Chan Zuckerberg Initiative | omega

Strambio-De-Castillia Lab @ Program in Molecular Medicine | University of Massachusetts Medical School | *caterina.strambio@umassmed.edu a*

Grunwald Lab @ RNA Therapeutics Institute | University of Massachusetts Medical School | *david.grunwald@umassmed.edu*

OME | NIH | 4D Nucleome

Metadata and Performance Tracking for Fluorescent Microscopes – Towards a Shared Microscopy Ontology

The 4DN-OME Ontology, an OME-OWL Extension with Emphasis on Usability, Minimum Information Guidelines and Quality Control for Super-Resolution Fluorescence Microscopy

Mathias Hammer[1], Norio Kobayashi[2], Josh Moore[3], **Alessandro Rigano**[4], **Farzin Farzam**[1], Shuichi Onami[5], Jason R. Swedlow[3], **David Grunwald**[1], **Caterina Strambio-De-Castillia**[4]

1 RNA Therapeutics Institute, UMass Medical School, Worcester MA 01605, USA
2 Head Office for Information Systems and Cybersecurity (ISC), RIKEN, Wako, Saitama 351-0198, Japan
3 Division of Computational Biology School of Life Sciences, University of Dundee, Dundee DND DD1 4HN, UK
4 Program in Molecular Medicine, UMass Medical School, Worcester MA 01605, USA
5 RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Hyogo 650-0047, Japan

Abstract: One of the central goals of the 4D Nucleome consists of mapping the localization of single genomic loci obtained by fluorescence microscopy onto global chromatin folding maps obtained by Chromatin Conformation Capture (CCC) experiments [1]. A significant challenge with the reproducibility of microscopy data and with its integration with multi-modality 'omics' data lies in the large variability of what is recorded by different microscopes, the absence of quality control standards and the lack of shared guidelines. The Open Microscopy Environment (OME) model [2], is a specification for sharing biological imaging data which stores metadata part as OME-XML [3]. The OME Consortium recently introduced the OME Core Ontology [4] as a basis to facilitate the introduction of domain-specific extensions. In synergy with this effort, the 4DN IWG has proposed an extension [5, 6] of the OME model, which is tailored at enhancing the comparability of single-molecule super-resolution fluorescence microscopy experiments. Here we present the current status of development of the 4DN-OME ontology prototype [7]. This semantic extension of the OME Core Ontology has the following features: 1) a tiered-system of reporting guidelines that scales required metadata content with experimental complexity. 2) A metadata model designed to better capture the technical complexity of high-resolution single-molecule localization and single-particle tracking experiments. 3) The introduction of standards for fluorescence microscope calibration and quantitative instrument performance assessment. Specifically, the 4DN-OME proposal extends the existing OME core-classes `Instrument' and `Image' to reflect the technological advances and the quality control requirements associated with single-molecule super-resolution microscopy. To this aim, the proposal puts forth several types of modifications, including the creation of additional classes and attributes to capture the complexity of microscope hardware commonly encountered in the field and their calibration requirements.

*This work is accompanied by other contributions describing the proposed **4DN standards for optical and performance calibration** and the development of a software tool termed **Micro-Meta App** that facilitates the collection of microscopy metadata.*

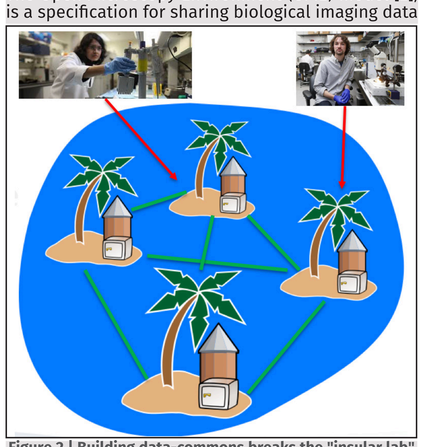


Figure 2 | Building data-commons breaks the "insular lab" pradigm, improves reproducibility and facilitate data re-use. Documentation and quality-control metadata standards, such as those proposed by the 4DN Imaging-Standards Working Group (IWG), together with data management infrastructure and integrated processing pipelines are essential pre-requisites to facilitate sharing imaging data and integration with **Chromatin Conformation Capture** (3C) results.

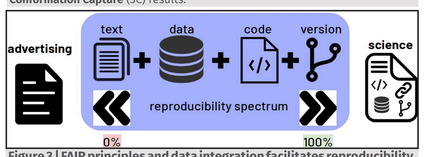
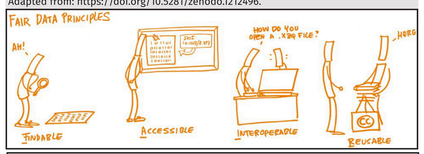
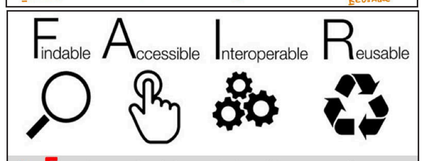


Figure 3 | FAIR principles and data integration facilitates reproducibility. The basic idea is that objects that belong together (e.g., an article with its associated code, data and workflows) should have some means of being aggregated, so that all associated research objects can be discovered together. Although this might seem to be obvious, as research objects are scattered across different repositories on the web, the connections between them are often lost.
Adapted from: https://doi.org/10.5281/zenodo.1212496.





- **F** – Searchable original image data
- **A** – Open, free and accessible retrieval procedures
- **I** – Community standards, shared vocabularies, shared analysis
- **R** – QC, rich metadata, data provenance

https://www.nature.com/articles/sdata201618

Figure 4 | FAIR data principles. FAIR principles describe characteristics that data should exhibit to assist discovery and reuse through the web. 1) FAIR does not imply Open.2) While the different facets of FAIR are related, they are in fact independent from each other allowing to define different degrees of FAIR-ness. 3) FAIR-ness can be achieved with a wide range of technologies and implementations). Adapted from: https://doi.org/10.5281/zenodo.1212496.

1- Introduction

Because the information content of image data is not machine-readable, microscopy images need to be accompanied by thorough documentation of the microscope hard-ware and imaging settings to ensure a correct interpretation of the results.

A significant challenge with the reproducibility of microscopy results and with their integration with chromatin folding maps generated by the 4DN consortium lies in the lack of shared super-resolution microscopy reporting guidelines and of instrument performance and calibration standards. The proposed 4DN-OME ontology is put forth as an extension of the OME core-ontology to help address this challenge.

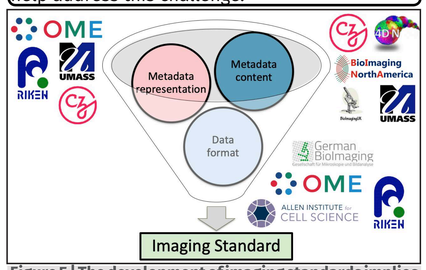


Figure 5 | The development of imaging standards implies the integration of three complementary elements. A significant global effort is under way to develop next-generation of imaging standards, which includes work conducted by members of the 4DN IWG in collaboration with the indicated organanizations and initiatives. Such efforts includes parallel work on three complementary fronts: 1) Development of next-generation data storage standards (Figure 6). 2) Improvement of **metadata representation** to facilitate standard flexibility and adaptation to technological advances (Figures 7-10). 3) Development of **shared community metadata and quality-control guidelines** (Figures 12-14).

2- OME core ontology

The **Open Microscopy Environment (OME)** is a consortium that produces format standards for microscopy data, and open-source tools for data management.

The OME model is a specification for sharing biological imaging data in two parts: metadata as **OME-XML** and binary data as **OME-TIFF**. OMERO is an image data repository that leverages the OME model to specify imaging metadata and is fast becoming a de facto-standard imaging tool.

An **Web Ontology Language (OWL)** ontology translation of the OME-XML specification called **OME core ontology** was built to catch up with the rapid progress and diversification of imaging technology and to facilitate the integrated analysis of various life science data [4]. The **OME core ontology** (Figure 7) is a translation of the OME-XML representation of the OME data model version 2016-06 that covers all its concepts and attributes.

In order to integrate imaging data produced using different imaging modalities, a **multi-modal ontology** for electron microscopy, X-ray computed tomography and magnetic resonance imaging was developed as an OWL-based extension of the OME core ontology.

Having confirmed the feasibility of this approach, the OME consortium has recently adopted the newly developed OWL-based OME model as an official companion to their XSD-based model.

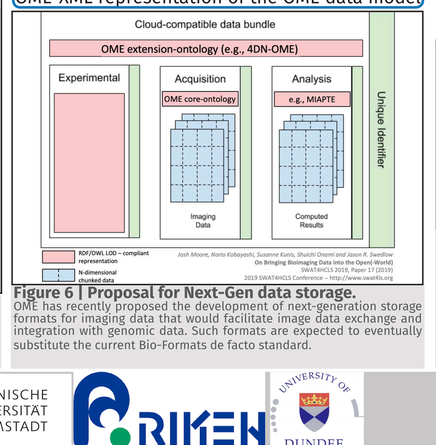


Figure 6 | Proposal for Next-Gen data storage. OME has recently proposed the development of next-generation storage formats for imaging data that would facilitate image data exchange and integration with genomic data. Such formats are expected to eventually substitute the current Bio-Formats de facto standard.

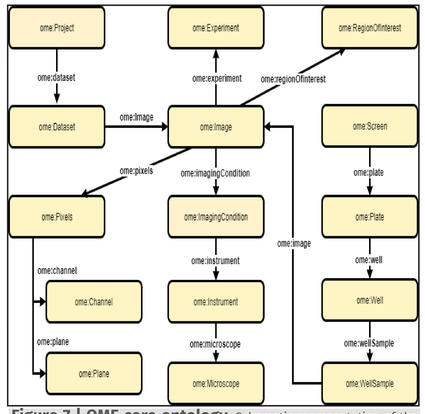


Figure 7 | OME core ontology. Schematic representation of the recently developed OWL/RDF representation of the OME core ontology. Adapted from Kobayashi, N. et al., OME Core Ontology: An OWL-based Life Science Imaging Data Model. **SWAT4HCLS-2019, Paper 29 (2019)**. 2019 SWAT4HCLS Conference. http://www.swat4ls.org

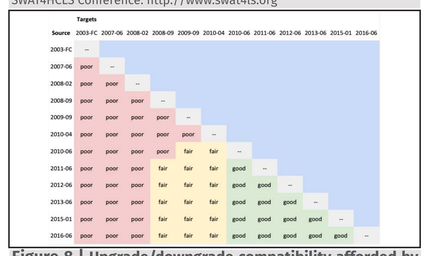


Figure 8 | Upgrade/downgrade compatibility afforded by the XSL transforms. The XML Schema Definition (XSD) representation of the OME model limits its ability to evolve in reponse to the development of new technology. While, upgrading any document is supported perfectly (blue). Downgrading documents becomes progressively harder and more lossy the older a target version is (green, yellow and red).

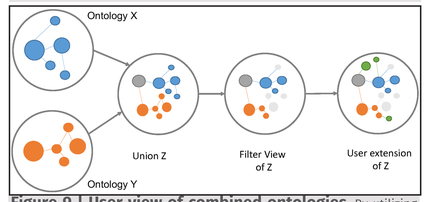


Figure 9 | User view of combined ontologies. By utilizing OWL/RDF technology it is possible to extend the OME XSD model locally with any objects and define one's own reference points to OME elements. In progress is an OWL and RDF-based update of the interface to make it easier for the user to use one or more ontologies. In particular, the user can restrict ontologies to the areas they require ("filter view") for better clarity. For more usability the user is given the opportunity to integrate her "own world" into existing ontologies.
Figures 6, 7, 8 and 9 are adapted from Moore et al., On Bringing Bioimaging Data into the Open(-World). **SWAT4HCLS-2019, Paper 17 (2019)**. 2019 SWAT4HCLS Conference. http://www.swat4ls.org.

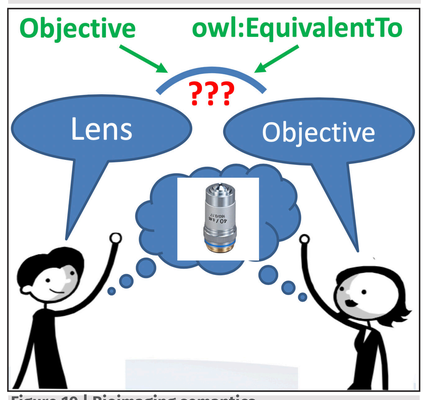


Figure 10 | Bioimaging semantics. Semantic modelling technologies developed for the modern World Wide Web, such as OWL and RDF, provide "translation" tools that can be used to improve bioimaging metadata flexibility and to produce future-proof imaging standards.

3- Bringing bioimaging data to the Open-World

In the absence of common, up-to-date metadata specifications, the publication of modern bioimaging data in a form that meets FAIR data principles is extremely challenging.

OME in collaboration with the 4DN IWG, **German BioImaging** and **RIKEN**, aims to address this fundamental block by providing tools to properly annotate and describe these new methodologies [4]. The idea is to bring semantic modelling technologies developed for the modern **World Wide Web** to bioimaging.

This will be achieved by expressing OME metadata as **Linked Open Data (LOD) in OWL/Resource Description Framework (RDF) triples**. In adherence with the open-world assumption (OWA), input from diverse sources can be combined to capture a complete record of modern imaging systems and analysis workflows, improving the ability of users to find and access data for re-analysis and integration (Figure 9).

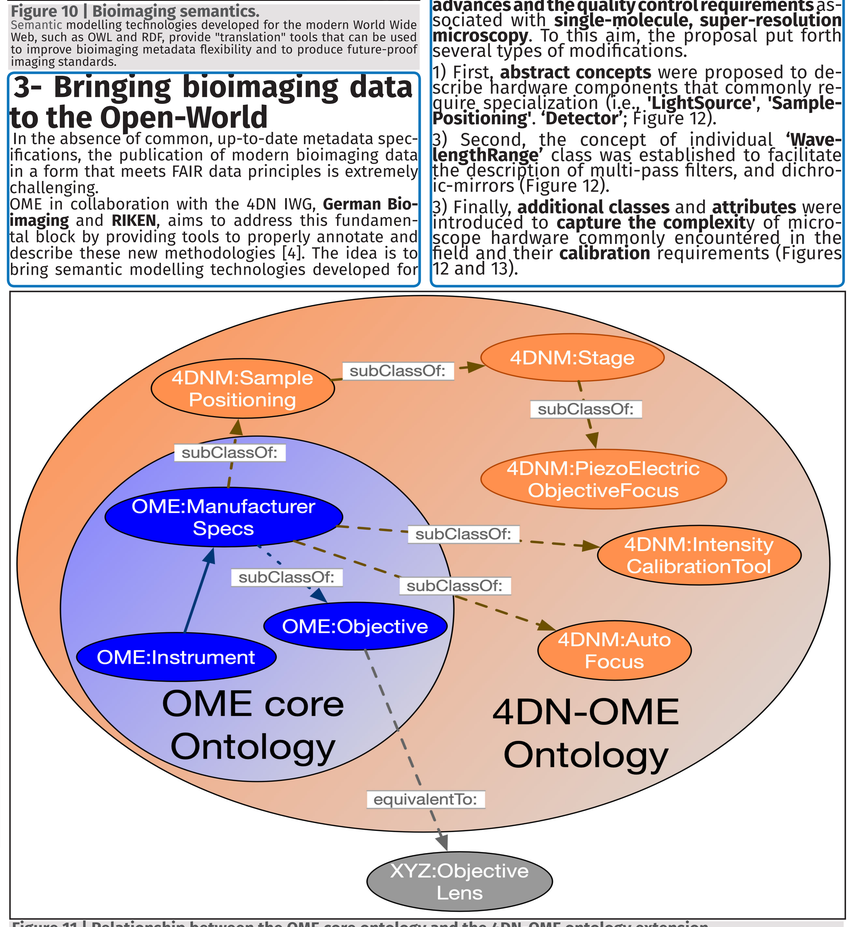


Figure 11 | Relationship between the OME core ontology and the 4DN-OME ontology extension. The OWL/RDF semantic modelling technology stack developed for the World Wide Web the allows the construction of flexible data models using the "core/extension" paradigm (Figure 9). In the specific 4DN-OME ontology extension example, novel model elements proposed by the 4DN IWG can be added to an extension package, which can be used or not depending on the needs of specific imaging communities.

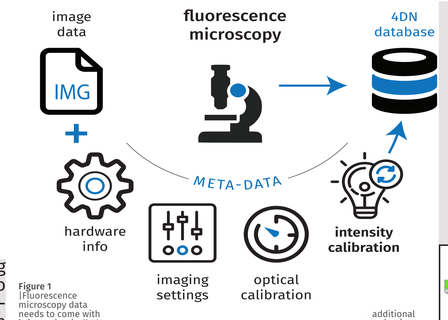
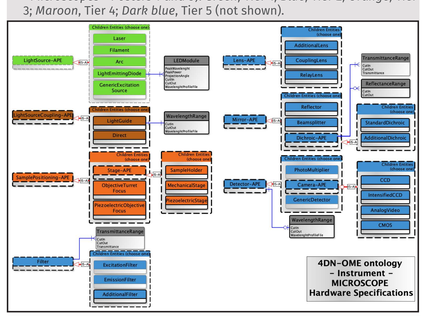


Figure 1 |Fluorescence microscopy data needs to come with information (called accompanying the image standardized in a tier-based system experiment and imaging technique. additional *meta-data*) data files, which is and depends on the type of

Figure 12 | Modifications of the OME core ontology structure proposed in the 4DN-OME ontology.
Mimicking the hierarchical structure of <LightSource>, several additional **Abstract Parents Elements (APE**; boxes with dashed lines) were proposed in the model to describe hardware components that commonly require specialization (i.e., <LightSourceCoupling>, <Filter>, <Mirror>, and <Detector> etc.). This streamlined the structure of the model and reduced data duplication.
The concept of individual **<Wavelength Range>** and **<LEDModule>** classes were established to facilitate the description of multi-pass excitation sources, filters and dichroic-mirrors.
Boxes with thick-black lines indicate **newly introduced classes** of the model. Color codes represent the proposed Tier level of each element (see, *Metadata and Performance Tracking for Fluorescent Microscopes - Posters 1 and 5*): *Green*, Tier 1; *Blue*, Tier 2; *Orange*, Tier 3; *Maroon*, Tier 4; *Dark blue*, Tier 5 (not shown).



4- The 4DN-OME ontology

The 4DN-OME ontology is being developed on the basis of the proposed 4DN exten-sion [5] of the OME xml model. This proposed ontology has the following key features:

1) A tiered-system of reporting guidelines that scales required metadata content with expermental complexity.

2) A metadata model designed to better capture the technical complexity of high-resolution single-molecule localization and single-particle tracking experiments.

3) The introduction of standards for fluorescence microscope calibration and quantitative instrument performance assessment.

In addition to introducing the concept of graded documentation requirements based on a tiered-system of guidelines, the 4DN-OME proposal extends the existing the OME core-classes **'Instrument'** and **'Image'** to reflect the **technological advances and the quality control requirements** associated with **single-molecule, super-resolution microscopy**. To this aim, the proposal put forth several types of modifications.

1) First, **abstract concepts** were proposed to describe hardware components that commonly require specialization (i.e., **'LightSource'**, **'Sample-Positioning'**. **'Detector'**; Figure 12).

3) Second, the concept of individual **'WavelengthRange'** class was established to facilitate the description of multi-pass filters, and dichroic-mirrors (Figure 12).

3) Finally, **additional classes** and **attributes** were introduced to **capture the complexity** of microscope hardware commonly encountered in the field and their **calibration** requirements (Figures 12 and 13).

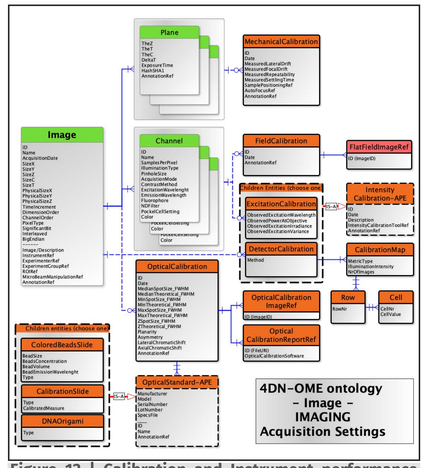


Figure 13 | Calibration and Instrument performance metadata model as proposed in the 4DN-OME ontology.
Additional classes and **attributes** were introduced in the 4DN-OME ontology extension in order to **capture the complexity** of microscope hardware commonly encountered in the field and their **calibration** requirements. Color codes and other annotations are as indicated in Figure 12.

BioImaging NorthAmerica
QUALITY CONTROL AND DATA MANAGEMENT WORKING GROUP

The BINA Quality Control and Data Management Working Group (QC-DM WG) was created to help drive the North American and international imaging communities' efforts to improve the reproducibility, comparability, downstream analysis and re-use of image data through rigorous record-keeping, quality control, and data management.

As such the QC-DM WG will concentrated its efforts in two directions:
1. Work in close connection with similar global efforts to generate shared guidelines and usable tools to facilitate the quantitative assessment and calibration of microscope performance, rigorous record-keeping of image acquisition conditions, and management of data in a manner that facilitates connecting resulting imaging dataset with metadata describing its "provenance".
2. Work in close connection with the BINA Communication WG to promote the dissemination of emerging best practices to the entire imaging community in North America and at large.

Caterina Strambio De Castillia, chair
UMass Medical School

Alison North, executive committee liaison
The Rockefeller University

Claire Brown
McGill University

Jay Copeland
Harvard Medical School

Alex Laude, advisor
Newcastle University

Pablo Ariel
University of North Carolina at Chapel Hill

Alex Corbett, advisor
University of Exeter

Peter Bajcsy
National Institute of Standards and Technology (NIST)

Glyn Nelson, advisor
Newcastle University

Judith Lacoste, corporate member
MIA Cellavie Inc.

Roland Nitschke, advisor
University of Freiburg

Figure 14 | Bio-Imaging North America (BINA) Quality Control and Data Management Working Group (QC-DM-WG). Dr. Strambio De Castillia was recently asked to chair the BINA QC-DM-WG, which was created to help drive the North American and international imaging communities' efforts to improve the reproducibility, comparability, downstream analysis and re-use of image data through rigorous record-keeping, quality control, and data management. As such the QC-DM WG will concentrated its efforts in two directions:

1) Work in close connection with similar global efforts to generate shared guidelines and usable tools to facilitate the quantitative assessment and calibration of microscope performance, rigorous record-keeping of image acquisition conditions, and management of data in a manner that facilitates connecting resulting imaging dataset with metadata describing its "provenance".

2) Work in close connection with the BINA Communication WG to promote the dissemination of emerging best practices to the entire imaging community in North America and at large.
Participation of Dr. Strambio De Castillia in this effort will ensure that the 4DN-OME ontology will undergo a thorough vetting process, which will include all majory community stake-holders (i.e. imaging scientists, vendors, funding agencies, regulatory agencies and journals) ensuring the creation of a community accepted microscopy metadata model.

[1] Dekker J, et al., The 4D nucleome project. Nature, 549:219–26 (2017)
[2] Goldberg IG, et al., The OME Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biol, 6, R47, (2005)
[3] Linkert M, et al., Metadata matters: access to metadata in the real world. JCB, 189, 777-782 (2010)
[4] https://gitlab.com/openmicroscopy/incubator/ome-owl/tree/master/ontology/owl/ome_core; Kobayashi N, et al. SWAT4HCLS 2019, Paper 29 (2019) and Moore J, et al. SWAT4HCLS 2019, Paper 17 (2019), 2019 SWAT4HCLS Conference. http://www.swat4ls.org
[5] Huisman M, et al. Minimum Information guidelines for fluorescence microscopy: increasing the value, quality, and fidelity of image data. arXiv:1910.11370 [q-bio.QM] (2019)
[6] https://github.com/WU-BIMAC/MicroscopyMetadata4DNGuidelines
[7] https://gitlab.com/openmicroscopy/incubator/ome-owl/tree/master/ontology/owl/4DNucleomeMicroscopyOntology

University of Massachusetts Medical School | TECHNISCHE UNIVERSITÄT DARMSTADT | RIKEN | UNIVERSITY OF DUNDEE | Program in Molecular Medicine | RNA Therapeutics Institute