preferably as a single file.
2. Microscope systems must not store metadata in proprietary databases that are available only on the data acquisition system.
3. Metadata must be readable by third party software using a common, openly accessible software package or library. PFF developers must work with developers of open translation libraries to ensure their format is correctly interpreted.
4. Scientists must use image processing and analysis tools that preserve image metadata.
5. Image data must reflect the original measurement. If compression is supported, the user must be given the option of saving uncompressed or losslessly compressed images (which allows the exact original data to be reconstructed after compression). If compression or encryption is used, the algorithm and parameters must be stated and stored in the metadata.
6. Commercial software programs must provide data export to an open metadata specification. To ensure that commercial software writes these formats correctly, open, freely available libraries and format validators must be available to enable compliance.
7. Public and charity funding for imaging systems must include a requirement that the system writes data in an open, accessible format, wherever possible.
8. All file formats must use versioning to reflect any changes in the data model.
9. When PFFs must be used, new versions must be announced to the scientific community, and users and funding bodies must predicate their purchases on this type of support for the scientific community.
10. Once a standardized repository is available, journals must require deposition of original data supporting scientific manuscripts.

Goal: **Common Framework**



```
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameHeight #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameWidth #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtLinePeriod #1=3.0000000000000001e-005
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetX #1=0
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetY #1=0
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionHeight #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionWidth #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtSuperSampling #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtZoom #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingX #1=4.1513291872589499e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingY #1=4.1513291872589499e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingZ #1=9.9999999999999995e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|SimRotations #1=3
Experiment|AcquisitionBlock|AcquisitionModeSetup|TimeSeries #1=false
Experiment|AcquisitionBlock|AcquisitionModeSetup|TrackMultiplexType #1=ZStack
Experiment|AcquisitionBlock|AcquisitionModeSetup|UseRois #1=false
Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomX #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomY #1=1
Experiment|AcquisitionBlock|Laser|LaserName #1=HeNe633
Experiment|AcquisitionBlock|Laser|LaserName #2=Diode 405-30
Experiment|AcquisitionBlock|Laser|LaserName #3=Argon
Experiment|AcquisitionBlock|Laser|LaserName #4=HeNe594
Experiment|AcquisitionBlock|Laser|LaserPower #1=0.0050000000000000001
Experiment|AcquisitionBlock|Laser|LaserPower #2=0.029999999999999999
Experiment|AcquisitionBlock|Laser|LaserPower #3=0.025000000000000001
Experiment|AcquisitionBlock|Laser|LaserPower #4=0.002

...
```

Goal: **Common Framework**

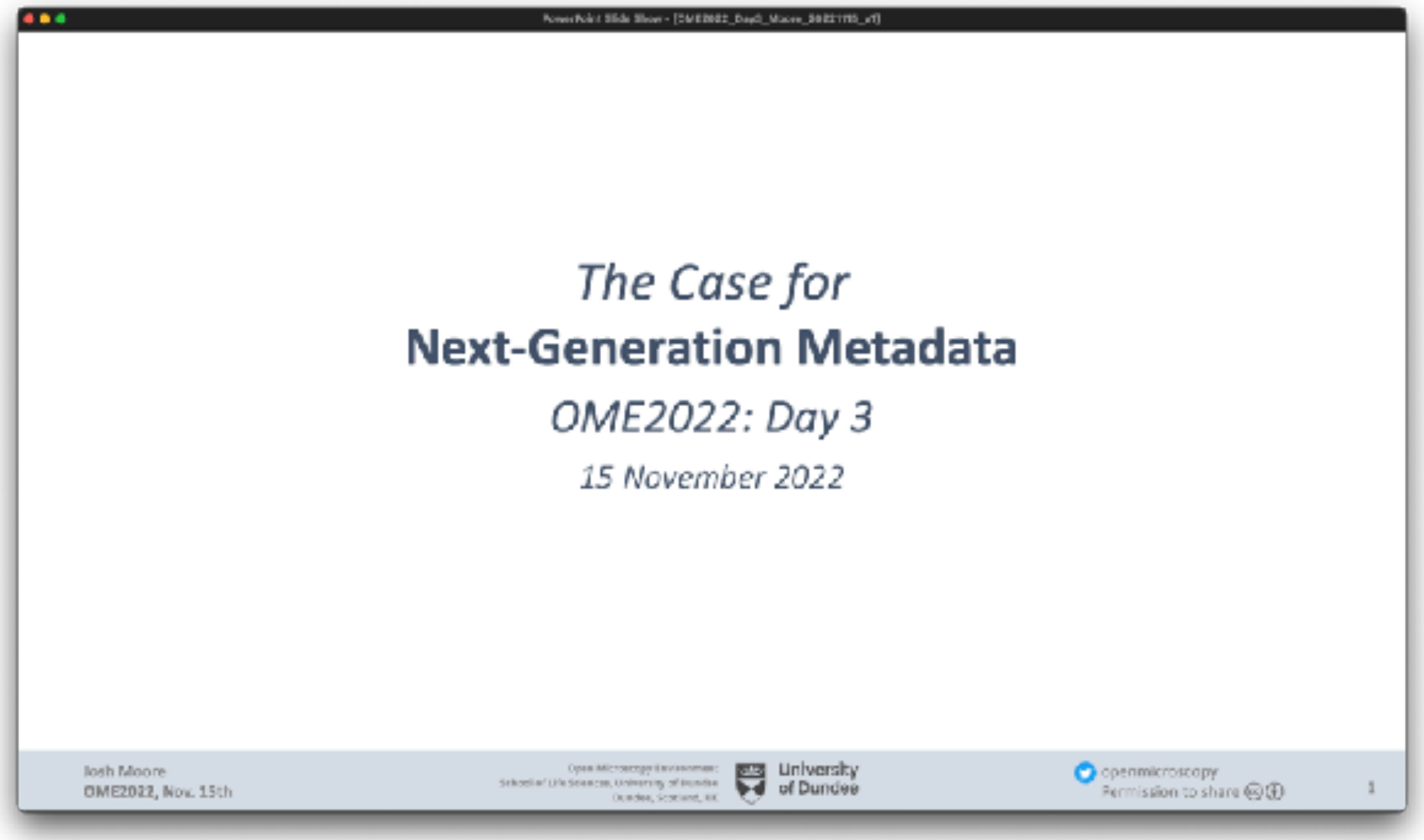
```
{
  @context: {
    “ome”: “https://openmicroscopy.org/namespace”,
    “vendor”: “https://vendor.com/namespace”
  },
  “ome:Image”: …
  “vendor:acquisitionModeSetup”: [
    {
      “RtFrameHeight”: 512
      …
    }, …
  ]
}
```

JSON-LD (RDF)

(make it simple)

https://linkml.io/

OME**2022**



The Case for

Next-Generation Metadata

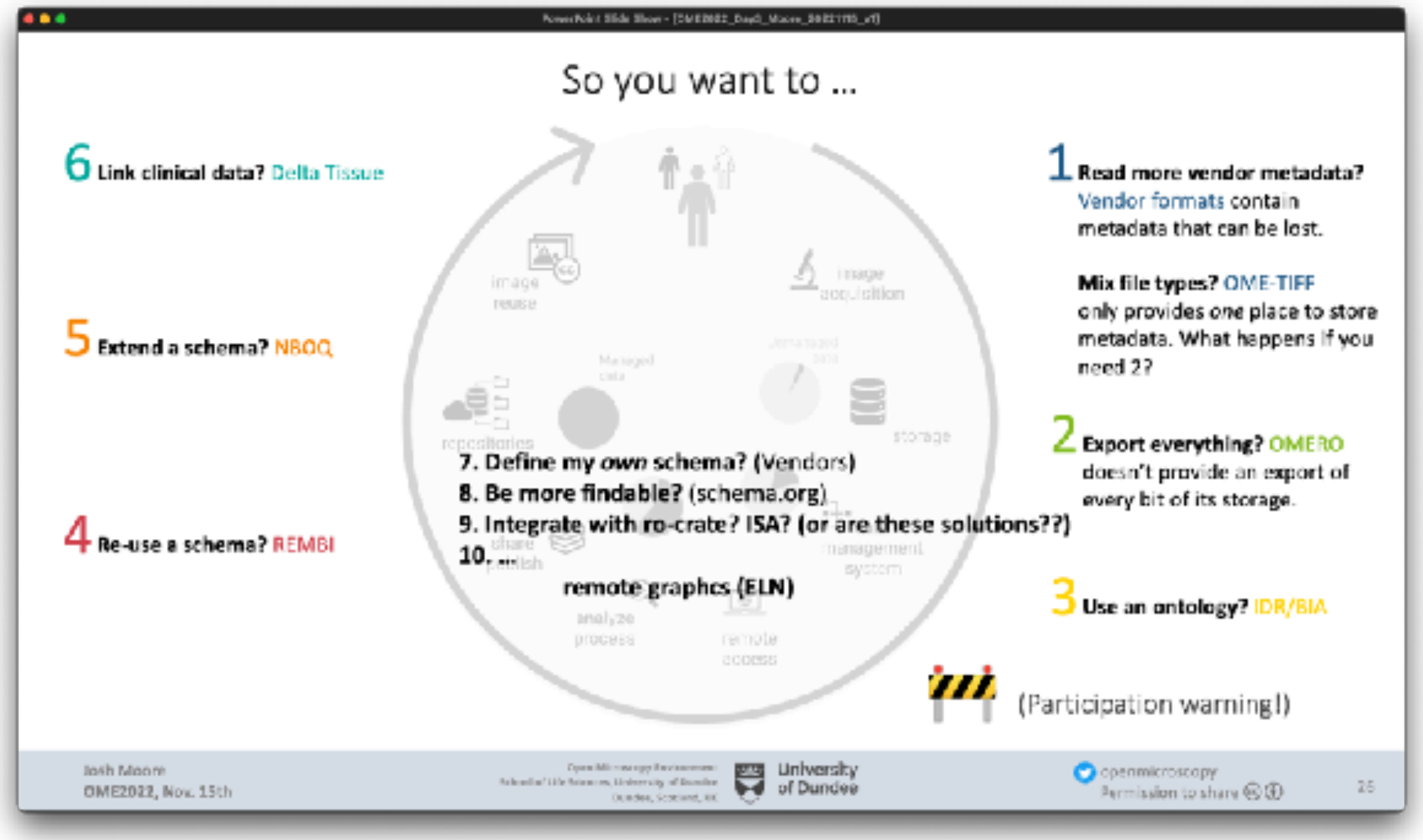
OME2022: Day 3

15 November 2022

Josh Moore
OME2022, Nov. 15th

University of Dundee

openmicroscopy
Permission to share



So you want to …

6 **Link clinical data?** Delta Tissue

5 **Extend a schema?** NBOQ

4 **Re-use a schema?** REMBI

1 **Read more vendor metadata?**
Vendor formats contain metadata that can be lost.

Mix file types? OME-TIFF only provides one place to store metadata. What happens if you need 2?

2 **Export everything?** OMERO doesn't provide an export of every bit of its storage.

3 **Use an ontology?** IDR/BIA

7. Define my *own* schema? (Vendors)
8. Be more findable? (schema.org)
9. Integrate with ro-crate? ISA? (or are these solutions??)
10. …

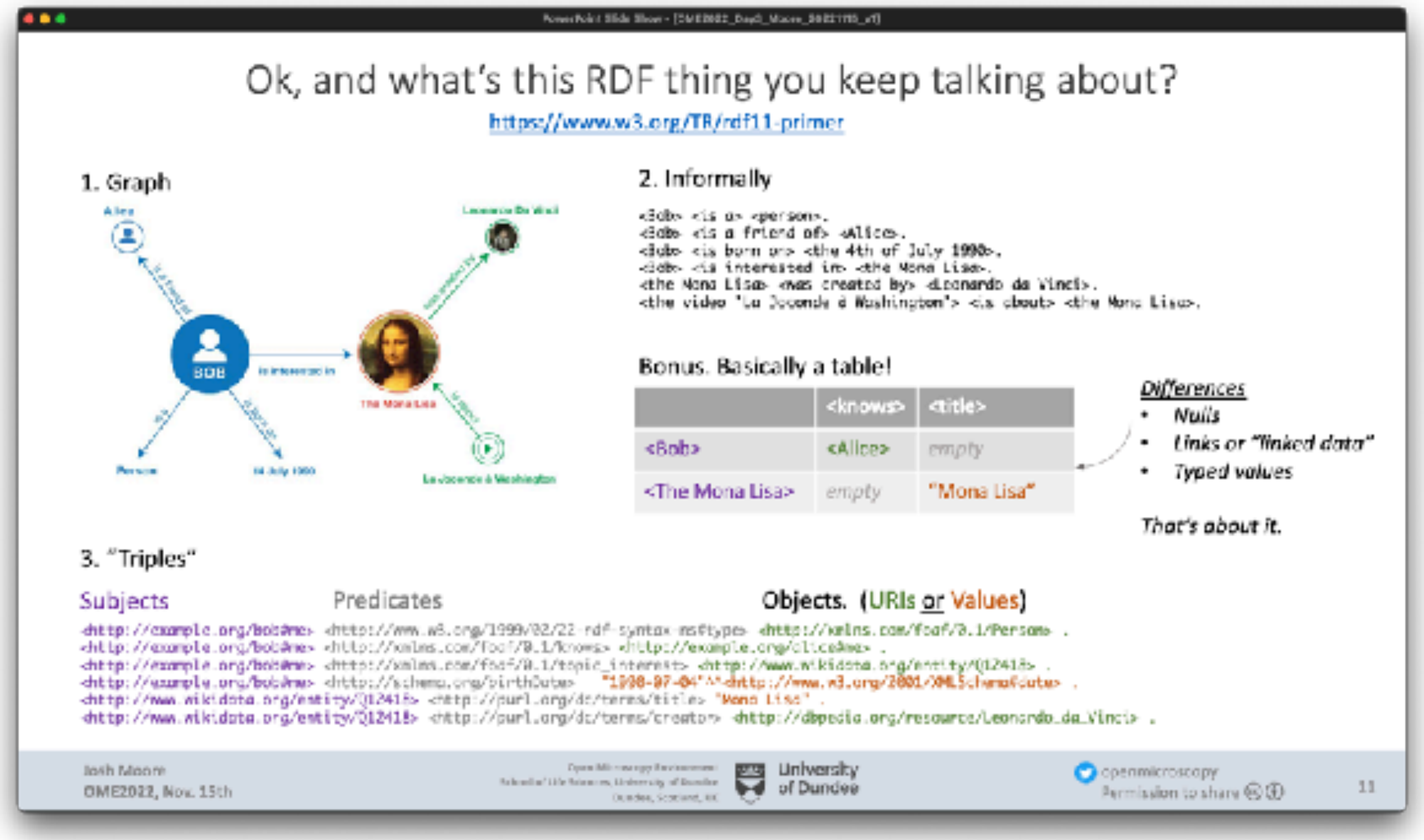
remote graphs (ELN)

(Participation warning!)

Josh Moore
OME2022, Nov. 15th

University of Dundee

openmicroscopy
Permission to share



Ok, and what's this RDF thing you keep talking about?

https://www.w3.org/TR/rdf11-primer

1. Graph

2. Informally

```
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>.
```

Bonus. Basically a table!

	<knows>	<title>
<Bob>	<Alice>	*empty*
<The Mona Lisa>	*empty*	"Mona Lisa"

Differences
- *Nulls*
- *Links or "linked data"*
- *Typed values*

That's about it.

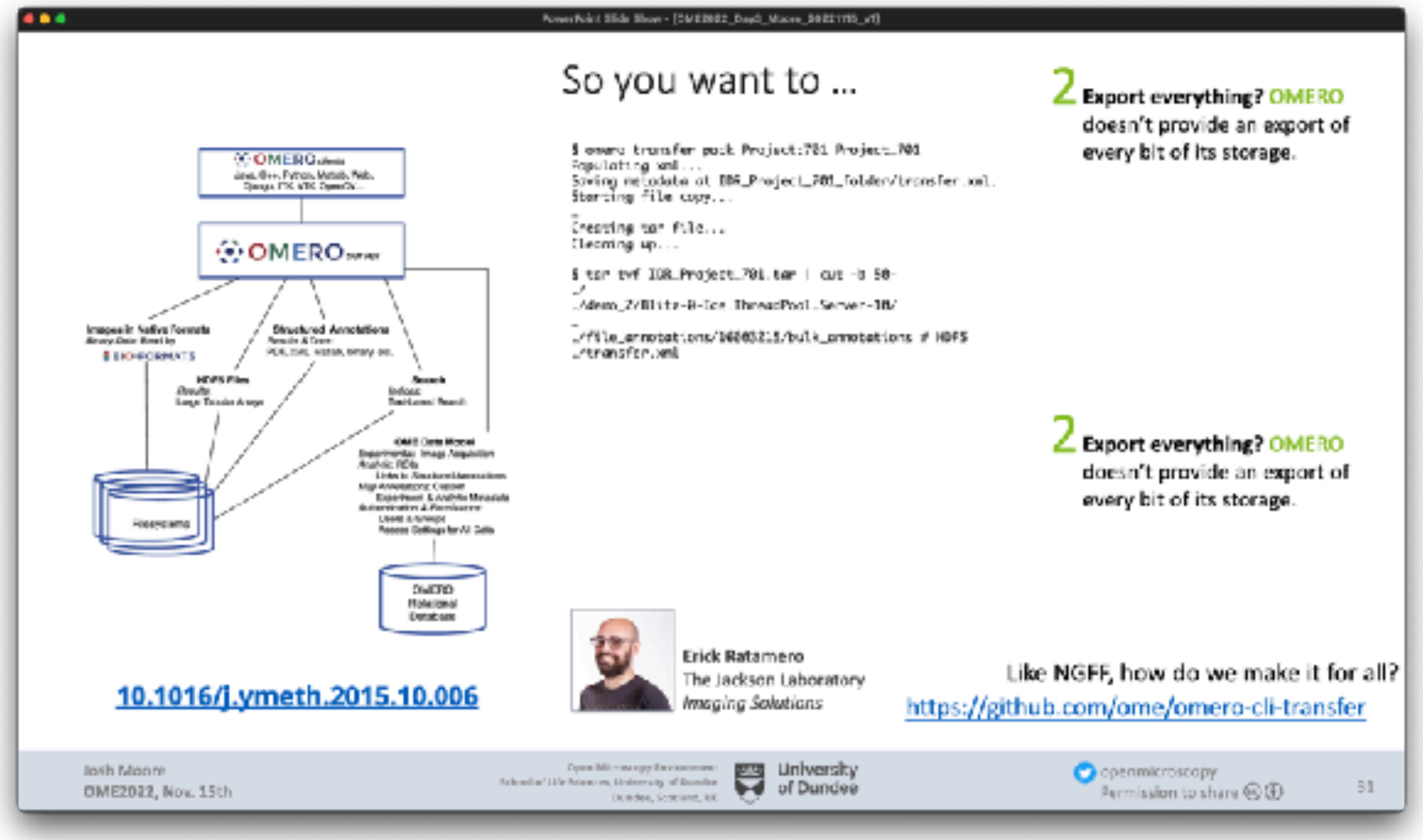
3. "Triples"

Subjects Predicates **Objects. (URIs or Values)**

Josh Moore
OME2022, Nov. 15th

University of Dundee

openmicroscopy
Permission to share



So you want to …

2 **Export everything?** OMERO doesn't provide an export of every bit of its storage.

2 **Export everything?** OMERO doesn't provide an export of every bit of its storage.

10.1016/j.ymeth.2015.10.006

Erick Ratamero
The Jackson Laboratory
Imaging Solutions

Like NGFF, how do we make it for all?
https://github.com/ome/omero-cli-transfer

Josh Moore
OME2022, Nov. 15th

University of Dundee

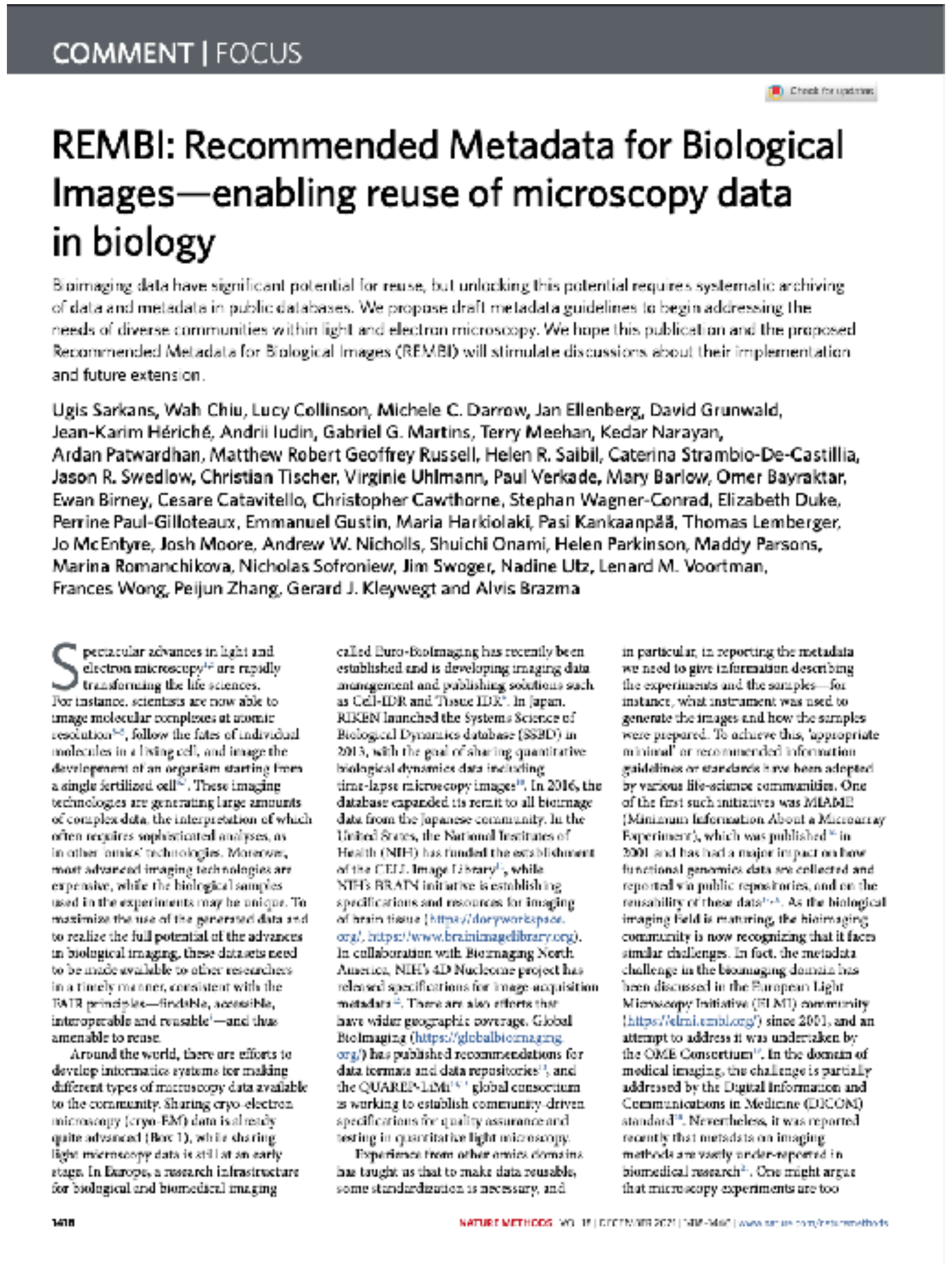
openmicroscopy
Permission to share

https://www.youtube.com/playlist?list=PL-oOCWFUMH5033INnAE-P4HBm2e6Mn6FA

What’s been going on?

(tl;dr)

REMBI (2019)



COMMENT | FOCUS

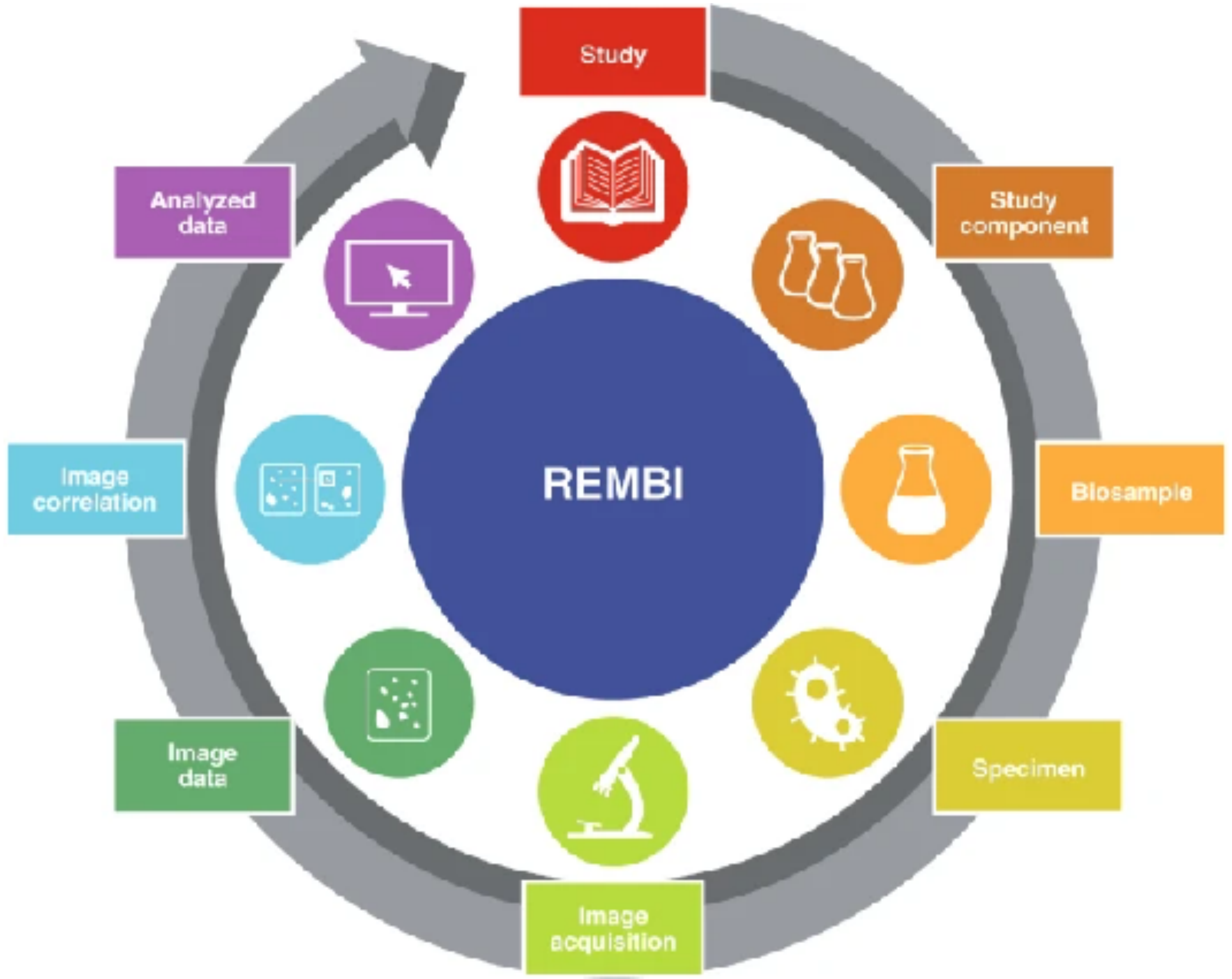
Check for updates

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

Bioimaging data have significant potential for reuse, but unlocking this potential requires systematic archiving of data and metadata in public databases. We propose draft metadata guidelines to begin addressing the needs of diverse communities within light and electron microscopy. We hope this publication and the proposed Recommended Metadata for Biological Images (REMBI) will stimulate discussions about their implementation and future extension.

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-Castillia, Jason R. Swedlow, Christian Tischer, Virginie Uhlmann, Paul Verkade, Mary Barlow, Omer Bayraktar, Ewan Birney, Cesare Catavitello, 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 Ultz, Lenard M. Voortman, Frances Wong, Peijun Zhang, Gerard J. Kleywegt and Alvis Brazma

https://doi.org/10.1038/s41592-021-01166-8



BioImage Archive

Study Type	Protein localisation using fluorescence microscopy
Organism	Homo sapiens
Experiment Type	Immunocytochemistry
Imaging Method	Deconvolution widefield fluorescence microscopy
Data Analysis	OMERO.mtools http://help.openmicroscopy.org/workflows-mtools.html
Publication Title	The Ndc80 complex targets Bod1 to human mitotic kinetochores
Publication Authors	Katharina Schleicher, Michael Porter, Sara ten Have, Ramasubramanian Sundaramoorthy, Iain M Porter, Jason R Swedlow
PubMed ID	29142109
PMC ID	tba
Publication DOI	10.1098/rsob.170099
License	Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0/
Data Publisher	University of Dundee
Data DOI	http://dx.doi.org/10.17867/10000109

Matthew Hartley

QUAREP-LiMi (2020)

Quality Assessment and Reproducibility for Instruments & Images in Light Microscopy

QUAREP-LiMi – who are we?

WG 7 – Metadata

Home / Working Groups / WG 7 – Metadata

WG 7 - Objectives

The Work Group 7 focuses on various aspects of Image Metadata (initially primarily Microscopy Metadata but also Sample and Analysis Metadata). These span from defining and continually revising the 4DN-BINA-OME (NBO) Microscopy Metadata model, to facilitating metadata capture through user-friendly tools, to integrating the captured metadata into cloud-ready storage containers such as the newly developed Open Microscopy Environment (OME) Next-Generation File Format (NGFF). To promote the uptake of metadata tools and models, educational material and outreach strategies are also developed.

To carry out these tasks, WG7 is organized into four Subgroups (see "Subgroup" tab below), each of which carries out specific objectives. These subgroups are created, paused, or dissolved on a needs basis as agreed by the WG7 as a whole.

The general meeting serves as a space for reporting back on the progress that has been achieved within the Subgroups, connecting the dots, strategic planning, and general discussions.

The Subgroups (see "Subgroup" tab below) identify specific tasks to be carried out to move our work forward. To participate, please first fill in the general QUAREP-LiMi membership form and then enter your name here. Still confused? No worries, we have a specific list of tasks that you can check out to know exactly how you can lend a hand.

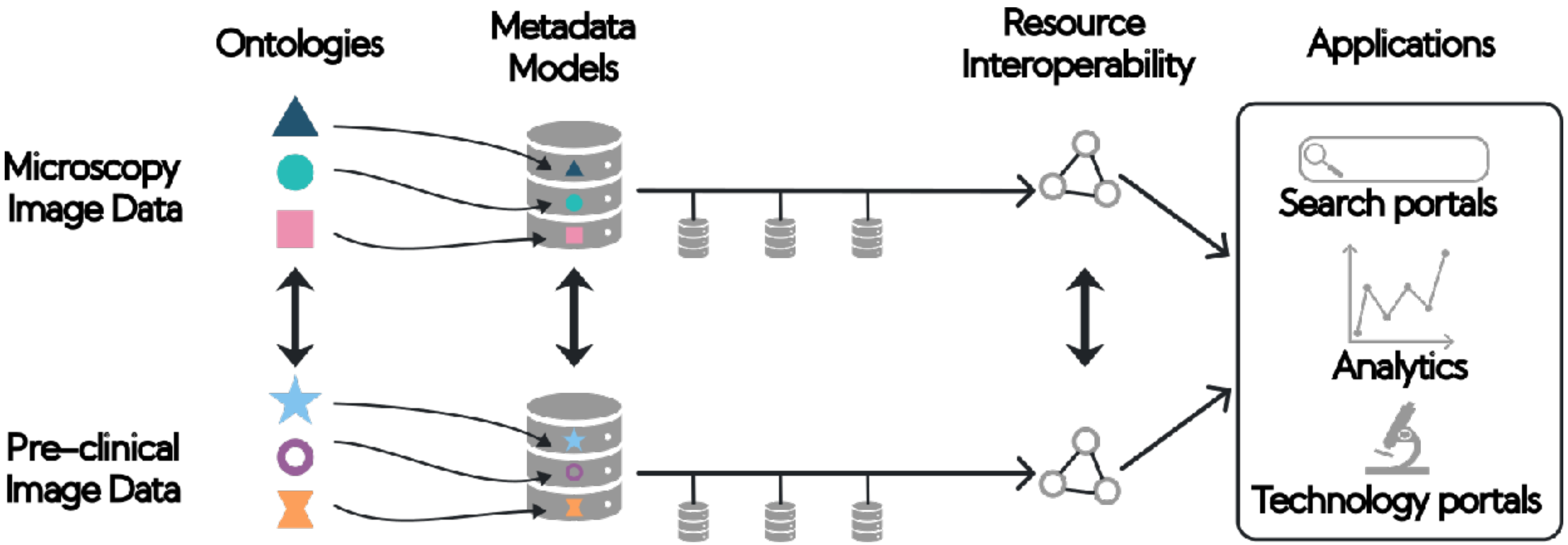
WG 7 - Co-Chairs

Kenneth Ho
Big Image Data Systems Engineer at the Francis Crick Institute
Co-Chair of WG 7
Call:
Email: kenneth.ho@crick.ac.uk

Aastha Mathur
Coordinator Image Data Services at EuroBioImaging
Co-Chair & Web Content Manager of WG 7
Call:
Email:
aastha.mathur@eurobioimaging.eu

Caterina Strambio De Castillia
Ass. Professor in Mol. Medicine
Chan Zuckerberg Imaging Scientist
Co-Chair of WG 7
Call: +1 (508) 856-2383
Email:
caterina.strambio@umassmed.edu

FoundingGIDE (2024-2026)



foundingGIDE (**G**lobal **I**mage **D**ata **E**cosystem)

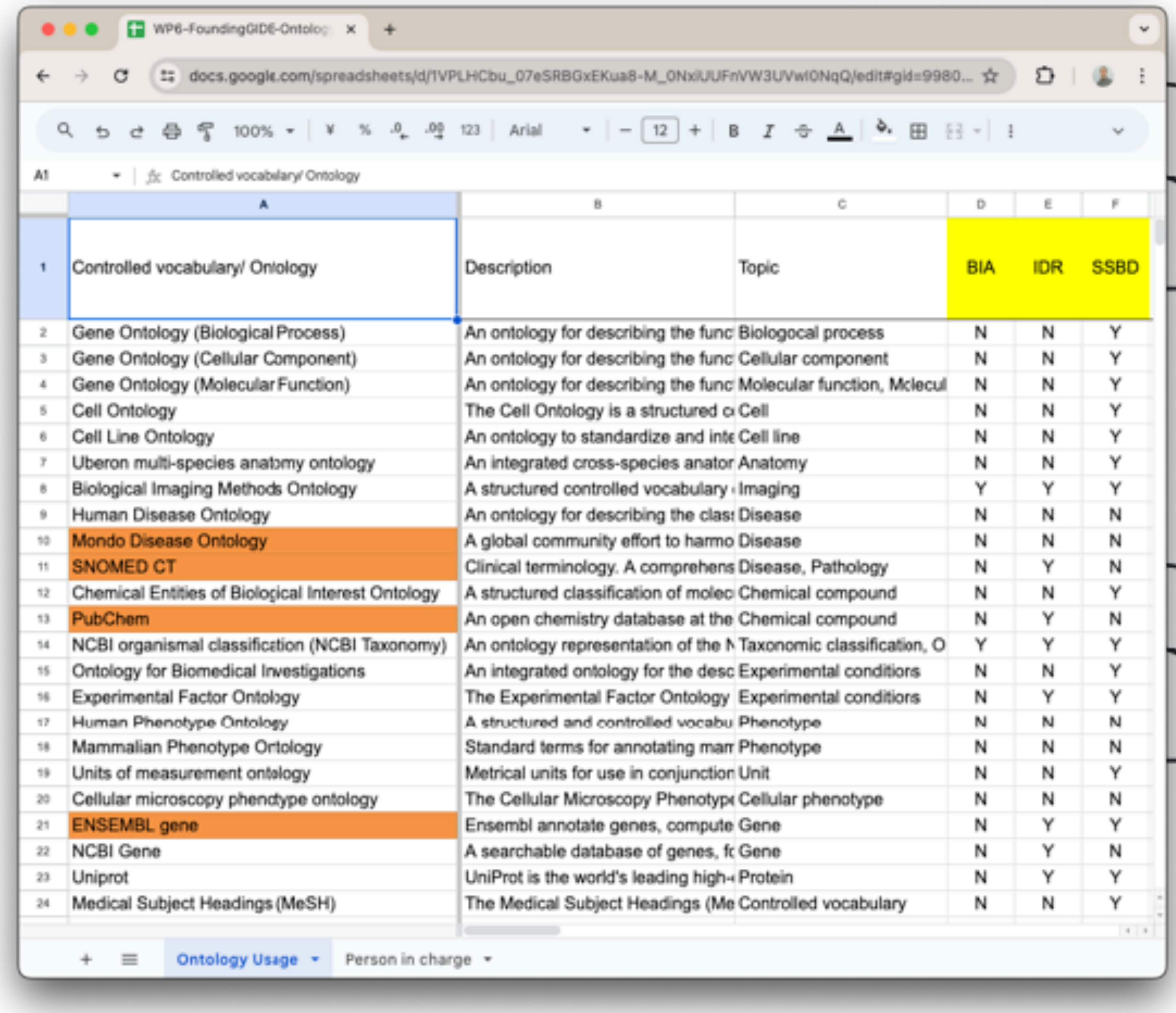
Funded by the European Union


Matthew Hartley

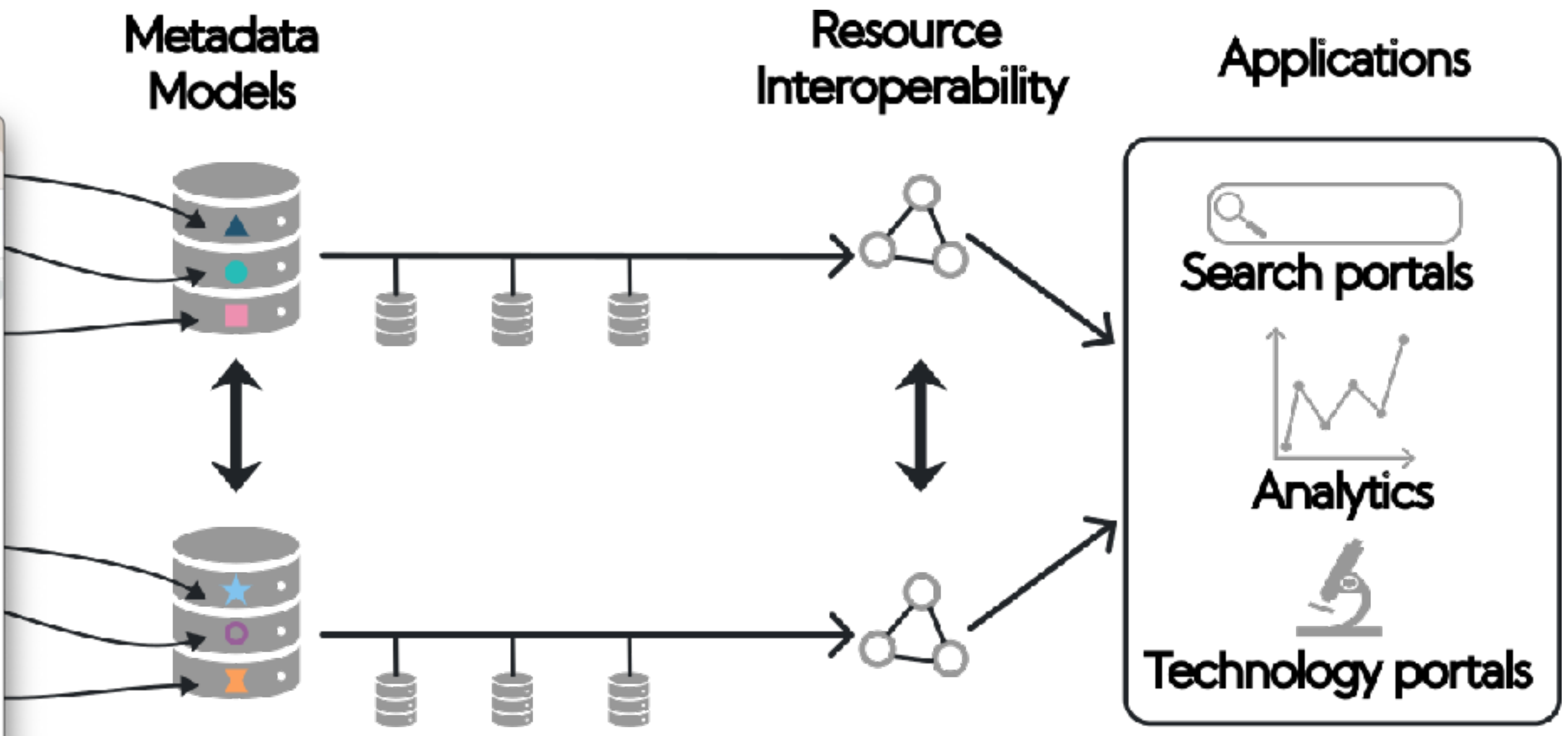

Shuichi Onami

FoundingGIDE (2024-2026)

NFDI4BIOIMAGE



	A	B	C	D	E	F
1	Controlled vocabulary/ Ontology	Description	Topic	BIA	IDR	SSBD
2	Gene Ontology (Biological Process)	An ontology for describing the func	Biologocal process	N	N	Y
3	Gene Ontology (Cellular Component)	An ontology for describing the func	Cellular component	N	N	Y
4	Gene Ontology (Molecular Function)	An ontology for describing the func	Molecular function, Mclecul	N	N	Y
5	Cell Ontology	The Cell Ontology is a structured c	Cell	N	N	Y
6	Cell Line Ontology	An ontology to standardize and inte	Cell line	N	N	Y
7	Uberon multi-species anatomy ontology	An integrated cross-species anator	Anatomy	N	N	Y
8	Biological Imaging Methods Ontology	A structured controlled vocabulary	Imaging	Y	Y	Y
9	Human Disease Ontology	An ontology for describing the clas	Disease	N	N	N
10	Mondo Disease Ontology	A global community effort to harmo	Disease	N	N	N
11	SNOMED CT	Clinical terminology. A comprehens	Disease, Pathology	N	Y	N
12	Chemical Entities of Biological Interest Ontology	A structured classification of molec	Chemical compound	N	N	Y
13	PubChem	An open chemistry database at the	Chemical compound	N	Y	N
14	NCBI organismal classification (NCBI Taxonomy)	An ontology representation of the N	Taxonomic classification, O	Y	Y	Y
15	Ontology for Biomedical Investigations	An integrated ontology for the desc	Experimental conditions	N	N	Y
16	Experimental Factor Ontology	The Experimental Factor Ontology	Experimental conditions	N	Y	Y
17	Human Phenotype Ontology	A structured and controlled vocabu	Phenotype	N	N	N
18	Mammalian Phenotype Ontology	Standard terms for annotating mam	Phenotype	N	N	N
19	Units of measurement ontology	Metrical units for use in conjunction	Unit	N	N	Y
20	Cellular microscopy phenotype ontology	The Cellular Microscopy Phenotype	Cellular phenotype	N	N	N
21	ENSEMBL gene	Ensembl annotate genes, compute	Gene	N	Y	Y
22	NCBI Gene	A searchable database of genes, fc	Gene	N	Y	N
23	Uniprot	UniProt is the world's leading high-	Protein	N	Y	Y
24	Medical Subject Headings (MeSH)	The Medical Subject Headings (Me	Controlled vocabulary	N	N	Y



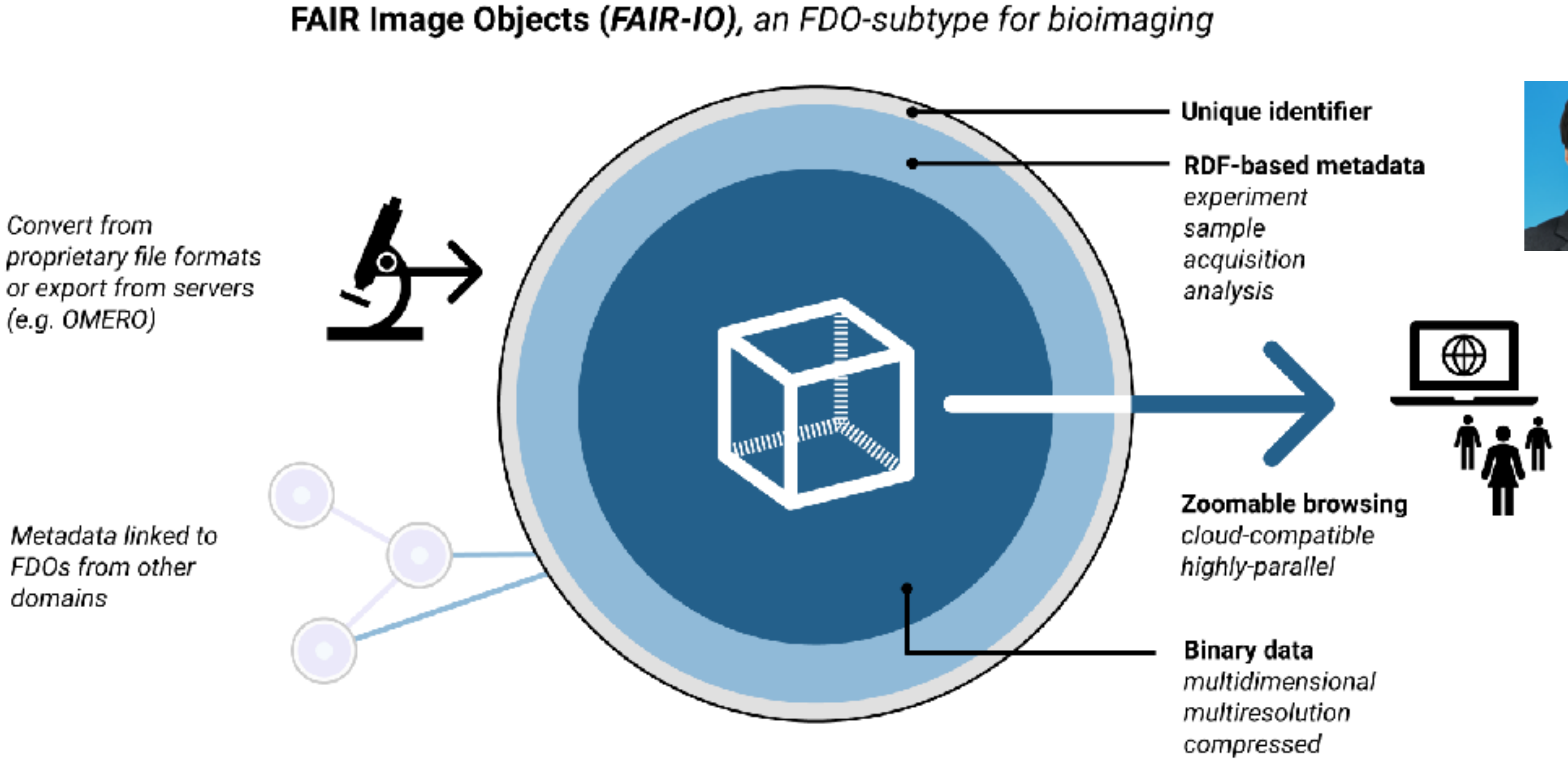
foundingGIDE (**G**lobal **I**mage **D**ata **E**cosystem)

Funded by the European Union


Matthew Hartley


Shuichi Onami

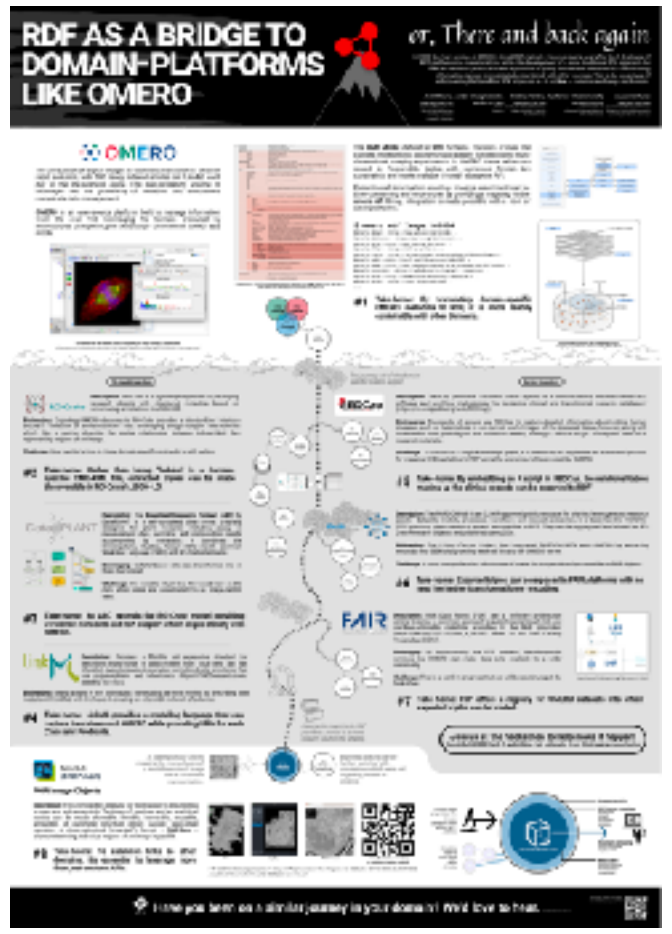
FAIR-IO (2023-2028)



https://zenodo.org/doi/10.5281/zenodo.7394674

(from the NFDI4BIOIMAGE grant proposal)

Norio Kobayashi
RIKEN
SWAT4HCLS
2018 & 2019

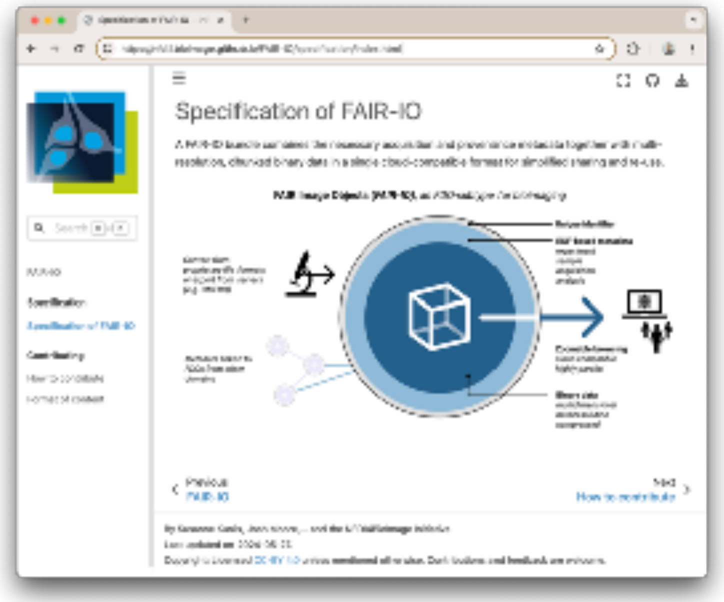


RDF AS A BRIDGE TO DOMAIN-PLATFORMS LIKE OMERO

or, There and back again

& omero-rdf
(PyPI)

https://zenodo.org/doi/10.5281/zenodo.10687658

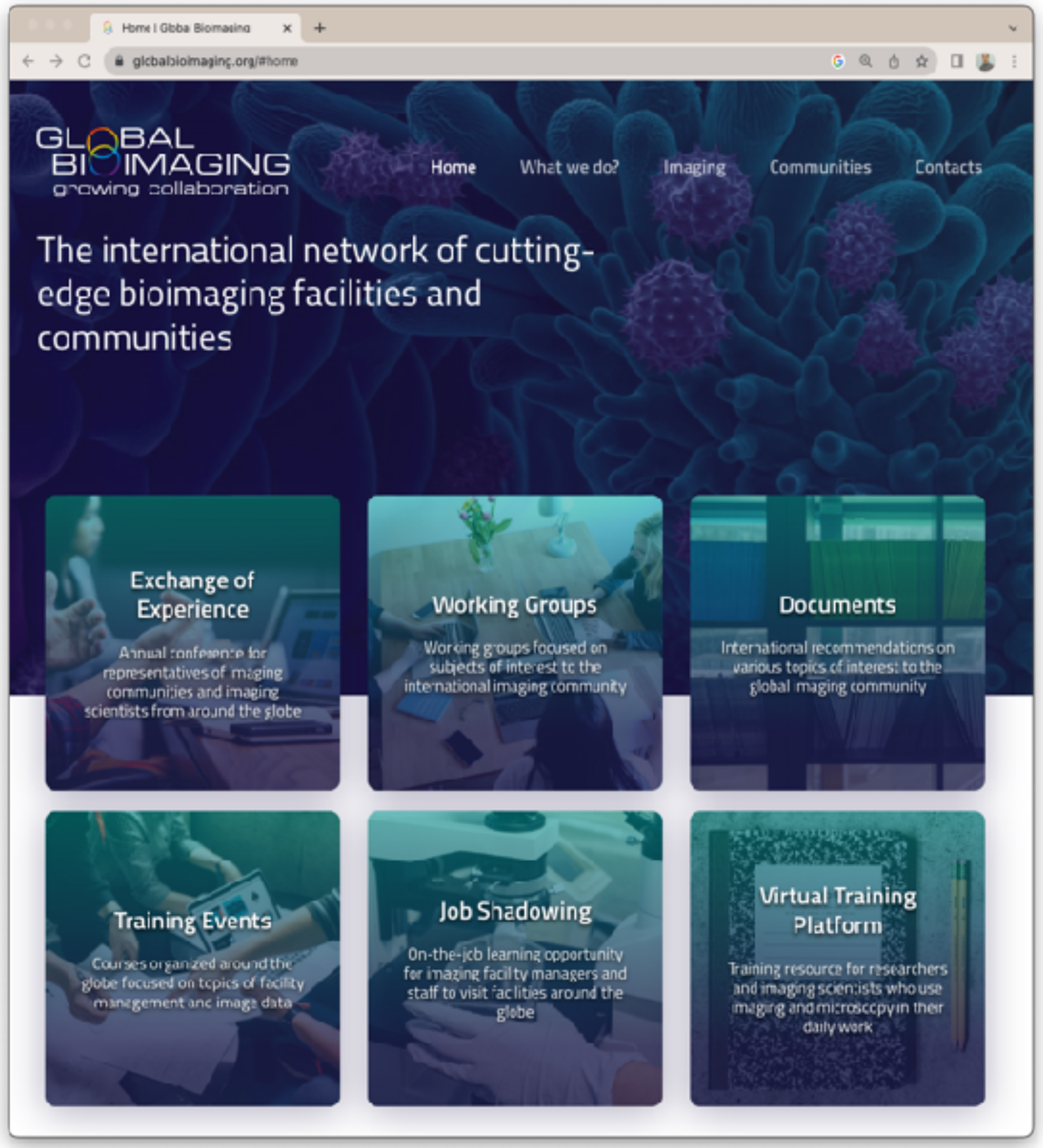


Specification of FAIR-IO

https://nfdi4bioimage.github.io/FAIR-IO/specification/

Susanne Kunis

Upcoming GBI working group



The international network of cutting-edge bioimaging facilities and communities

Exchange of Experience
Annual conference for representatives of imaging communities and imaging scientists from around the globe

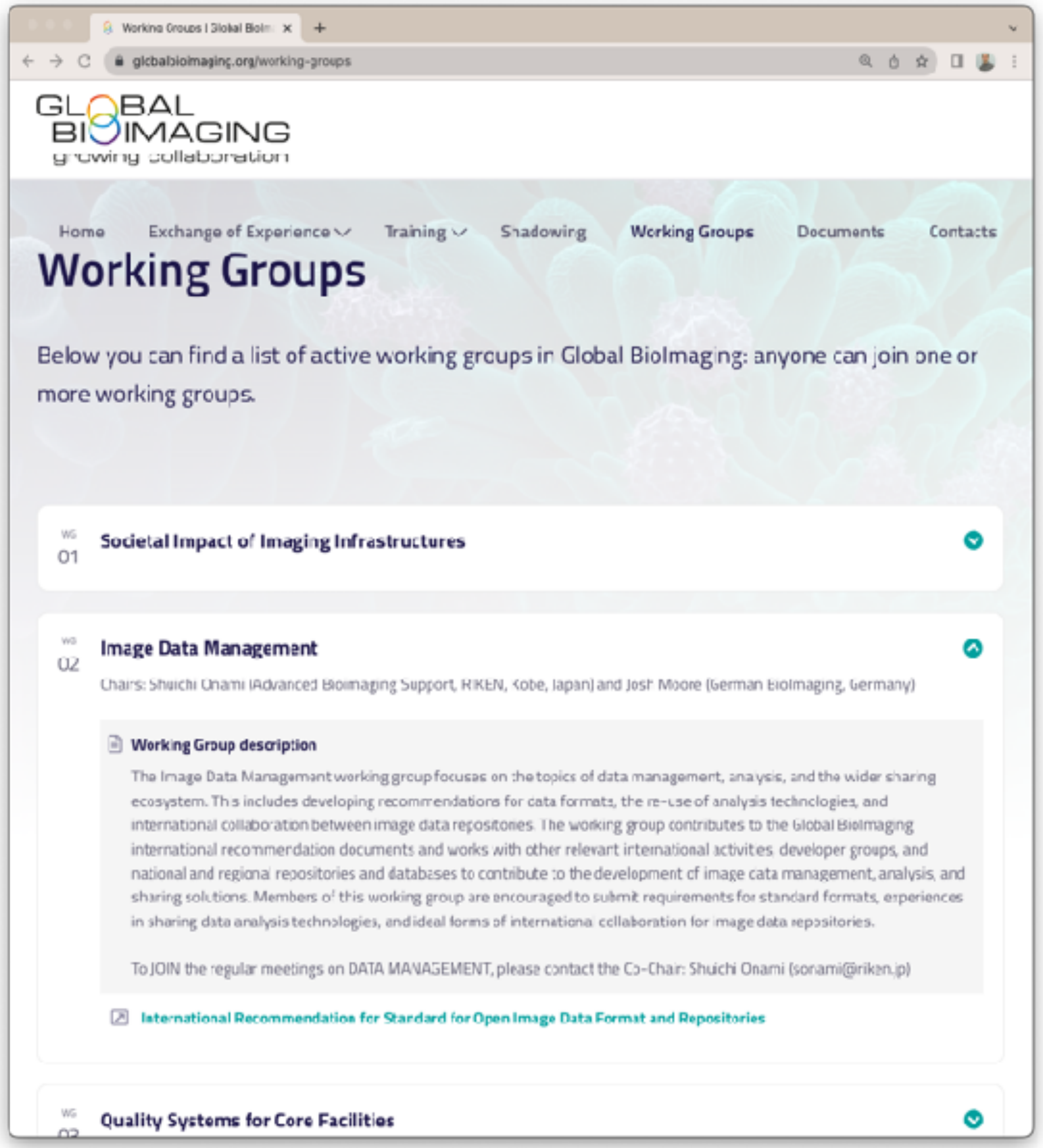
Working Groups
Working groups focused on subjects of interest to the international imaging community

Documents
International recommendations on various topics of interest to the global imaging community

Training Events
Courses organized around the globe focused on topics of facility management and image data

Job Shadowing
On-the-job learning opportunity for imaging facility managers and staff to visit facilities around the globe

Virtual Training Platform
Training resource for researchers and imaging scientists who use imaging and microscopy in their daily work



Working Groups

Below you can find a list of active working groups in Global BioImaging: anyone can join one or more working groups.

WG 01 **Societal Impact of Imaging Infrastructures**

WG 02 **Image Data Management**

Chairs: Shuichi Onami (Advanced Bioimaging Support, RIKEN, Kobe, Japan) and Josh Moore (German BioImaging, Germany)

Working Group description

The Image Data Management working group focuses on the topics of data management, analysis, and the wider sharing ecosystem. This includes developing recommendations for data formats, the re-use of analysis technologies, and international collaboration between image data repositories. The working group contributes to the Global BioImaging international recommendation documents and works with other relevant international activities, developer groups, and national and regional repositories and databases to contribute to the development of image data management, analysis, and sharing solutions. Members of this working group are encouraged to submit requirements for standard formats, experiences in sharing data analysis technologies, and ideal forms of international collaboration for image data repositories.

To JOIN the regular meetings on DATA MANAGEMENT, please contact the Co-Chair: Shuichi Onami (sonami@riken.jp)

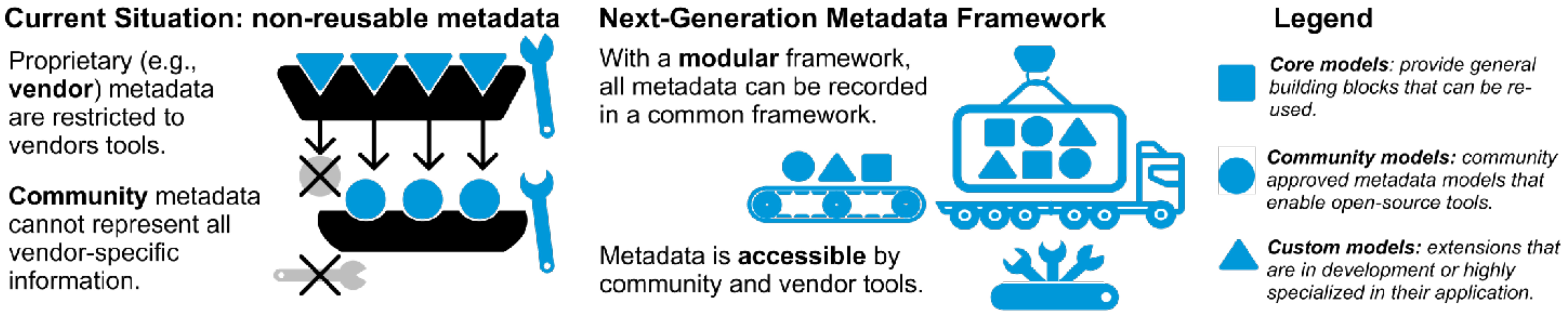
International Recommendation for Standard for Open Image Data Format and Repositories

WG 03 **Quality Systems for Core Facilities**

https://globalbioimaging.org/working-groups

Persistent Hardware Descriptors (2023)

(NSF proposal)



Current Situation: non-reusable metadata

Proprietary (e.g., **vendor**) metadata are restricted to vendors tools.

Community metadata cannot represent all vendor-specific information.

Next-Generation Metadata Framework

With a **modular** framework, all metadata can be recorded in a common framework.

Metadata is **accessible** by community and vendor tools.

Legend

- ■ ***Core models***: *provide general building blocks that can be re-used.*
- ● ***Community models:*** *community approved metadata models that enable open-source tools.*
- ▲ ***Custom models:*** *extensions that are in development or highly specialized in their application.*

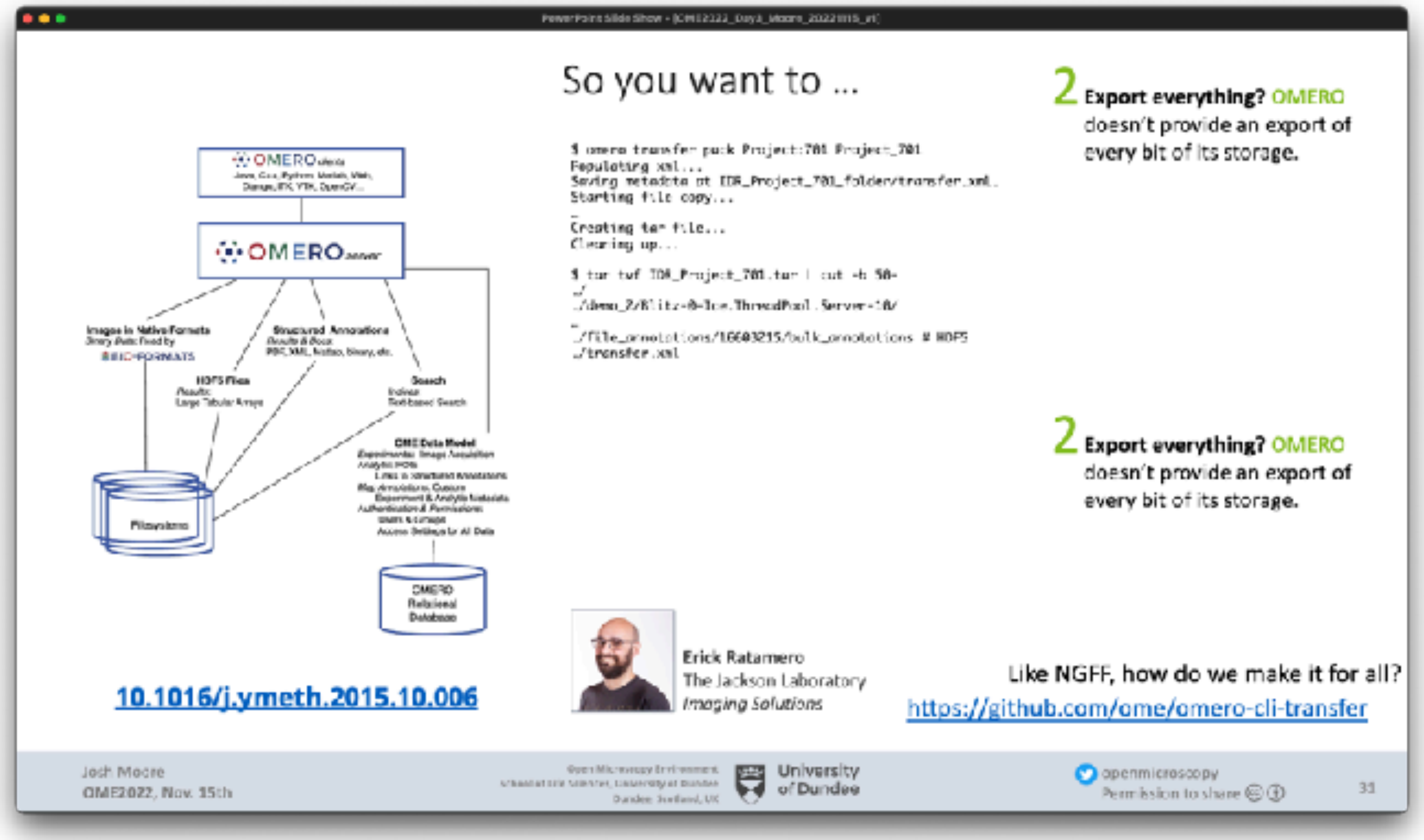
https://zenodo.org/doi/10.5281/zenodo.11265016

Caterina
Strambio De Castillia

What’s the plan?

(short, middle, long)

Mini-schemas (XSD)



So you want to ...

```
$ omero transfer pack Project:701 Project_701
Populating xml...
Saving metadata at IDR_Project_701_folder/transfer.xml.
Starting file copy...

Creating tar file...
Cleaning up...
```

2 Export everything? OMERO doesn't provide an export of every bit of its storage.

2 Export everything? OMERO doesn't provide an export of every bit of its storage.

10.1016/j.ymeth.2015.10.006

Erick Ratamero
The Jackson Laboratory
Imaging Solutions

Like NGFF, how do we make it for all?
https://github.com/ome/omero-cli-transfer

Josh Moore
OME2022, Nov. 15th

University of Dundee

openmicroscopy
Permission to share

33

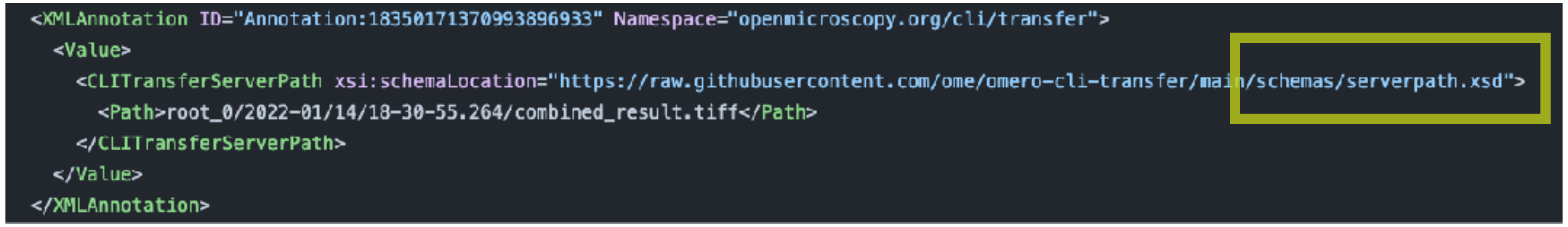
omero-cli-transfer

Installation

Usage

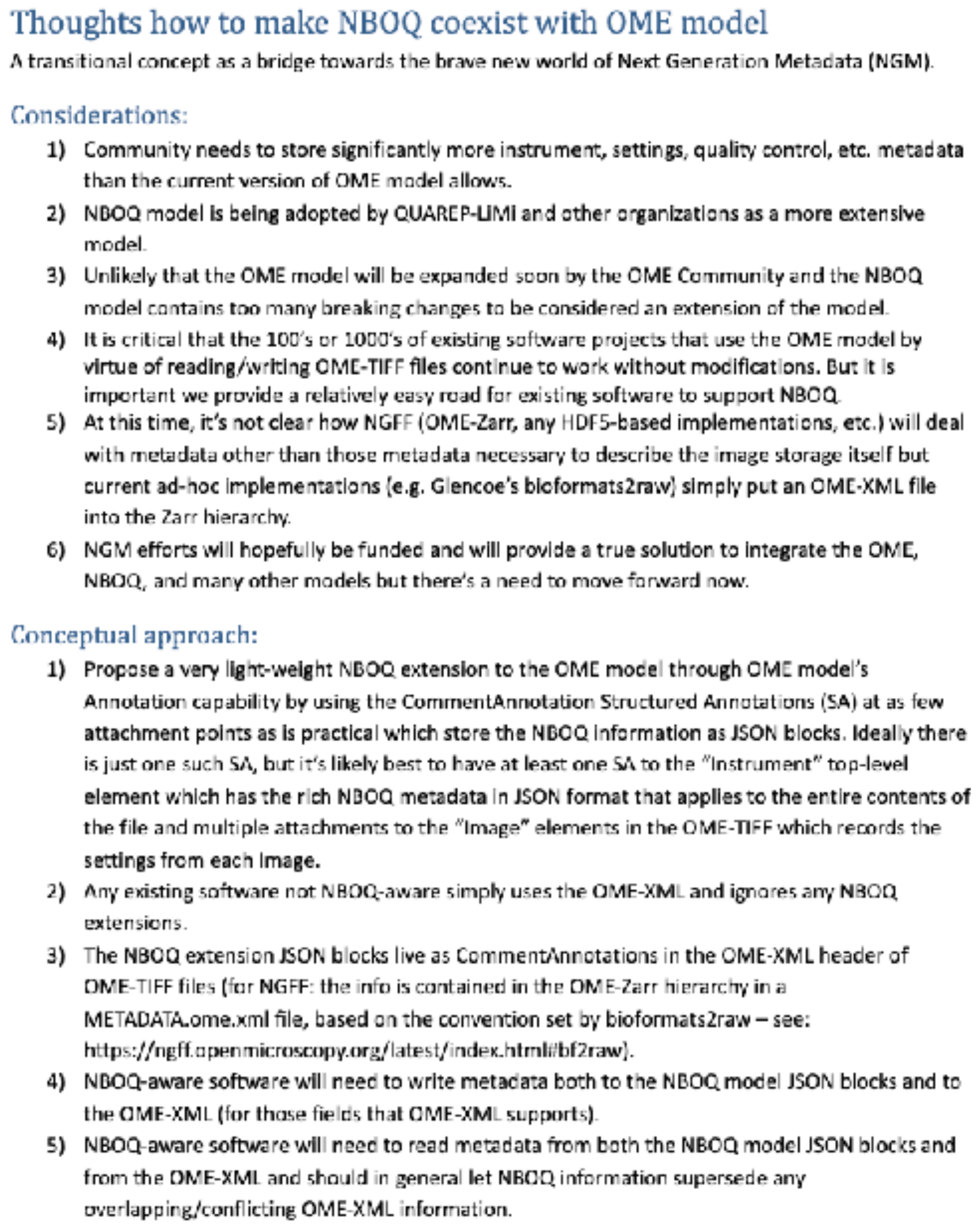
ome-types

Installation



```
<XMLAnnotation ID="Annotation:18350171370993896933" Namespace="openmicroscopy.org/cli/transfer">
  <Value>
    <CLITransferServerPath xsi:schemaLocation="https://raw.githubusercontent.com/ome/omero-cli-transfer/main/schemas/serverpath.xsd">
      <Path>root_0/2022-01/14/18-30-55.264/combined_result.tiff</Path>
    </CLITransferServerPath>
  </Value>
</XMLAnnotation>
```

https://github.com/ome/omero-cli-transfer/blob/main/test/data/transfer.xml



Thoughts how to make NBOQ coexist with OME model

A transitional concept as a bridge towards the brave new world of Next Generation Metadata (NGM).

Considerations:

1) Community needs to store significantly more instrument, settings, quality control, etc. metadata than the current version of OME model allows.
2) NBOQ model is being adopted by QUAREP-LIMI and other organizations as a more extensive model.
3) Unlikely that the OME model will be expanded soon by the OME Community and the NBOQ model contains too many breaking changes to be considered an extension of the model.
4) It is critical that the 100's or 1000's of existing software projects that use the OME model by virtue of reading/writing OME-TIFF files continue to work without modifications. But it is important we provide a relatively easy road for existing software to support NBOQ.
5) At this time, it's not clear how NGFF (OME-Zarr, any HDF5-based implementations, etc.) will deal with metadata other than those metadata necessary to describe the image storage itself but current ad-hoc implementations (e.g. Glencoe's bioformats2raw) simply put an OME-XML file into the Zarr hierarchy.
6) NGM efforts will hopefully be funded and will provide a true solution to integrate the OME, NBOQ, and many other models but there's a need to move forward now.

Conceptual approach:

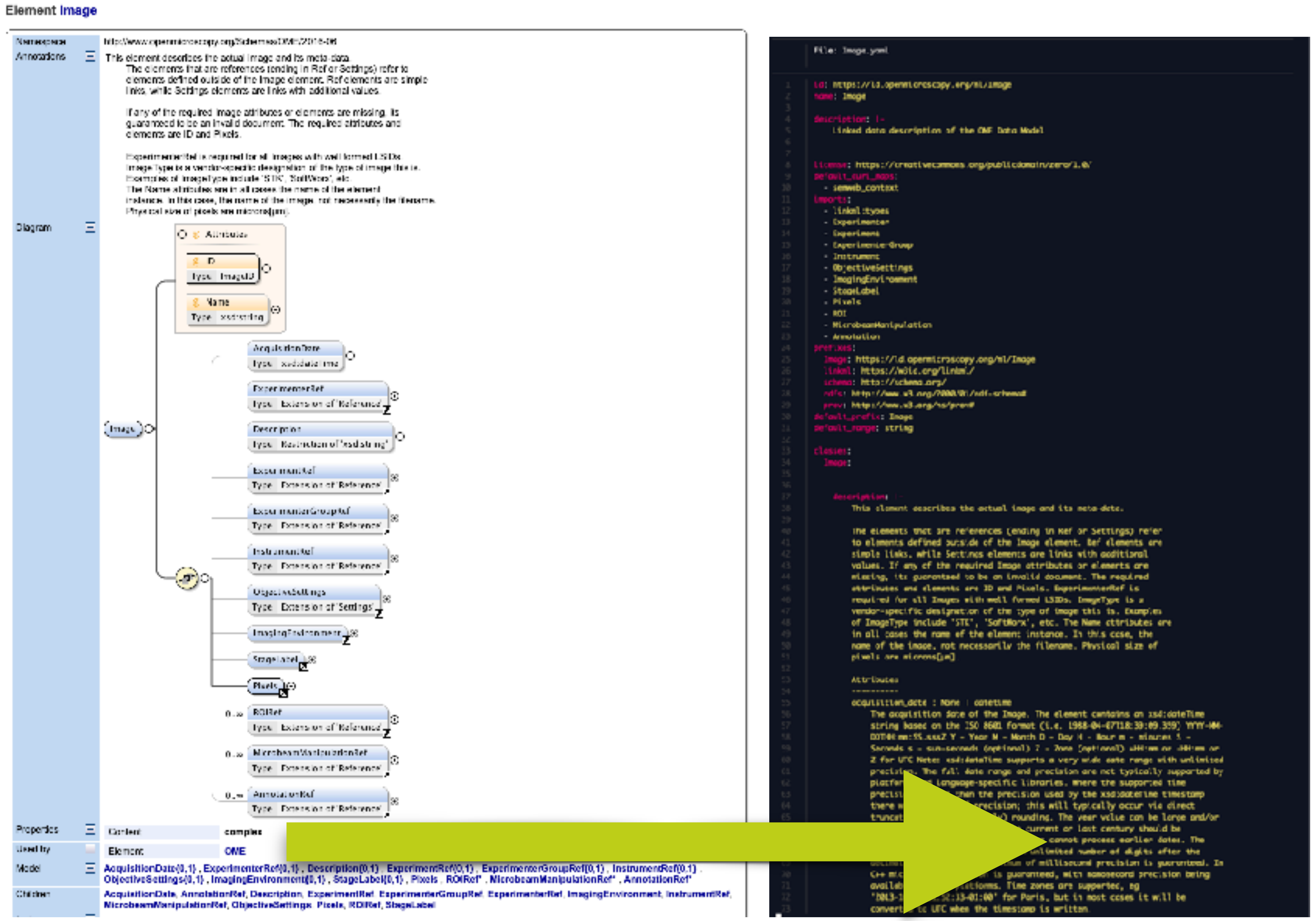
1) Propose a very light-weight NBOQ extension to the OME model through OME model's Annotation capability by using the CommentAnnotation Structured Annotations (SA) at as few attachment points as is practical which store the NBOQ information as JSON blocks. Ideally there is just one such SA, but it's likely best to have at least one SA to the "Instrument" top-level element which has the rich NBOQ metadata in JSON format that applies to the entire contents of the file and multiple attachments to the "Image" elements in the OME-TIFF which records the settings from each image.
2) Any existing software not NBOQ-aware simply uses the OME-XML and ignores any NBOQ extensions.
3) The NBOQ extension JSON blocks live as CommentAnnotations in the OME-XML header of OME-TIFF files (for NGFF: the info is contained in the OME-Zarr hierarchy in a METADATA.ome.xml file, based on the convention set by bioformats2raw – see: https://ngff.openmicroscopy.org/latest/index.html#bf2raw).
4) NBOQ-aware software will need to write metadata both to the NBOQ model JSON blocks and to the OME-XML (for those fields that OME-XML supports).
5) NBOQ-aware software will need to read metadata from both the NBOQ model JSON blocks and from the OME-XML and should in general let NBOQ information supersede any overlapping/conflicting OME-XML information.

Erick Ratamero

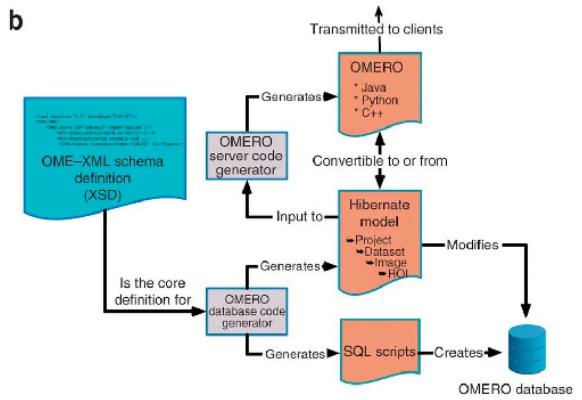
Caterina
Strambio De Castillia

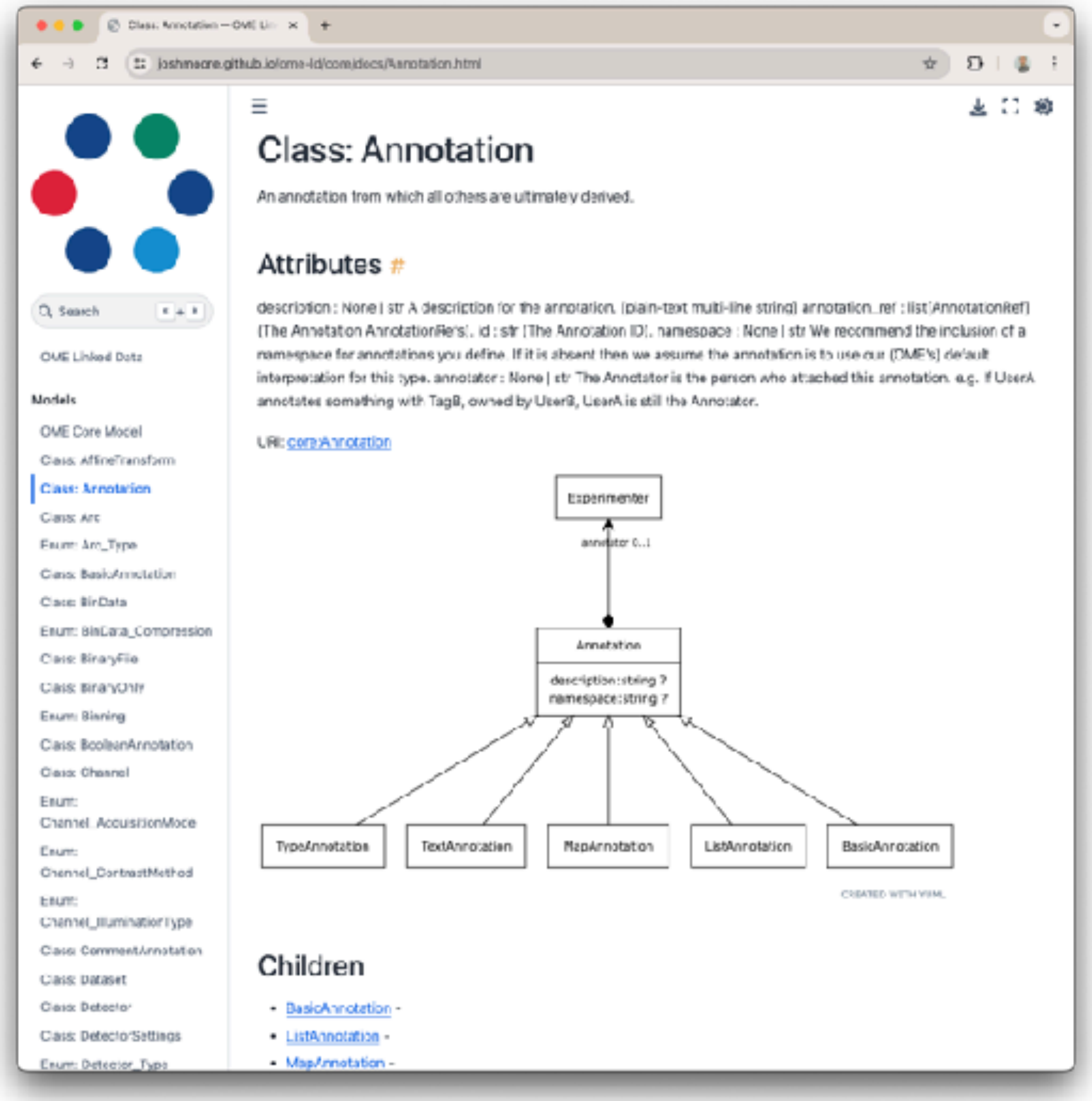
Damir Sudar

Linked Data (OME-LD)



Conversion of XSD to LinkML





Class: Annotation

An annotation from which all others are ultimately derived.

Attributes

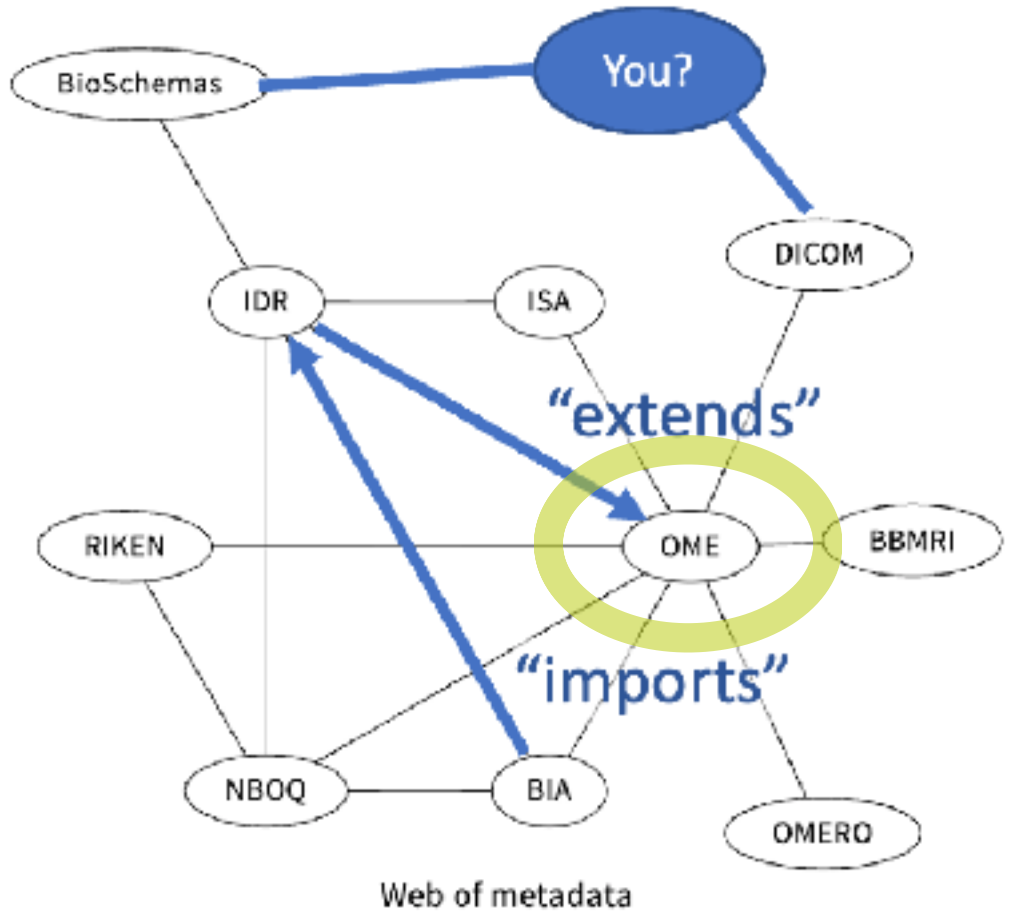
description : None | str A description for the annotation. (plain-text multi-line string) annotation_ref : list[AnnotationRef] [The Annotation AnnotationRefs]. id : str [The Annotation ID]. namespace : None | str We recommend the inclusion of a namespace for annotations you define. If it is absent then we assume the annotation is to use our (OME's) default interpretation for this type. annotator : None | str The Annotator is the person who attached this annotation. e.g. If UserA annotates something with TagB, owned by UserB, UserA is still the Annotator.

URI: core:Annotation

Children

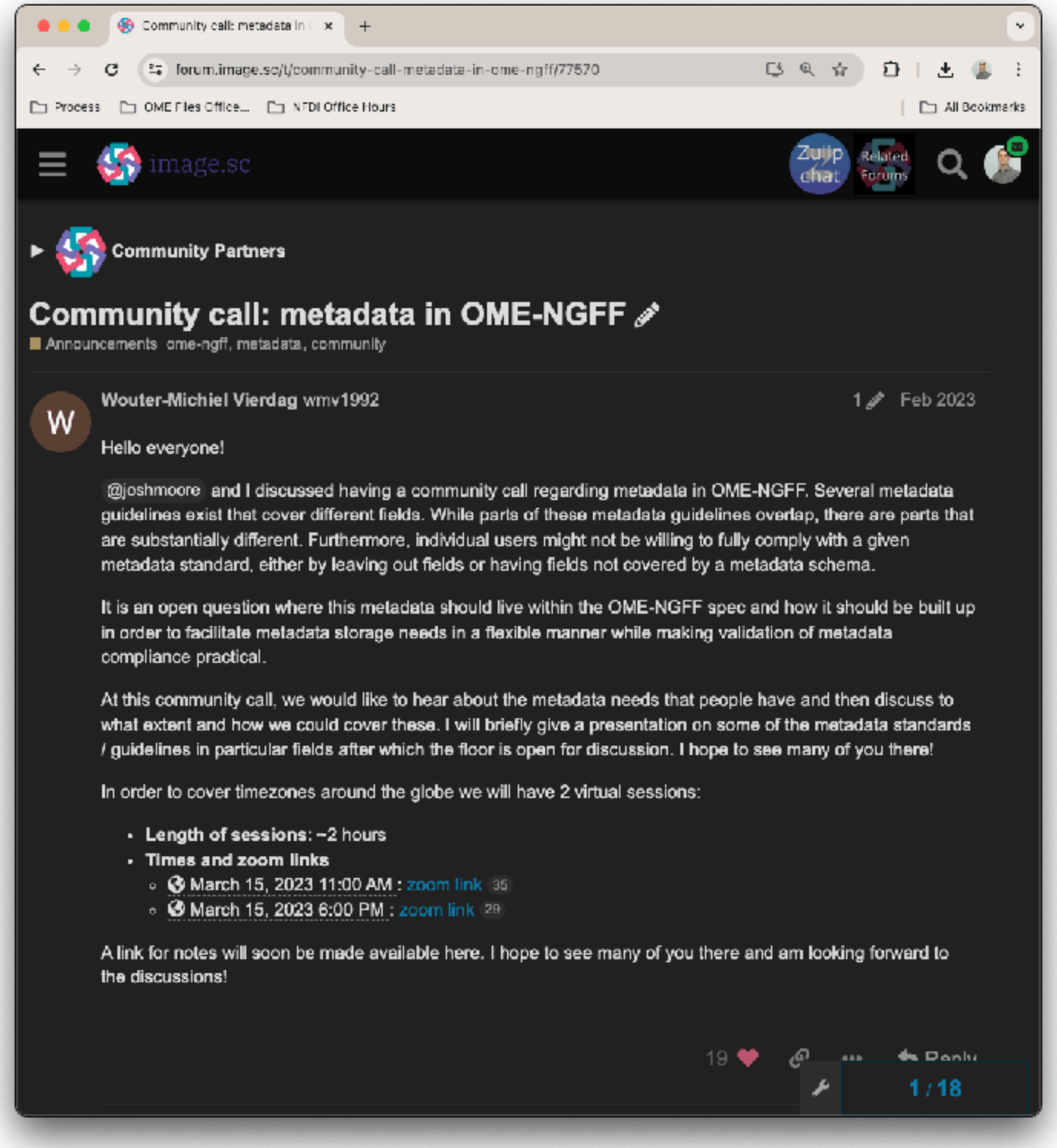
- BasicAnnotation -
- ListAnnotation -
- MapAnnotation -

https://joshmoore.github.io/ome-ld/core/docs/Annotation.html



OME**2022**

How to tie it all together?



Community call: metadata in OME-NGFF

Announcements ome-ngff, metadata, community

Wouter-Michiel Vierdag wmv1992 1 Feb 2023

Hello everyone!

@joshmoore and I discussed having a community call regarding metadata in OME-NGFF. Several metadata guidelines exist that cover different fields. While parts of these metadata guidelines overlap, there are parts that are substantially different. Furthermore, individual users might not be willing to fully comply with a given metadata standard, either by leaving out fields or having fields not covered by a metadata schema.

It is an open question where this metadata should live within the OME-NGFF spec and how it should be built up in order to facilitate metadata storage needs in a flexible manner while making validation of metadata compliance practical.

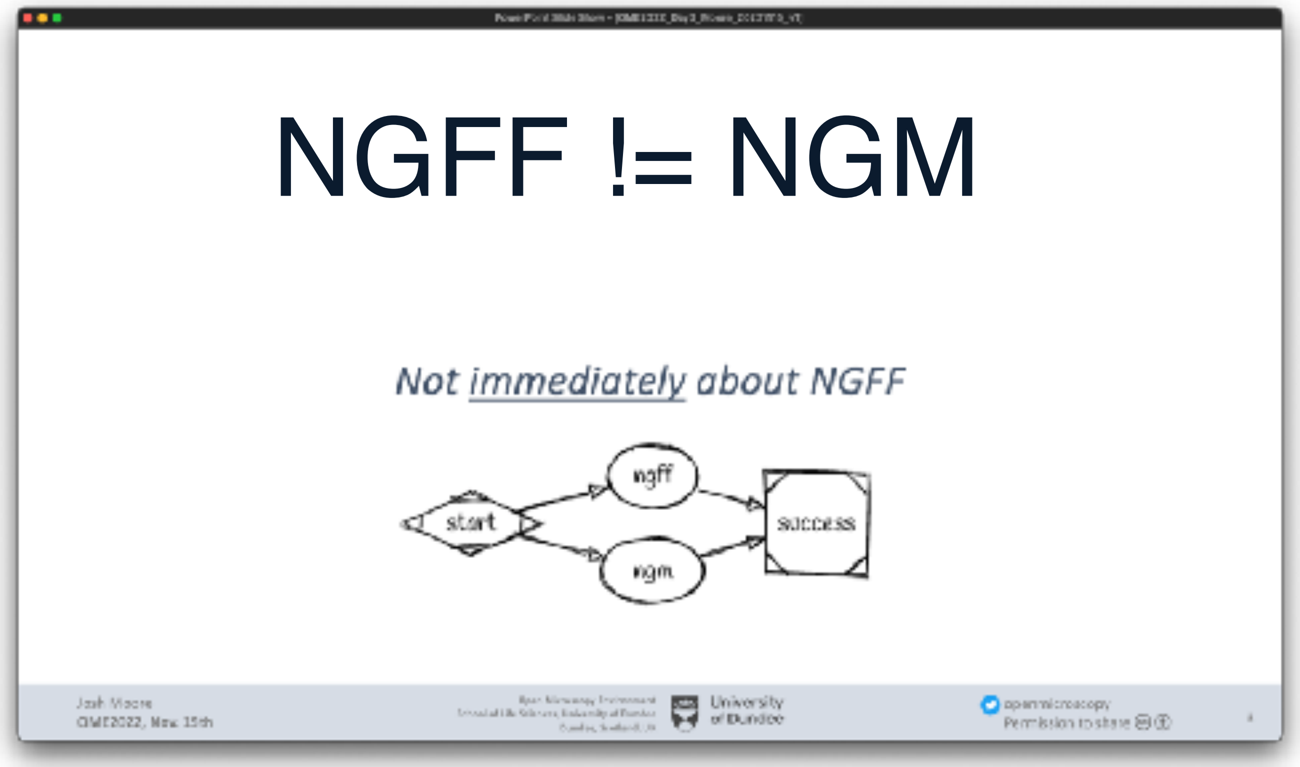
At this community call, we would like to hear about the metadata needs that people have and then discuss to what extent and how we could cover these. I will briefly give a presentation on some of the metadata standards / guidelines in particular fields after which the floor is open for discussion. I hope to see many of you there!

In order to cover timezones around the globe we will have 2 virtual sessions:

- **Length of sessions**: ~2 hours
- **Times and zoom links**
    - March 15, 2023 11:00 AM : zoom link 35
    - March 15, 2023 6:00 PM : zoom link 29

A link for notes will soon be made available here. I hope to see many of you there and am looking forward to the discussions!

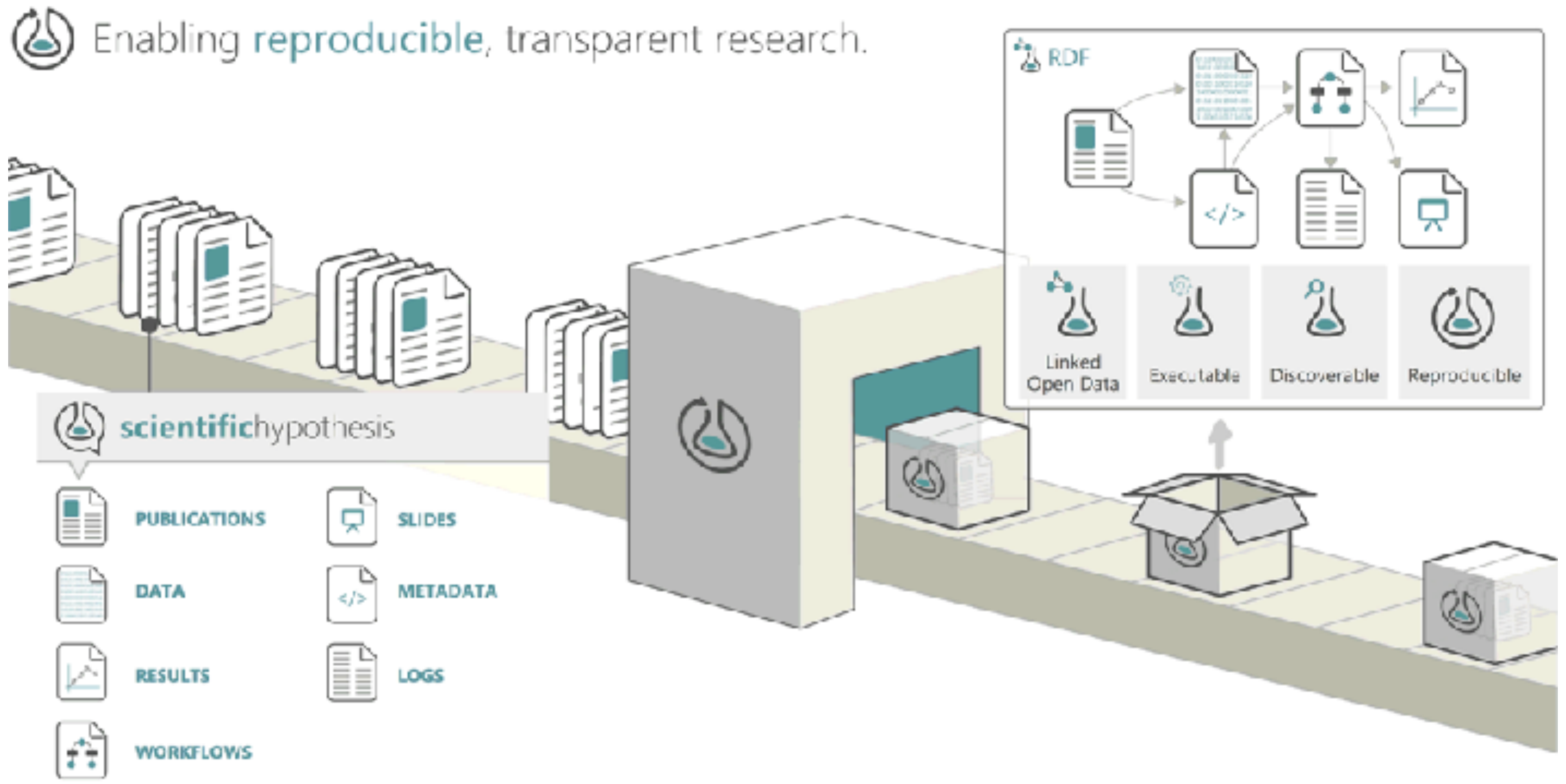
https://forum.image.sc/t/community-call-metadata-in-ome-ngff/77570



NGFF != NGM

Not immediately about NGFF

Zarr



Enabling reproducible, transparent research.

https://www.researchobject.org/

How to tie it all together?

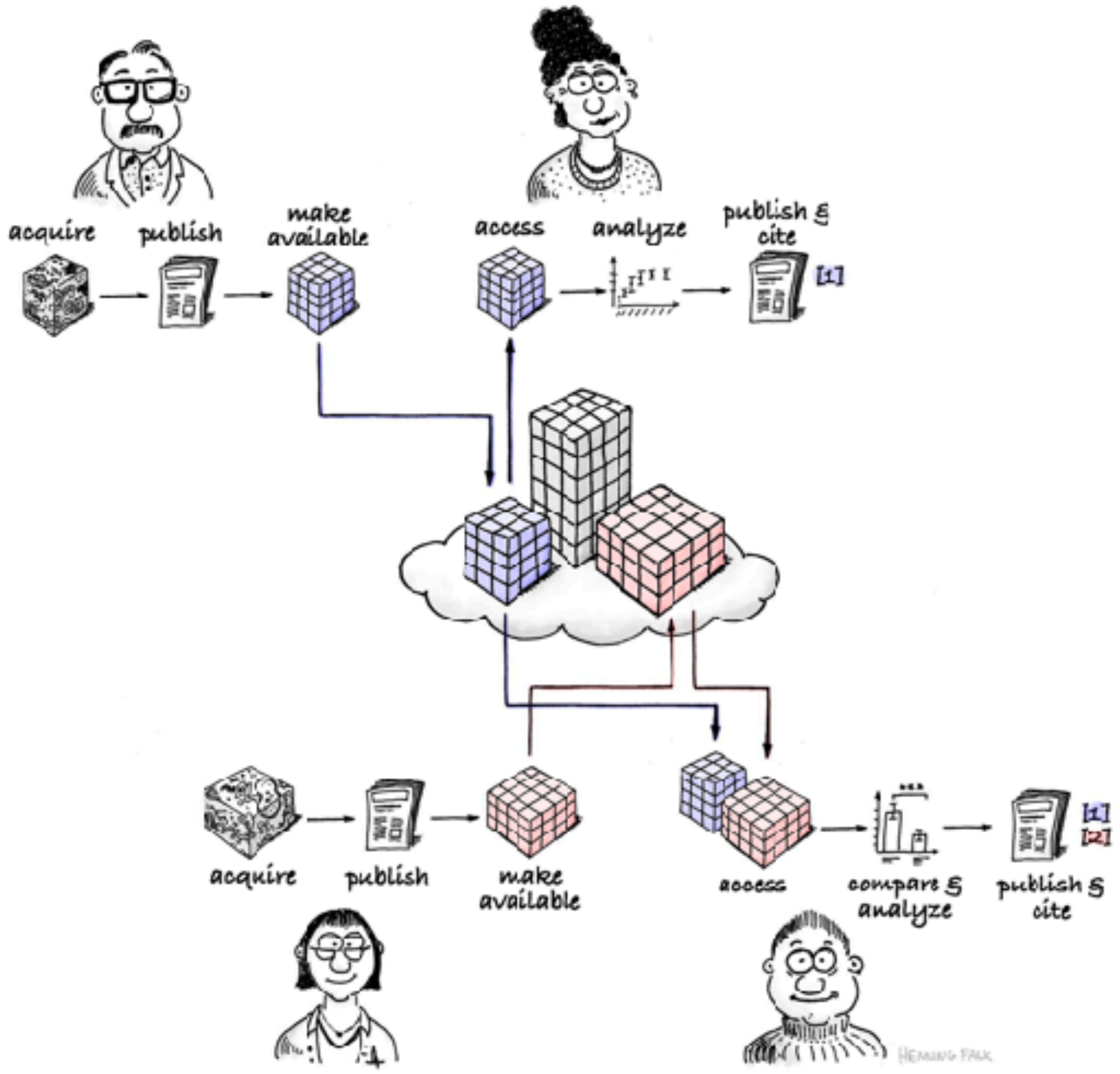
RO-Crate *is to* ***NGM*** *what* ***Zarr*** *is to* ***NGFF***

```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
    {
      "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
      "about": {"@id": "./"}
    },
    {
      "@id": "./",
      "@type": [
        "Dataset"
      ],
      "publisher": {“@id":
          “https://ror.org/05tpnw772”},

      "hasPart": [
        {
          "@id": “my.ome.zarr/”
        },
        {
          "@id": “tiffs/”
        },
        {
          "@id": “conversion.cwl/“
        }
      ],
   },

...
```

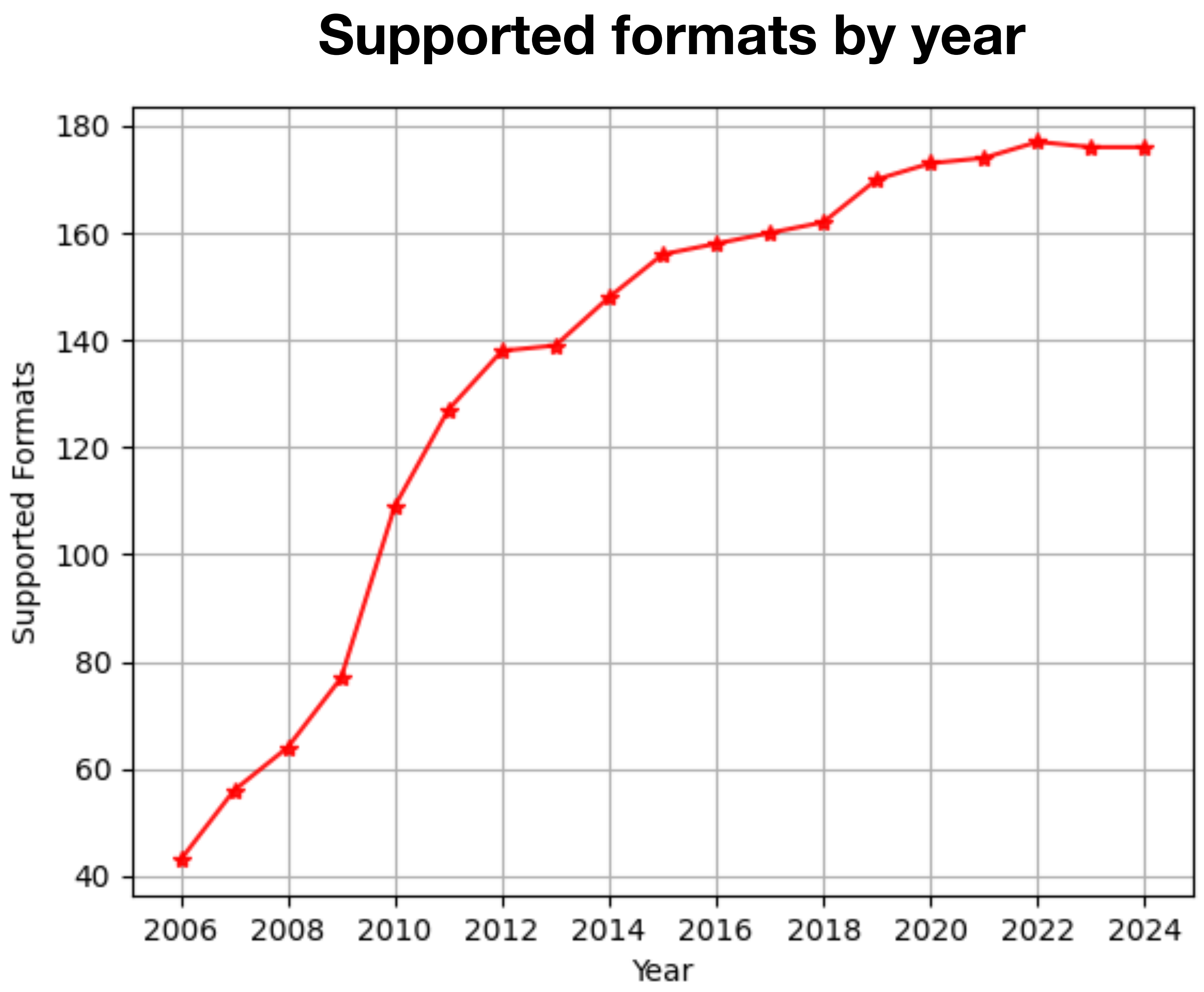
```
...
    {
      "@id": "https://ror.org/05tpnw772",
      "@type": "Organization",
      "name": “German Bio-Imaging“,
      "url": "https://ror.org/05tpnw772"
    },
    {
     "@id": “my.ome.zarr/”,
     "@type": "Dataset",
     "name": "Diagram showing trend to increase",
     "description": "Illustrator file for Glop Pot",
     "encodingFormat": “application/zarr”,
        {“@id": “https://ngff.openmicroscopy.org/0.4”}
    },
    {
     "@id": "tiffs/",
     "@type": "Dataset",
     "name": “Original image stack",
     "description": "This directory contains lots of TIFFs.”
    },
    {
      "@id": "conversion.cwl",
      "@type": "File",
      "name": “Process to create the OME-Zarr“,
      "contentSize": “120",
      "encodingFormat": ["text/plain", {"@id": "https://www.commonwl.org/v1.0/Workflow.html"}]
    }
  ]
}
```

JSON-LD (RDF)

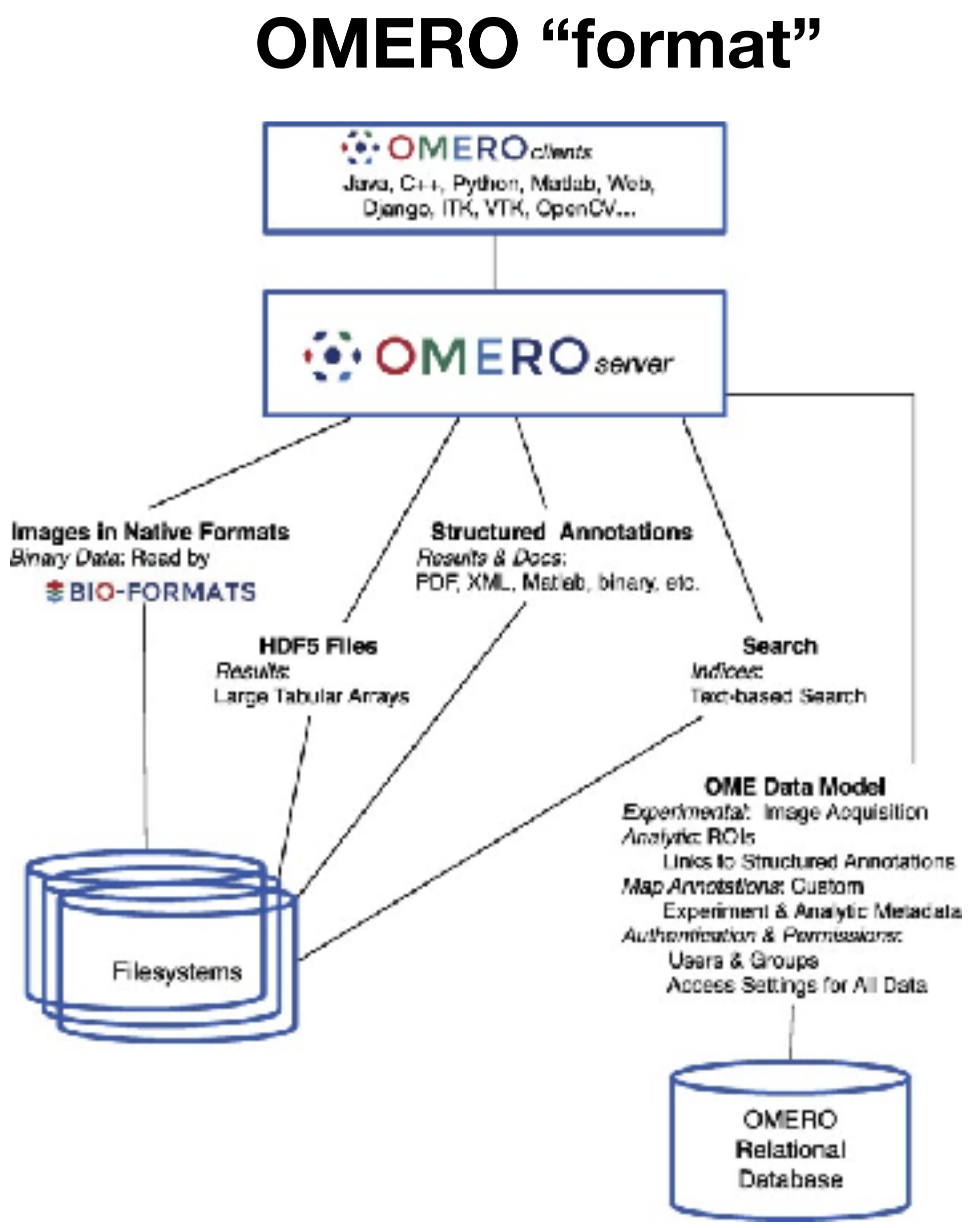
The FAIR Vision



"FAIR re-use" by @DrHenningFalk,
©2022 @NumFOCUS is used under a CC-BY 4.0 license.


Supported formats by year

https://bit.ly/bio-formats-list


OMERO "format"

https://doi.org/10.1016/j.ymeth.2015.10.006

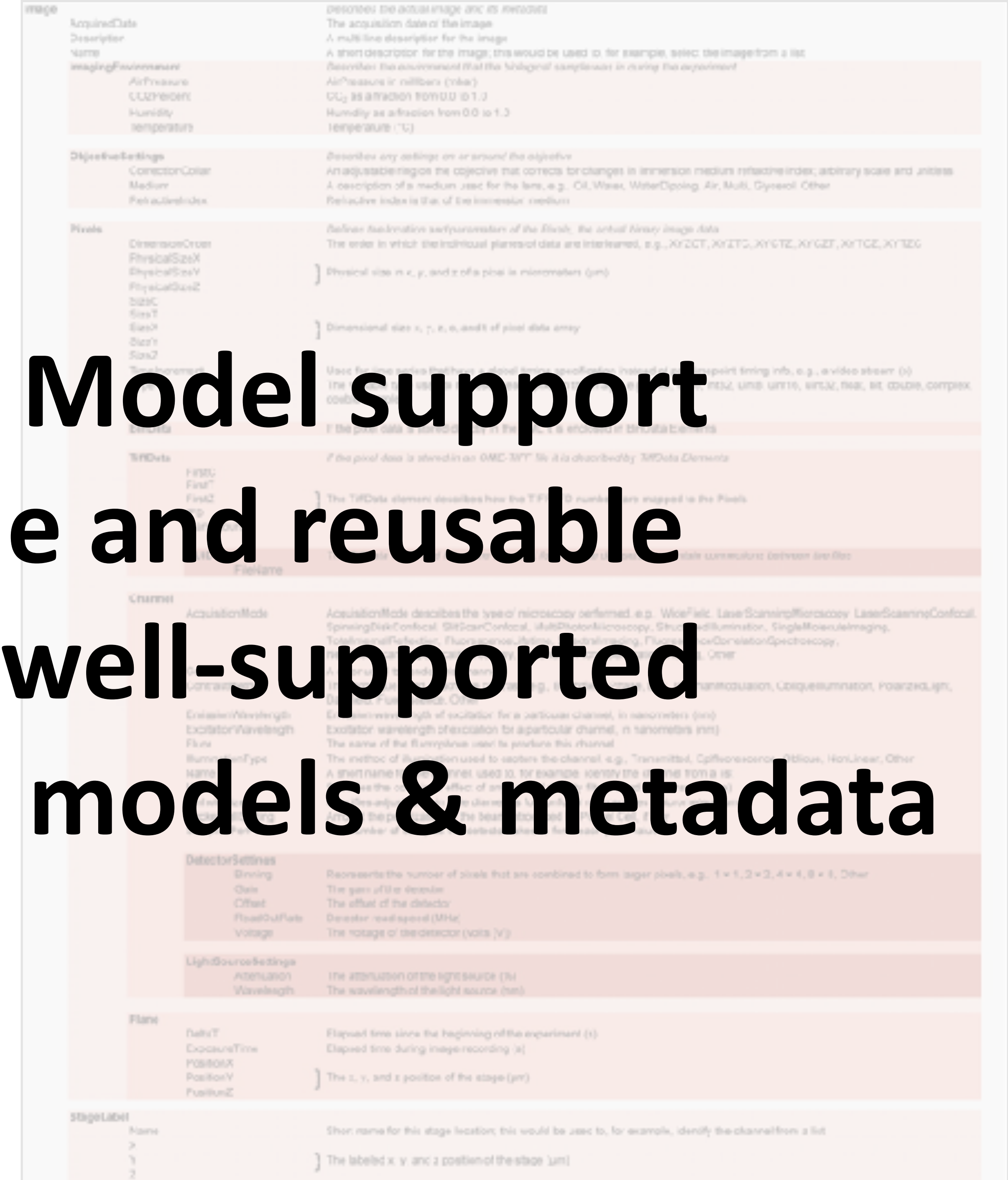
Metadata Matters 2024

- **Superset of OME Model support**
- **Models extensible and reusable**
- **Base models are well-supported**
- **Vendors produce models & metadata**

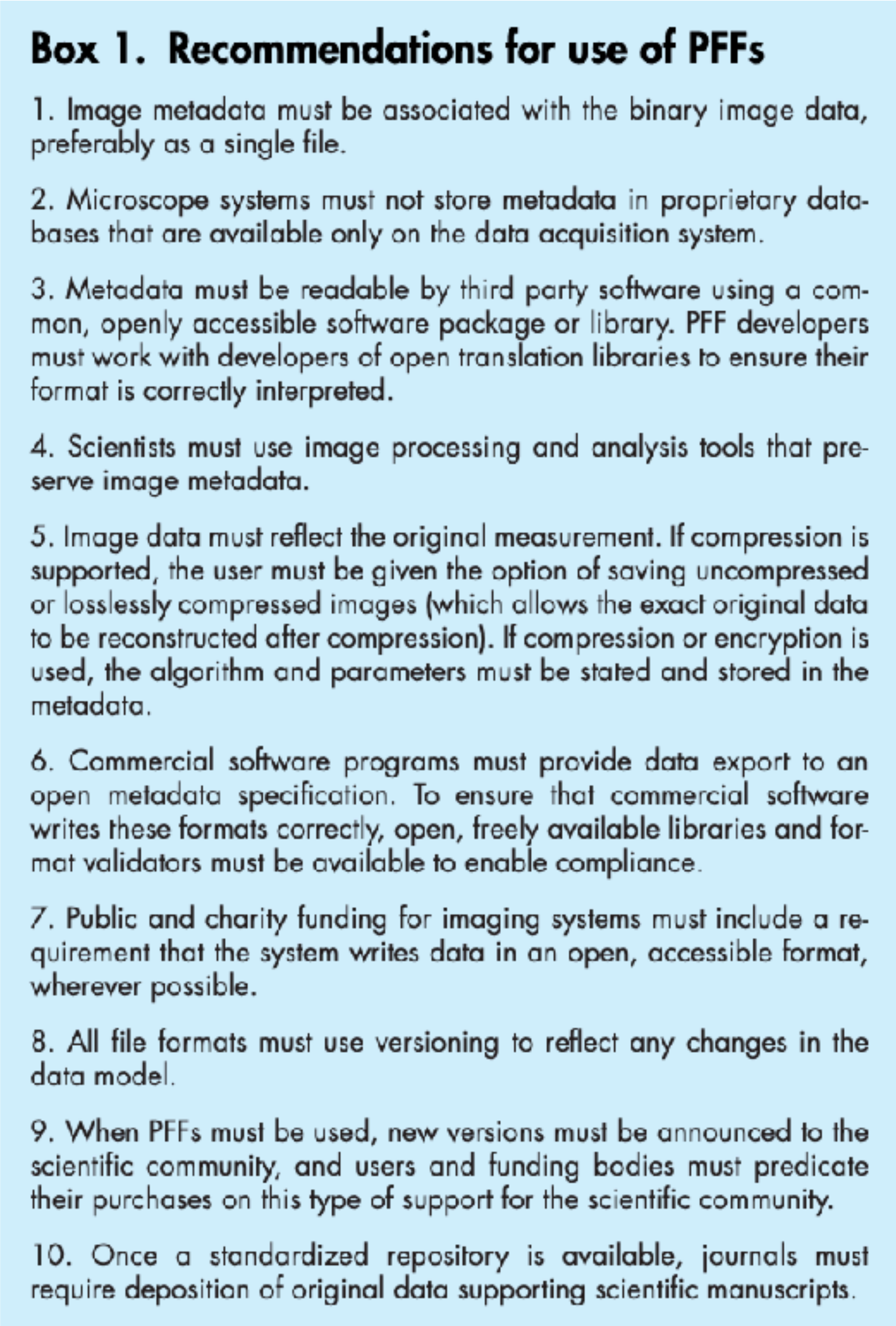
JCB: Comment

Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[5] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]



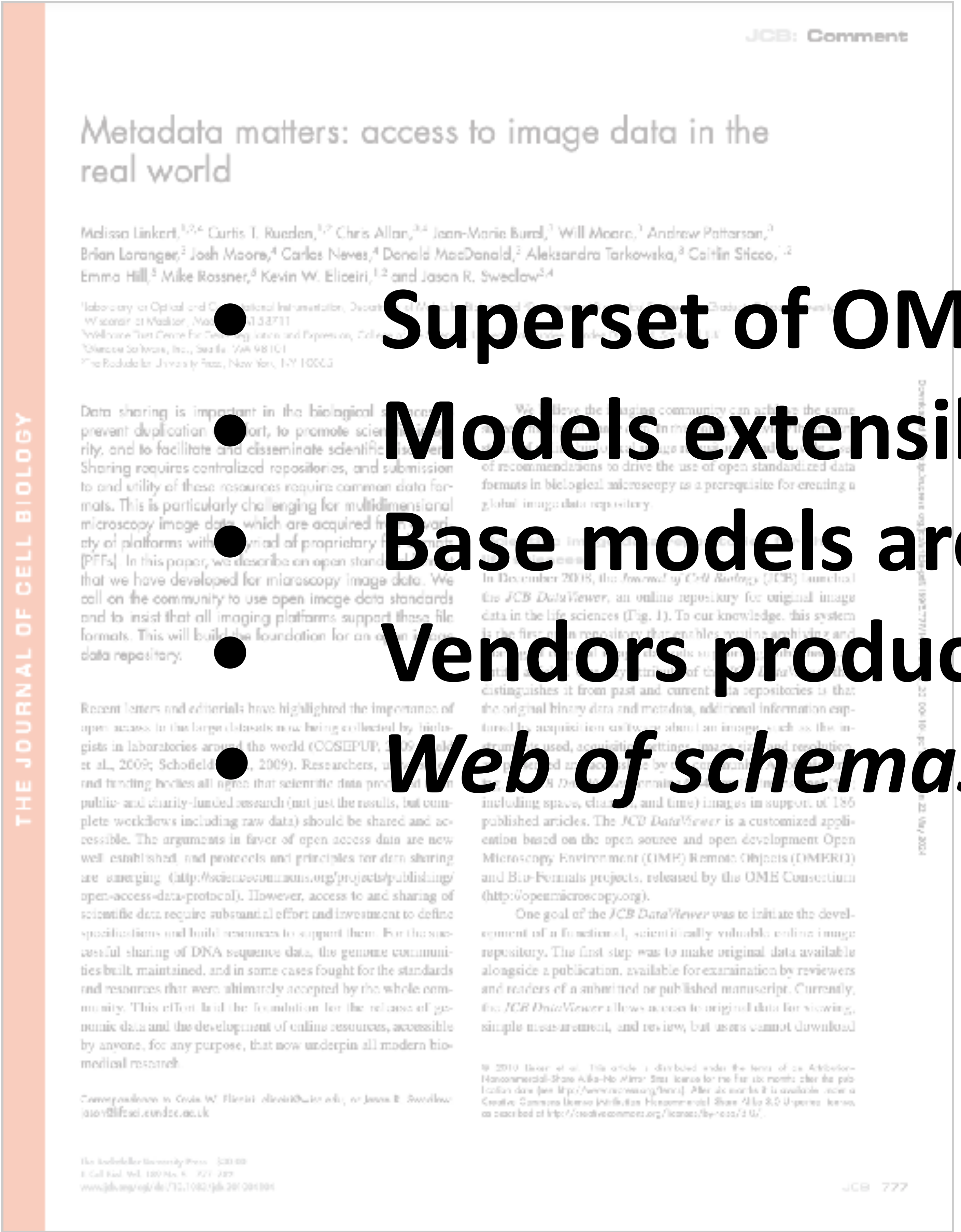
Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).



Box 1. Recommendations for use of PFFs

1. Image metadata must be associated with the binary image data, preferably as a single file.
2. Microscope systems must not store metadata in proprietary databases that are available only on the data acquisition system.
3. Metadata must be readable by third party software using a common, openly accessible software package or library. PFF developers must work with developers of open translation libraries to ensure their format is correctly interpreted.
4. Scientists must use image processing and analysis tools that preserve image metadata.
5. Image data must reflect the original measurement. If compression is supported, the user must be given the option of saving uncompressed or losslessly compressed images (which allows the exact original data to be reconstructed after compression). If compression or encryption is used, the algorithm and parameters must be stated and stored in the metadata.
6. Commercial software programs must provide data export to an open metadata specification. To ensure that commercial software writes these formats correctly, open, freely available libraries and format validators must be available to enable compliance.
7. Public and charity funding for imaging systems must include a requirement that the system writes data in an open, accessible format, wherever possible.
8. All file formats must use versioning to reflect any changes in the data model.
9. When PFFs must be used, new versions must be announced to the scientific community, and users and funding bodies must predicate their purchases on this type of support for the scientific community.
10. Once a standardized repository is available, journals must require deposition of original data supporting scientific manuscripts.

Metadata Matters 2024

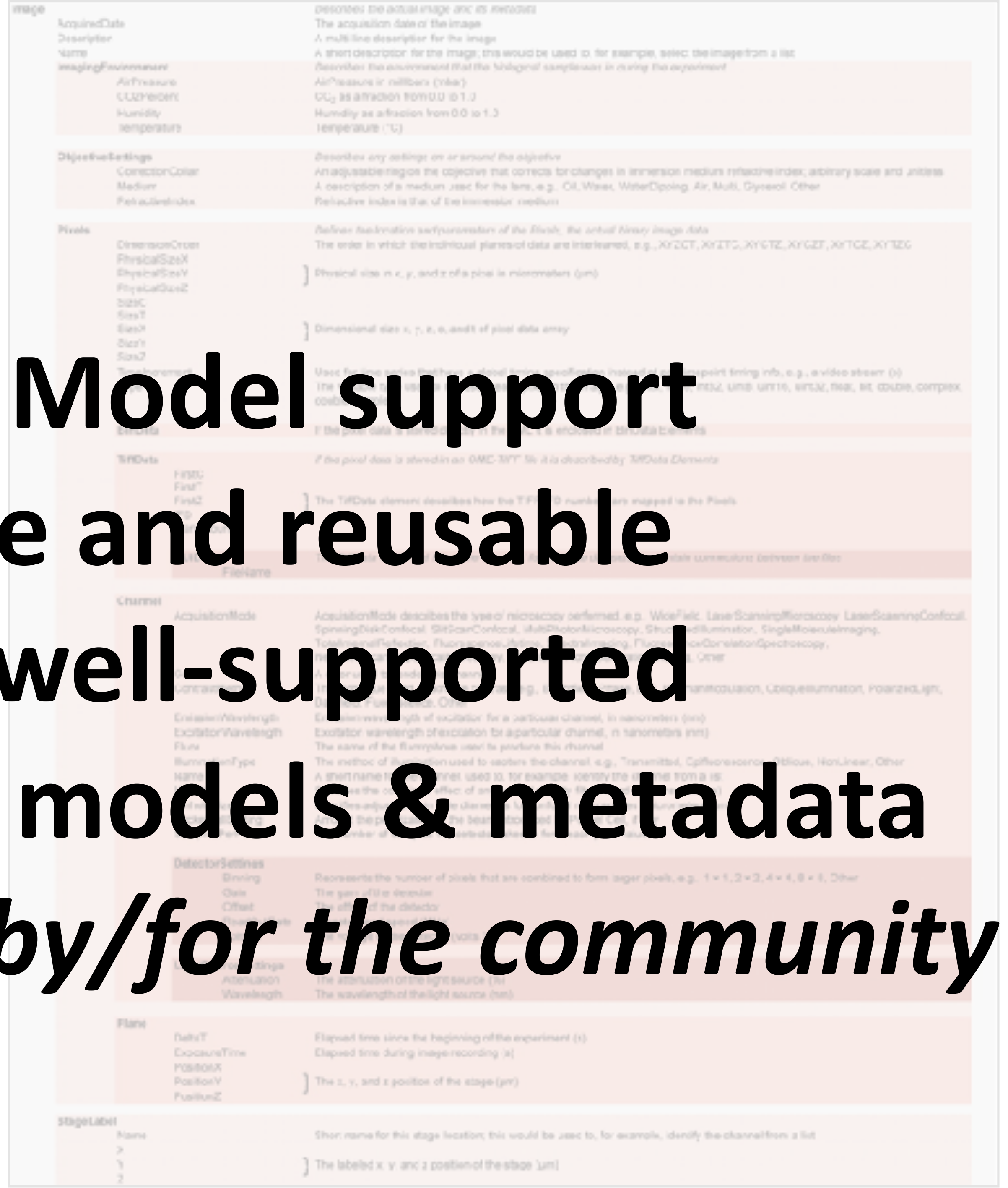


JCB: Comment

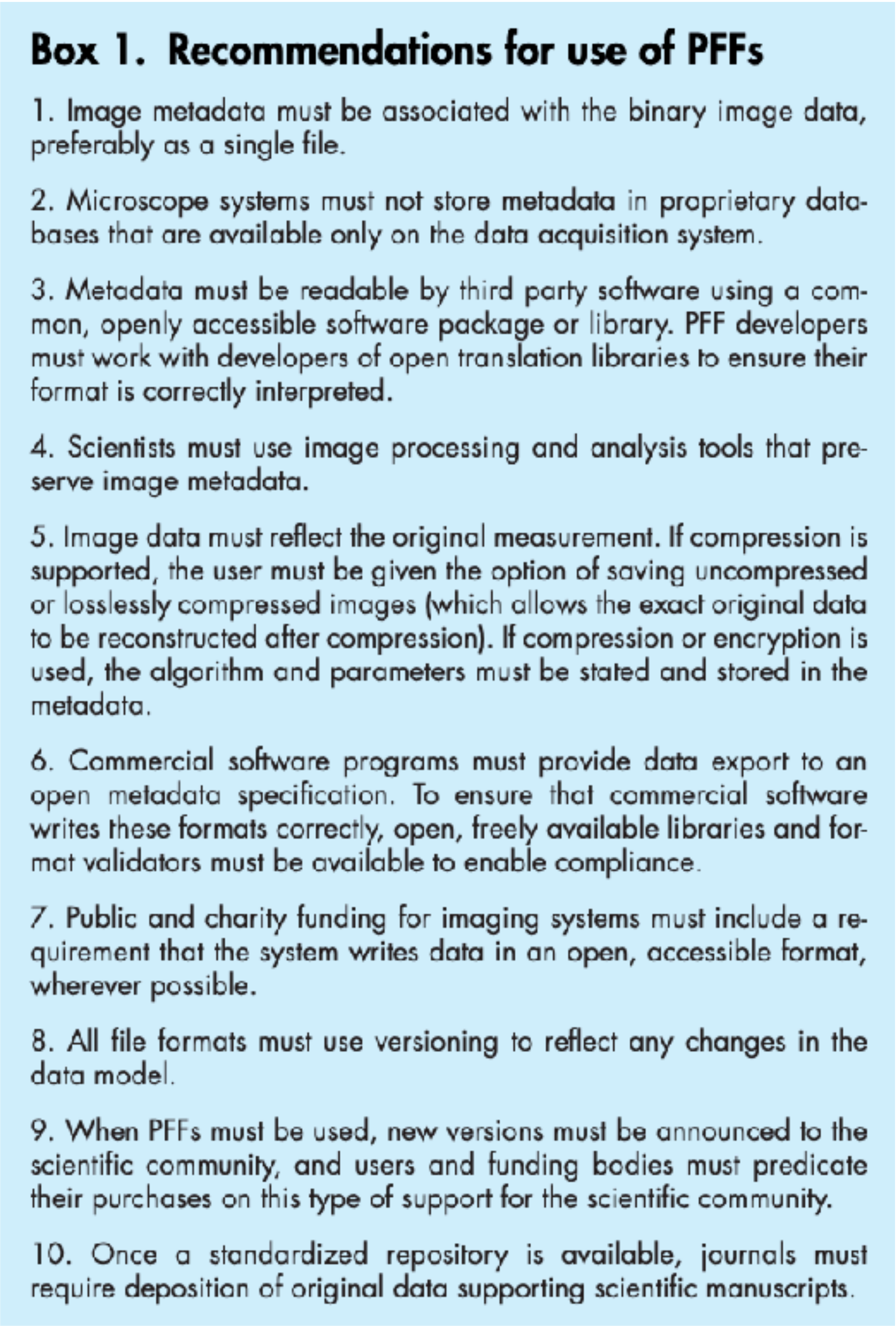
Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[5] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]

- **Superset of OME Model support**
- **Models extensible and reusable**
- **Base models are well-supported**
- **Vendors produce models & metadata**
- ***Web of schemas by/for the community***



Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).



Box 1. Recommendations for use of PFFs

1. Image metadata must be associated with the binary image data, preferably as a single file.
2. Microscope systems must not store metadata in proprietary databases that are available only on the data acquisition system.
3. Metadata must be readable by third party software using a common, openly accessible software package or library. PFF developers must work with developers of open translation libraries to ensure their format is correctly interpreted.
4. Scientists must use image processing and analysis tools that preserve image metadata.
5. Image data must reflect the original measurement. If compression is supported, the user must be given the option of saving uncompressed or losslessly compressed images (which allows the exact original data to be reconstructed after compression). If compression or encryption is used, the algorithm and parameters must be stated and stored in the metadata.
6. Commercial software programs must provide data export to an open metadata specification. To ensure that commercial software writes these formats correctly, open, freely available libraries and format validators must be available to enable compliance.
7. Public and charity funding for imaging systems must include a requirement that the system writes data in an open, accessible format, wherever possible.
8. All file formats must use versioning to reflect any changes in the data model.
9. When PFFs must be used, new versions must be announced to the scientific community, and users and funding bodies must predicate their purchases on this type of support for the scientific community.
10. Once a standardized repository is available, journals must require deposition of original data supporting scientific manuscripts.

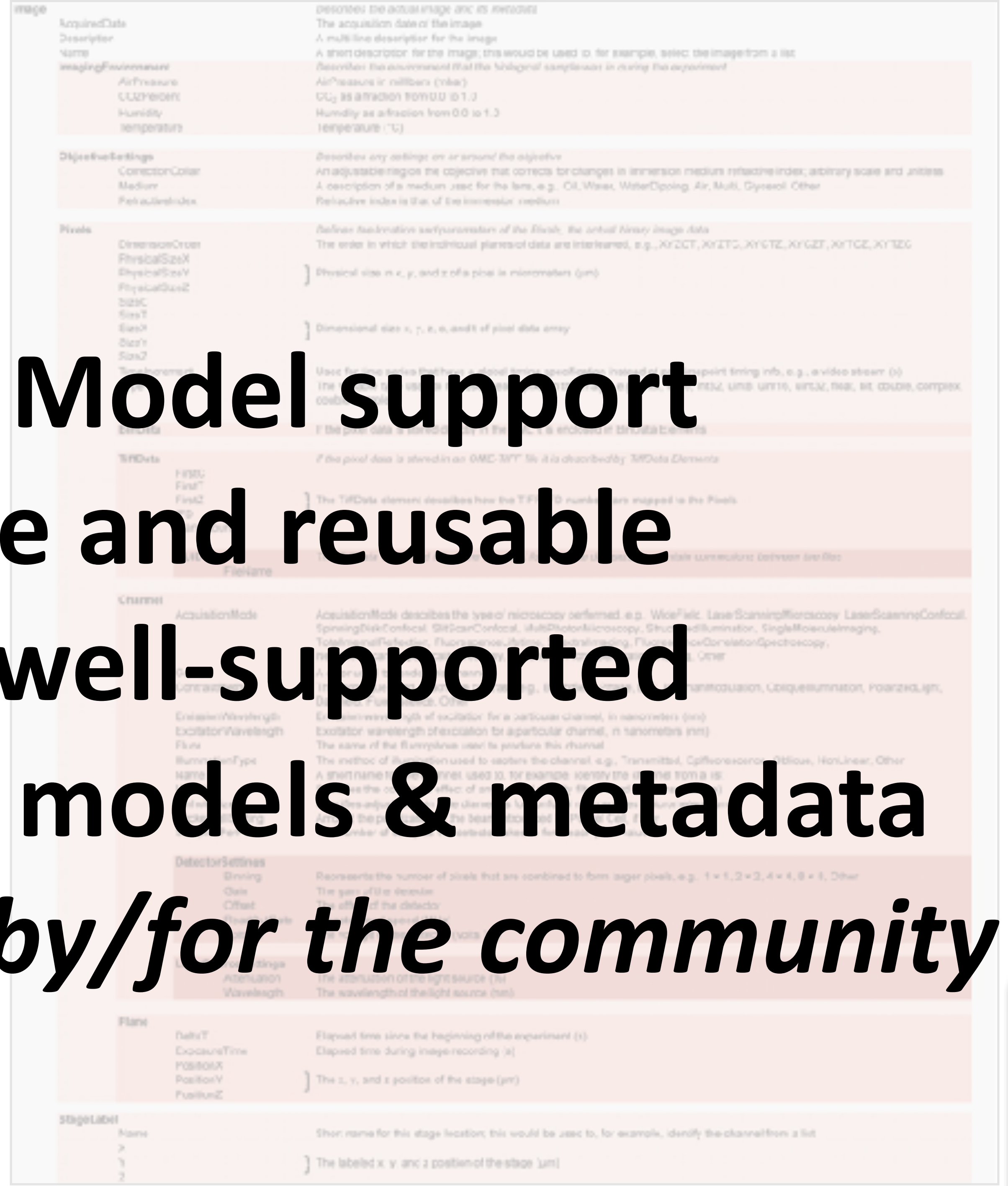
Metadata Matters 2024

- Superset of OME Model support
- Models extensible and reusable
- Base models are well-supported
- Vendors produce models & metadata
- *Web of schemas by/for the community*

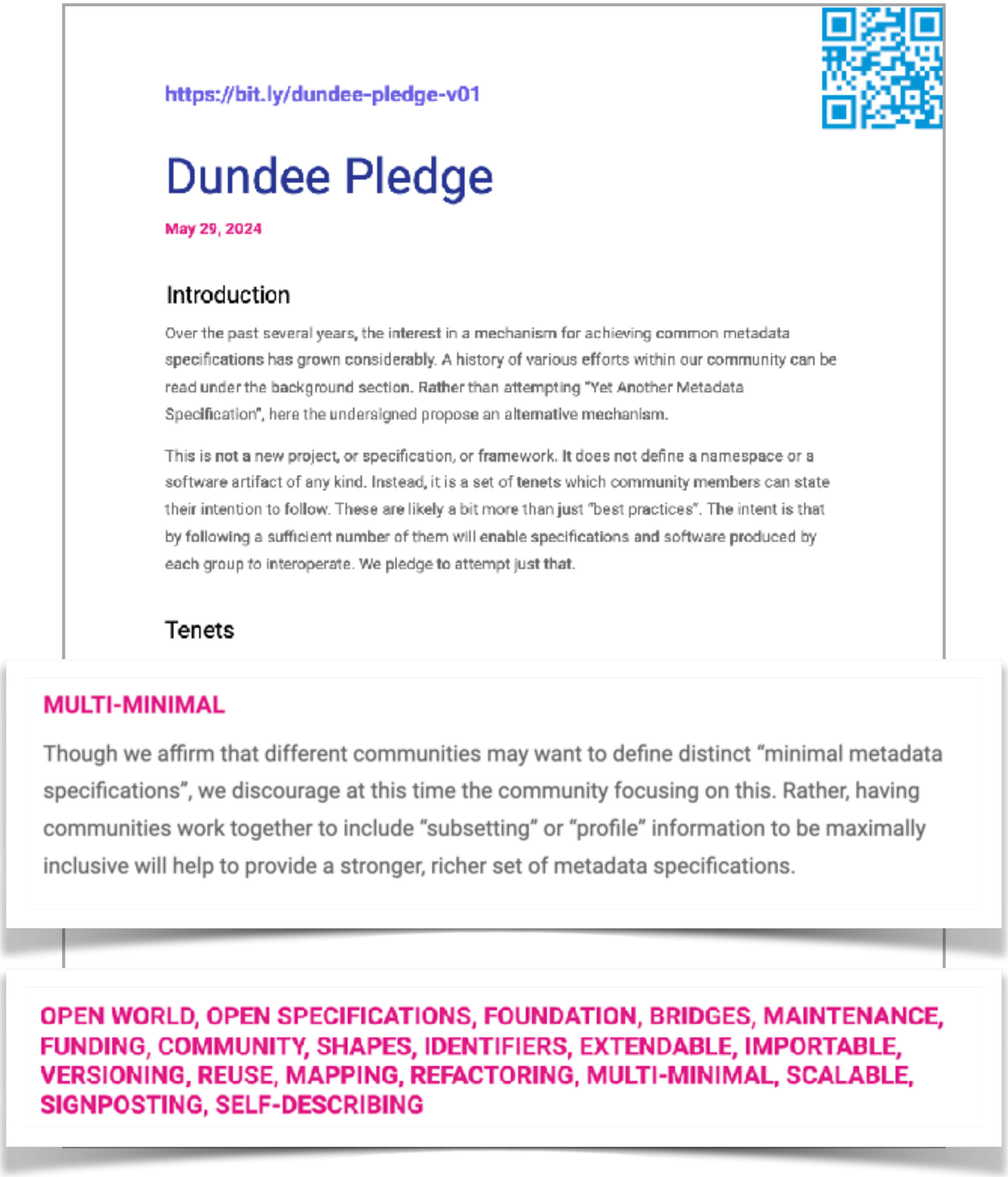
JCB: Comment

Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[6] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]



Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).



https://bit.ly/dundee-pledge-v01

Dundee Pledge

May 29, 2024

Introduction

Over the past several years, the interest in a mechanism for achieving common metadata specifications has grown considerably. A history of various efforts within our community can be read under the background section. Rather than attempting "Yet Another Metadata Specification", here the undersigned propose an alternative mechanism.

This is not a new project, or specification, or framework. It does not define a namespace or a software artifact of any kind. Instead, it is a set of tenets which community members can state their intention to follow. These are likely a bit more than just "best practices". The intent is that by following a sufficient number of them will enable specifications and software produced by each group to interoperate. We pledge to attempt just that.

Tenets

MULTI-MINIMAL

Though we affirm that different communities may want to define distinct "minimal metadata specifications", we discourage at this time the community focusing on this. Rather, having communities work together to include "subsetting" or "profile" information to be maximally inclusive will help to provide a stronger, richer set of metadata specifications.

OPEN WORLD, OPEN SPECIFICATIONS, FOUNDATION, BRIDGES, MAINTENANCE, FUNDING, COMMUNITY, SHAPES, IDENTIFIERS, EXTENDABLE, IMPORTABLE, VERSIONING, REUSE, MAPPING, REFACTORING, MULTI-MINIMAL, SCALABLE, SIGNPOSTING, SELF-DESCRIBING

https://bit.ly/dundee-pledge-v01

Acknowledgments

NFDI4BIOIMAGE

University of Dundee
(2005-2022)

wellcome

BBSRC

Chan Zuckerberg Initiative

EOSC-Life

CORBEL

GLOBAL BIOIMAGING growing collaboration

EURO-BIOIMAGING

German BioImaging
Gesellschaft für Mikroskopie und Bildanalyse

Chan Zuckerberg Initiative

leap

nfdi
Nationale Forschungsdaten Infrastruktur

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Funded by the European Union

SPRINGER LINK

Find a journal | Publish with us | Search

Home › Histochemistry and Cell Biology › Article

Original Paper | Open Access | Published: 10 July 2023

OME-Zarr: a cloud-optimized bioimaging file format with international community support

Josh Moore, Daniela Basurto-Lozada, Sébastien Besson, John Bogovic, Jordão Bragantini, Eva M. Brown, Jean-Marie Burel, Xavier Casas Moreno, Gustavo de Medeiros, Erin E. Diel, David Gault, Satrajit S. Ghosh, Ilan Gold, Yaroslav O. Halchenko, Matthew Hartley, Dave Horsfall, Mark S. Keller, Mark Kittisopikul, Gabor Kovacs, Aybüke Küpcü Yoldaş, Koji Kyoda, Albane le Tournoulx de la Villegeorges, Tong Li, Prisca Liberali, Dominik Lindner, Melissa Linkert, Joel Lüthi, Jeremy Maitin-Shepard, Trevor Manz, Luca Marconato, Matthew McCormick, Merlin Lange, Khaled Mohamed, William Moore, Nils Norlin, Wei Ouyang, Bugra Özdemir, Giovanni Palla, Constantin Pape, Lucas Pelkmans, Tobias Pietzsch, Stephan Preibisch, Martin Prete, Norman Rzepka, Sameeul Samee, Nicholas Schaub, Hythem Sidky, Ahmet Can Solak, David R. Stirling, Jonathan Striebel, Christian Tischer, Daniel Toloudis, Isaac Virshup, Petr Walczysko, Alan M. Watson, Erin Weisbart, Frances Wong, Kevin A. Yamauchi, Omer Bayraktar, Beth A. Cimini, Nils Gehlenborg, Muzlifah Haniffa, Nathan Hotaling, Shuichi Onami, Loic A. Royer, Stephan Saalfeld, Oliver Stegle, Fabian J. Theis & Jason R. Swedlow — Show fewer authors

Histochemistry and Cell Biology (2023) | Cite this article

224 Accesses | Metrics

BioImaging North America

EURO-BIOIMAGING

GLOBAL BIOIMAGING growing collaboration

QUAREP-LiMi

4D Nucleome

(many, many more)

& openmicroscopy.org/about/#consortium

Slides: https://downloads.openmicroscopy.org/presentations/2024/Dundee/Day1/

*Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – **NFDI 46/1 – 501864659***