Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model

Mathias Hammer[^,*,1,2], **Maximiliaan Huisman**[*,1], **Alessandro Rigano**[^,*,3], **Ulrike Boehm**[#,^,4], **James J. Chambers**[#,^,5], **Nathalie Gaudreault**[#,^,*,6], **Alison J. North**[#,^,7], **Jaime A. Pimentel**[#,^,8], **Damir Sudar**[#,^,9], **Peter Bajcsy**[#,^,10], **Claire M. Brown**[#,^,11], **Alexander D. Corbett**[#,^,12], **Orestis Faklaris**[#,^,13], **Judith Lacoste**[#,^,14], **Alex Laude**[#,^,15], **Glyn Nelson**[#,^,15], **Roland Nitschke**[#,^,16], **Farzin Farzam**[1], **Carlas S. Smith**[17], **David Grunwald**[^,*,1], **Caterina Strambio-De-Castillia**[#,^,*,3]

[1.]RNA Therapeutics Institute, UMass Chan Medical School, Worcester MA 01605, USA

[2.]Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany

[3.]Program in Molecular Medicine, UMass Chan Medical School, Worcester MA 01605, USA

[4.]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA

[5.]Institute for Applied Life Sciences, University of Massachusetts, Amherst, MA 01003, USA

[6.]Allen Institute for Cell Science, Seattle, WA 98109, USA

[7.]The Rockefeller University, New York, NY 10065, USA

[8.]Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, 62210, México

9.Quantitative Imaging Systems LLC, Portland, OR 97209, USA

10.National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

11.Advanced BioImaging Facility (ABIF), McGill University, Montreal, Quebec, H3G 0B1, Canada

12.Department of Physics and Astronomy, University of Exeter, Exeter, EX4 4QL, UK

13.MRI, BCM, Univ. Montpellier, CNRS, INSERM, Montpellier 34293, France

14.MIA Cellavie Inc., Montreal, Quebec, H1K 4G6, Canada

15.Bioimaging Unit, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

16.Life Imaging Center and BIOSS Centre for Biological Signaling Studies, Albert-Ludwigs-University Freiburg, Freiburg, 79104, Germany

17.Delft Center for Systems and Control & Department of Imaging Physics, Delft University of Technology, Delft, 2628 CN, The Netherlands

Abstract

Rigorous record-keeping and quality control are required to ensure the quality, reproducibility and value of imaging data. The 4DN Initiative and BINA here propose light Microscopy Metadata specifications that extend the OME data model, scale with experimental intent and complexity, and make it possible for scientists to create comprehensive records of imaging experiments.

Keywords

bioimage informatics; calibration; data formats; data model; data provenance; data standards; image quality; imaging; metadata; microscopy; open microscopy quality control; reproducibility; standards

1 - Summary

Digital light microscopy provides powerful tools for quantitatively probing the real-time dynamics of subcellular structures. Thorough documentation and quality assessment are required to ensure that imaging data may be properly interpreted (*quality*), reproduced (*reproducibility*), and used to extract reliable information and scientific knowledge which can be shared for further analysis (*value*). In the absence of community guidelines and tools, it is inherently difficult for manufacturers to incorporate standardized configuration information and performance metrics into image data and for scientists to produce comprehensive records of imaging experiments.

To solve this problem, the 4D Nucleome Initiative (**4DN**)[1] Imaging Standards Working Group (IWG), working in conjunction with the BioImaging North America (**BINA**) Quality Control and Data Management Working Group (QC-DM-WG)[2], here propose flexible light Microscopy Metadata specifications[3] that cover a spectrum of imaging modalities and scale with the complexity of the experimental design, instrumentation and analytical requirements. They consist of a set of three extensions of the Open Microscopy Environment (**OME**) Data Model[4,5], which forms the basis for the ubiquitous Bio-Formats library[5], and

because of their tiered nature they clearly specify which provenance[6] and quality control metadata should be recorded for a given experiment. This endeavor is closely aligned with the recently established QUAlity Assessment and REProducibility in Light Microscopy (QUAREP-LiMi) global community initiative[7,8]. As a result the ensuing 4D**N-B**INA-**O**ME (NBO) framework[3,9], alongside three interoperable metadata collection tools being developed in parallel (OMERO.mde, Micro-Meta App, MethodsJ2)[10–15], represents a major turning point towards increasing data fidelity, improving repeatability and reproducibility, easing future analysis, and facilitating the verifiable comparison of different datasets, experimental setups, and assays. The intention of this proposal is therefore to encourage participation, constructive feedback and contributions from the entire imaging community and all stakeholders, including research and imaging scientists, facility personnel, instrument manufacturers, software developers, standards organizations, scientific publishers, and funders.

2 - Introduction

The reproducibility crisis affecting the biological sciences is well-documented[16–20]. In the field of light microscopy, it can only be addressed if all published images are accompanied by complete descriptions of experimental procedures, biological samples, microscope hardware specifications, image acquisition settings, image analysis parameters and metrics detailing instrument performance and calibration[5,17,21,22]. This complete description, also known as ***Image Metadata***, consists of any and all information about an imaging experiment that ensures its rigorous interpretation, reproducibility and re-use, and should be recorded in scientific publications and alongside the actual image data in the file header or in supplementary files[23]. A fully developed metadata model would provide for consistent tracking of crucial information pertaining to the quality, reproducibility and scientific value of image data, and will allow the communication and comparison of such information in a Findable, Accessible, Interoperable, and Reproducible (FAIR) manner (see also Text Box I in: Huisman et al., 2021)[23,24]. However, as microscopy has evolved from a tool that generates purely descriptive or illustrative data to primary quantitative data acquired with ever more sophisticated and complex instruments, our practices to record this quantitative data and metadata faithfully and reproducibly have not kept up.

The OME consortium[25] has made significant advances with the development of the OME Data Model[4,5], which, together with the ubiquitous Bio-Formats image file format conversion library[5], serves as the only available *de facto* specification for accessing and exchanging image data. Nonetheless, the field of light microscopy still lacks much-needed community-mandated standards for imaging data and specifications for metadata (i.e., *Microscopy Image Data Standards*; Figure 1)[4,5] resulting in an unmanageable growth of proprietary and/or incompatible image file formats and metadata capture practices.

This manuscript is intended to launch a community-driven way forward to break the impasse. Specifically, it puts forth scalable specifications for light developed jointly by the 4DN[1] IWG and by the BINA QC-DM-WG[2]. In order to foster widespread adoption of the 4D**N-B**INA-**O**ME[3] framework (Figure 1A, *magenta bubble*), key components of this effort are: 1) user-friendly and when possible automated metadata-collection software

tools (OMERO-mde, MethodsJ2 and Micro-Meta App) that are presented in parallel manuscripts[10–15], and are coupled with standards for metadata representation and storage (Figure 1A, *yellow bubble*)[26–30]; and 2) sustainable roadmaps to switch from proprietary image data formats to common, cloud-ready OME Next-Generation File Formats (NGFF, Figure 1A, *blue bubble*)[31]. Importantly, all of these activities are carried out in the context of QUAREP-LiMi[7,8] and involve key members of the community, including microscope users, custodians, and manufacturers, imaging scientists, national and global bioimaging organizations, bio-image informaticians, standards organizations, funders and scientific publishers.

The proposed 4DN-BINA-OME specifications articulate along three mutually independent axes (Figure 1B):

1. Guideline Tiers - Metadata specification (Figure 3): A system of adaptable Tiers that spells-out which specific subset of metadata information should be included depending on experimental context and intent, technical complexity, and image analysis needs.

2. Core model and Extensions - Metadata extension (Figure 4): A suite of extensions that expand the **Core** of the OME Data Model[4,5] to comprehensively capture state-of-the-art transmitted light and widefield fluorescence microscopy (**Basic Extension**), and confocal and advanced fluorescence modalities (**Advanced and Confocal Extension**). Importantly to improve the management of quality control, a novel data model for capturing instrument calibration procedures (**Calibration and Performance Extension**) was developed in close collaboration with QUAREP-LiMi[7,8].

3. Metadata-requirement Levels - Metadata inclusion (Figure 5): Inherent flexibility in the inclusion of metadata is built in the model so that specific pieces of information will be considered as "required" (essential for rigor and reproducibility), and "recommended" (to improve image quality and to maximize scientific and sharing value).

While 4DN-BINA-OME is inherently adaptable, it provides all community stakeholders with clear and enforceable community-driven mandates for which information is required to ensure scientific rigor, experimental reproducibility, and maximal scientific value.

2.1 The metadata challenge in microscopy: the great variability of data formats and metadata reporting practices

The introduction of digital light detectors and computers has drastically improved the objectivity of optical observations and changed light microscopy in three profound ways. First, it has led to digital image formation, signal processing and computational methods that enable the extraction of quantitative information from images and have transformed light (and fluorescence) microscopy into a key quantification tool for biomedical research. Second, it has allowed the increasingly accurate recording of progressively lower amounts of light, enabling the visualization and quantitative measurement of sub-cellular and single-molecule (SM) events and molecular interactions with high specificity and temporal

resolution. Third, it has enabled imaging modalities, such as Confocal Laser Scanning Microscopy (CLSM), and super-resolution (SR) imaging techniques that allow high-resolution imaging of fixed and live samples in three dimensions.

Despite these advances and the employment of ever more sophisticated and complex instruments, practices to faithfully and reproducibly record quantitative image data and metadata have not kept up thus exacerbating existing challenges of quality control and reproducibility. The quality and scientific value of imaging data should be assessed not only based on the extent to which it can be used to answer the questions it was intended to address, but also trusted and re-used by others. It follows, that when performing imaging experiments, scientific rigor is inextricably tied to image quality, the reproducibility of experimental results, and the measure in which datasets can be integrated with other data and further analyzed to answer new questions.

Deriving valuable and rigorous information from images is completely dependent on the consistent recording and storage of information that captures the origin and subsequent processing of the data (i.e., 'data provenance')[6], as well as of metrics that quantitatively assess the quality of the microscope and of the images (i.e., 'quality control')[23]. A typical light microscopy experiment includes three (sometimes integrated) major steps centered around the production of image data (Figure 2): 1) **Sample Preparation**, i.e., all sample preparative steps for imaging. 2) **Image Data Acquisition**, i.e., light detection, image formation and recording; and 3) **Image Analysis**, i.e., the post-acquisition processing and quantification of images. Each procedure within these steps can add considerable variability to the final data. Thus, to document all possible sources of uncertainty, images need to be accompanied by **Image Metadata**[23] describing any and all information that allows the actual image data (i.e., quantitative values associated with the image pixels; Figure 2, *Pixel Image Data*) and imaging results to be evaluated, interpreted, reproduced, found, cited, compared and re-used as established by measurable data quality criteria (i.e., FAIR principles)[5,23,29]. As such, Image Metadata can be defined as metadata that documents all phases of a typical microscopy experiment (Figure 2) from: 1) experimental treatment, sample preparation and labelling (Figure 2A, *Experimental & Sample Metadata*)[32,33]; to 2) microscope hardware specifications, image acquisition settings, microscope performance metrics, and image-data structure (Figure 2A, *Microscopy Metadata*)[23]; to 3) details about any image analysis procedure employed to extract quantitative information from the images (Figure 2A, *Analysis Metadata*)[34–36]. As such, **Microscopy Metadata** consists of a subset of Image Metadata and, in turn, it can be subdivided into two sub-categories[23]: 1) **Microscopy data-Provenance Metadata (MPM)** describing the origin of the data microscope hardware specifications, image acquisition settings and image structure (Figure 2A, *Provenance*); and 2) **Microscopy Quality-control Metadata (MQM)** including calibration metrics that quantitatively assess the performance of the microscope (Figure 2A, *Quality Control*). In addition to capturing MPM and MQM (Figures 1 and 2), Microscopy Metadata standards should also address the following:

1. **Light microscopy utilizes a vast array of adaptable modalities**, each requiring different metadata to be reported as well as diverging quality control approaches.

2. A **microscope's theoretical performance and working conditions are difficult to assess** and are often unknown to the average user.

3. The **relevant hardware and software metadata can be difficult to retrieve** from available documentation.

4. The **paucity of automation and intuitive software tools** make record-keeping unduly burdensome, forcing experimental biologists to choose between scientific rigor and productivity.

5. The **variability of file formats** and the consequent need for raw data files to be converted into other formats prior to interpretation and comparison often yields a significant loss of metadata, or, worse still, inadvertently compromises the data during the conversion process.

Despite this apparent complexity, it is worth noting that the **Image Acquisition** step of an imaging experiment (Figure 2) is eminently manageable and quantifiable, as long as the microscope and imaging system are properly documented, maintained, and operated. Consequently, the development of community-sanctioned specifications for the compilation of Microscopy Metadata encompassing, MPM and MQM is not only essential for image data quality, reproducibility, and scientific and sharing value, but should be easy to obtain as described in more detail in an accompanying manuscript[23].

2.2 The importance and potential pitfalls of standardization

The value of Microscopy Image Data Standards (Figure 1) has been widely recognized, resulting in important efforts to establish best performance testing and instrument calibration practices[37–42], to unify data-submission requirements from journals[43–46] and to produce the exchange format between image data and metadata that forms the basis for this work[4,5,31,47,48].

Nonetheless, existing efforts have not yet reached normative value due primarily to the insufficiency of essential elements that are key components of this endeavour, including: 1) **Coordinated-community efforts** that lead to an easy-to-understand consensus on what specifications should be followed to ensure scientific rigor for imaging experiments[2,7,8]. 2) **Software tools** that make prescribed microscopy metadata models actionable by microscope manufacturers, custodians and users faced with the challenge of producing well-documented, high-quality, reproducible, and re-usable datasets, such as the ones being developed in complementary efforts (OMERO.mde, Micro-Meta App, MethodsJ2)[10–15]. 3) **Available endpoints** (i.e., deposition to image data repositories; data re-use pipelines) making the purpose and worth of good documentation clear to all members of the community[,48–51]. As a result, it remains challenging, for microscope manufacturers, custodians and users, to determine which parameters are relevant to a given technique and imaging experiment and best practice recommendations are often ignored due to their perception as too expensive, complicated and cumbersome.

Much would thus be gained from harmonizing the reporting standards in light microscopy. First, it would facilitate the documentation of any microscopy-based protocol, minimize error, and quantify residual uncertainty associated with each step of the procedure (Figure

2). This, in turn, would provide a wealth of valuable contextual information - collectively referred to as data provenance[6] – that would greatly increase the scientific and sharing value of the data. Such details would enable the reliable evaluation of scientific claims based on imaging data, facilitate comparisons within and between experiments, allow for reproducibility, and maximize the likelihood that data can be collated and analyzed by other scientists using current and future image processing and analysis methods. Furthermore, the increasing availability of public image repositories (e.g., Image Data Resource - IDR[48], Electron Microscopy Public Image Archive - EMPIAR[52], Bioimage Archive[49], the Cochin Image Database[53], the NIH Cell Image Library[54], the RIKEN Systems Science of Biological Dynamics database - SSBD[55]), will undoubtedly increase the need for community-wide documentation and quality control standards, which can adapt to new technologies. As a first step in this direction[51] the Recommended Metadata for Biological Images (REMBI)[50] guidelines were recently proposed that would maximize the possibility of making bioimaging datasets available to other researchers in a timely manner, consistent with the FAIR principles[24], and thus amenable for reuse.

Despite offering innumerable advantages, standardization also has its pitfalls. First, in the absence of software tools, it can significantly increase the administrative burden associated with imaging experiments. Second, because it is impossible to know *a priori* the complexity and diversity inherent to experimental details and imaging modalities that are yet to be developed, a lack of flexibility can severely limit the type of data that can be stored. It follows that it is critical that any proposed set of sustainable community specifications meet strict expandability requirements. Because of its inherent extensibility and the solid plans for modernization (see Text Box I), the OME Data Model[4,5] provides a robust foundation for Microscopy Metadata (Figure 2B) that can be extended by introducing information that is not yet covered (including experimental specific metadata, modality specific metadata, quality control metadata and analysis-specific metadata). As these extensions[28–30,34,36] become more commonly used, they can be incorporated into the core model through community announcements and related vetting processes to ensure they meet expanding community needs.

3 - A three-dimensional matrix of 4DN-BINA-OME microscopy metadata specifications

Since a one-size-fits-all solution for Microscopy Metadata requirements is clearly not tenable, here we propose the 4DN-BINA-OME framework in which microscopy documentation and quality control requirements are organized along three *orthogonal axes* that are largely independent from each other (Figure 1B). The first axis is based on the observation that different types of experiments have different reporting and quality control requirements based on technical complexity, experimental design, and image analysis needs. Hence, requirements along this axis are subdivided into Tiers depending on the three criteria listed above (Figure 1B, *Guideline Tiers*; Figure 3, Table I and Supplementary Table I). The second axis starts with and extends the OME Data Model[4,5] with additional metadata components that are introduced based on the microscopic modality (e.g., epifluorescence vs. confocal microscopy) and accommodates expansion as new technologies are developed

that are covered neither by the core model nor by the currently proposed extensions (Figure 1B, *OME Core vs. Extensions*; Figure 4). Finally, the third axis grades documentation requirements based on whether each piece of information is essential for rigor and reproducibility (*Must use*) or recommended to improve image quality and for maximizing scientific and sharing value (*Should use*; Figure 1B, *Metadata-requirement Levels*; Figure 5). The existence of these three axes will allow institutions, funding agencies, consortia, and scientific publishers to define best practices for light microscopy experiment documentation while concomitantly allowing individual scientists to find an appropriate position on the guideline matrix that both matches their needs and remains compatible with community-mandated guidelines. As an example, Table II lists where representative experiments would fall within the Microscopy Metadata guideline matrix (Figure 1B). It should be noted that the 4DN-BINA-OME and REMBI[50] metadata frameworks were developed in parallel and were deliberately designed to directly map with each other. Specifically, with the proviso that REMBI also defines metadata for Electron Microscopy and Correlative Imaging, regarding Light Microscopy the following correspondences exist between REMBI and 4DN-BINA-OME:

1. The REMBI "Instrument attributes" element maps with the `<Instrument>` core element of 4DN-BINA-OME (which captures Microscope Hardware Specifications metadata).

2. The REMBI "Image acquisition parameters" element maps with the `<Image>` core element of 4DN-BINA-OME, which captures Image Acquisition Settings metadata.

Because of this deliberate direct mapping, Microscopy Metadata specified by 4DN-BINA-OME intrinsically meets and exceeds the requirements imposed by REMBI for Light Microscopy. Hence, the adoption of 4DN-BINA-OME (especially through the use of the complementary software tools being simultaneously presented in related manuscripts)[10–15] would greatly facilitate the work of microscopists wanting to deposit imaging data on BioImage Archive[49].

3.1 The first axis: a tier-based system of guidelines for light Microscopy Metadata

To achieve rigor and reproducibility, increasingly elaborate imaging experiments require additional metadata on top of those required for more basic experiments. On this account, a graded system for metadata requirements is not only appropriate, but it also minimizes the burden of collecting metadata for each experiment whilst maximizing the opportunities for rigor, reproducibility, evaluation, analysis, and comparison. We envision a flexible system in which different imaging communities (i.e., individual research institutions, individual fields of knowledge or research consortia) would define their own sets of criteria whereby microscope hardware and imaging experiments are classified in Tiers based on experimental and image complexity, microscope technology and imaging modality, and analytical requirements. Hence the tiered system of guidelines presented here (Figure 3, Table I, Supplementary Table I, and Supplementary Material)[56], should be considered as an example of how different imaging experiment types could be placed on a complexity scale to facilitate the collection of the most appropriate minimum set of metadata required

for reproducibility and comparison of each category. We expect that this system will evolve organically to incorporate new imaging modalities. Active international initiatives such as QUAREP-LiMi[7,8] should help to ensure that new metadata specifications are agreed upon by the community and consistent with existing standards.

A robust, maximally useful, and efficient metadata standard would be tailored around the different reporting requirements of experiments of increasing complexity. We suggest here a system composed of one Descriptive (Tier 1) and two Analytical (Tiers 2 and 3) tiers (Figure 3; Table I and Supplementary Table I)[56], in which imaging instrumentation and datasets are classified based on the following sets of criteria:

1. Are results amenable to visual interpretation or is advanced image analysis (e.g., sub-pixel SM localization microscopy - SMLM) required for the full understanding of results?

2. Are biological samples fixed or alive during acquisition?

3. Are any parts of the quantitative microscopy pipeline (microscope instrument, acquisition modality and image analysis) relying on novel, rather than fully established technology?

4. Is the data provenance and quality control metadata tracked, documented, and reported by hardware manufacturers or instrument developers?

Consistent with minimum information principles, the system represents a minimal set of metadata required for each tier, covering only the information relevant for the interpretation of the specific imaging experiment (while more comprehensive information is always allowed and encouraged). As an example, the proposed specifications encompass information about the sample that directly impacts the imaging conditions (e.g., labeling method, mounting medium). However, due to the complexity of fully describing experimental and sample preparation procedures, such endeavour pertains more directly to the communities involved in the different research areas that utilize microscopy as an investigation method (i.e., cell biology, developmental biology, etc.,) and are clearly beyond the scope of this effort. While the initial impetus for developing such specifications will have to originate within individual research fields, coordination across domains will be necessary to develop consensus around overlapping areas and avoid splintering off in discordant directions. A detailed description of the 4DN-BINA-OME Tier-system[56] is available in Supplementary Material.

3.2 The second axis: a system of 4DN-BINA-sponsored community-driven OME extensions

In its simplest form, metadata can be easily represented as lists of key-value pairs, where the first term is a descriptive term for a specific attribute and the second term is the value of the attribute, including units for numerical values. However, lists of key-value pairs are often not sufficient to define rich metadata guidelines as they do not allow to capture the often-complex relationships between different real-world components and situations. A better method is the development of abstract models for the data that represent the scenario to be described. Ideally, such a data model would account for the components

of the system, the attributes that need to be recorded for each component to be fully documented, and the relationship between components (Figure 6A). A useful formalism for developing, describing, and viewing an appropriate data model is the Entity-Relationship (ER) diagram[57], which subsequently has to be translated into formalized schemas and file formats (Figure 6B) to facilitate implementing metadata capture and management tools.

The 4DN-BINA proposal: a suite of three extensions of the core OME ontology
—Due to its status as the only existing exchange format for imaging experiments, the robustness of its design, and the solid path forward toward modernization (Text Box I; Figure 2B; details in Supplementary Material), the OME Data Model (i.e., OME Core) represents the ideal starting point for the suite of 4DN-BINA extensions presented here (Figure 4). As such, the 4DN-BINA-OME specifications proposal consists of three extensions of the OME Core[4,5] each of which incorporates the concept of graded documentation requirements based on a tiered-system of guidelines (Figure 3; Table I and Supplementary Table I). A a first step towards this goal, the 4DN-BINA-OME Microscopy Metadata Specifications[9,58,59] extend the core OME elements <Instrument> and <Image> (Figure 2B; details in Supplementary Material) to reflect the technological advances and the Quality Control requirements associated with state-of-the-art transmitted light, widefield- and confocal-fluorescence microscopy. A detailed description of the system of three proposed 4DN-BINA-OME extensions is available in Supplementary Material. In summary:

1. The **Basic extension** is designed to better capture the technical complexity of transmitted light microscopy and wide-field fluorescence, including sub-pixel single particle localization and SPT experiments (Figure 4, blue and grey elements).

2. The **Advanced and Confocal extension** is designed to better capture experiments requiring tunable optics and confocal microscopy (Figure 4, green elements).

3. The **Calibration and Performance extension** introduces specifications for the capture of metrics required for microscope calibration and quantitative instrument performance assessment (Figure 4, maroon elements).

While it would be impracticable for the current version of the specifications to meet all emerging community needs, the proposed structure provides a flexible framework to easily accommodate future extensions that will be need to be developed in close collaboration with the community to capture sources of image data that our model does not yet fully define (e.g., light-sheet and airy scan confocal microscopy.

In order to facilitate understanding of the 4DN-BINA-OME by all relevant members of the community regardless of their information science expertise, while at the same time ensuring machine readability, formal representations of the 4DN-BINA-OME extensions are maintained on GitHub[9] in three formats (see also Data and Code Availability Statement): 1) a set of graphical ER schemas is used to facilitate an overall understanding of the model structure[59]; 2) an excel spreadsheet to express the details of the model in a human-readable form[59]; finally 3) XML Schema Definition (XSD) to represent the model schema in a machine-readable manner[58].

3.3   The third axis: metadata requirement levels

Along the third axis (Figure 5), individual metadata fields in these specifications are classified based on requirement level as described by the Request for Comment (RFC) document 2119[60]. The keyword MUST, or the terms "REQUIRED" or "SHALL," mean that the definition is an absolute requirement to validate experimental claims and ensure reproducibility. The keyword SHOULD, or the adjective "RECOMMENDED," means that while there may exist valid reasons in particular circumstances to ignore a particular field, they are highly recommended to maximize *Image Quality, scientific value and* FAIRness[24]. Two examples of the use of the third dimension to add flexibility to the proposed 4DN-BINA-OME Microscopy Metadata specifications are presented below:

Example 1) OME Core and 4DN-BINA Basic extension element `<Objective>` (Figure 5).—While the Manufacturer, Model, Magnification and Numerical Aperture (*LensNA*) of an objective are required to be able to interpret microscopy results and for reproducibility, other attributes such as a hardware component's Lot Number, a Lens's Back Focal Length and the Calibrated Magnification of an Objective are recommended to maximize Image Quality and scientific value, but they are not required because they are not essential for reproducing the experiment.

Example 2) 4DN- BINA Calibration and Performance extension element `<MultiColorBeads>`.—When using multicolored beads to prepare a colored-beads slide to use for the optical calibration of a microscope, the Manufacturer, Catalog Number, and Concentration of the beads preparation alongside the Diameter of the beads are essential for the interpretation of the calibration results and for reproducibility. However, the bead's Type and Material may be omitted because it can be argued that while that information improves the completeness of the data, they are not absolutely required for the correct interpretation of the results of the Optical calibration procedure in which the beads are utilized.

3.4   Model implementation: Material and Methods recommendations

A recent exploration about the quality of published Method sections in scientific articles containing images obtained with advanced microscopes, found that the quality of reporting was poor, with some articles containing no information about how images were obtained, and many articles lacking important basic details[17]. Nonetheless, there is ample evidence that the publication of full details about how each image was obtained is vital for rigor, reproducibility and maximal scientific and sharing value[18–20,44–46]. In this context, the 4DN-BINA-OME Microscopy Metadata specifications presented are intended to provide a major contribution towards the development of community-driven criteria for which information should be included in the Methods sections of scientific publications.

As a first step, in close agreement with the proposal presented in parallel efforts[46,23], we propose that Microscopy Metadata appropriate for Tier 1 (both MUST and SHOULD fields) should always be included in the Material and Methods section of any journal publication to meet minimal rigor and reproducibility criteria[17]. As such, the generalized and automated availability of Tier 1 metadata could save considerable effort both for authors, who would not need to search for information scattered across different data-files, hardware setups and

lab notebooks in preparation for publication, and for readers, who would not need to search the various sections of publications for information that may or may not have been included.

3.5 Model implementation: Facilitated metadata collection

The importance of rich metadata to ensure the quality, reproducibility, as well as scientific and sharing value of image data cannot be overstated. However, the collection of rich sets of microscopy metadata is time-consuming and, in the absence of active participation from hardware manufacturers, imposes an unfair burden on experimental scientists and is therefore difficult to enforce. Appropriate community-validated software tools and data management practices are essential to streamline and automate the documentation of microscopy experiments. In this context, in parallel with this proposal for Microscopy Metadata guidelines a suite of three complementary and interoperable software tools are being developed and are presented in related manuscripts. 1) OMERO.mde[10,11] focuses on facilitating the consistent handling of image metadata ahead of data publication and deposition based on shared community Microscopy Metadata specifications and according to the FAIR principles. In addition, OMERO.mde promotes the early development of Image Metadata extension specifications to allow testing and validation before incorporation in community-accepted standards. 2) Micro-Meta App[12,13] focuses on an easy-to-use, Graphical User Interface (GUI)-based platform that interactively guides users through the process of building tier-based records of microscope hardware, accessories and image acquisition settings containing all relevant Microscopy Metadata as sanctioned by the community specifications such as the ones described here. Because of its graphical nature, Micro-Meta App is particularly suited for imaging scientists to enter all microscope metadata and use the tool for teaching trainees about Microscopy, the importance of Microscopy Metadata and training microscope-users in imaging facilities. 3) Finally, MethodsJ2[14,15] focuses on automating the process of writing Microscopy Metadata guidelines-compliant Methods and Acknowledgement sections for scientific publications utilizing microscopy experiments. MethodsJ2, by design, operates in concert to automatically import Microscopy Metadata using BioFormats and from the Micro-Meta App[12,13].

3.6 Model implementation: Information required for basic image interpretability

To ensure the basic interpretability of image data acquired before the adoption of community-sanctioned guidelines, any data that might be shared or published should, at the very least, contain the required metadata fields stipulated by the intersection between Tier 1 and the Core of the OME Data Model. Thus, Tier 1/Core sanctions the baseline metadata requirements for any light microscopy experiment to be interpretable, utilized and shared for scientific purposes. Specifically, this includes minimal microscope Hardware Specifications (i.e., microscope, light source and objective manufacturer information and essential description), and essential information about the Image structure (i.e., number of planes, channels and time-points, pixel size, fluorophore name, emission, and excitation wavelength, etc.).

4 -　Conclusions

Light microscopy images need to be accompanied by thorough documentation of the microscope hardware and imaging settings to ensure a correct interpretation of the results. A significant challenge with the reproducibility of microscopy results and their integration with other data types, such as chromatin folding maps generated by the 4DN consortium[1], lies in the lack of standardized reporting guidelines for microscopy experiments as well as instrument performance and calibration standards[17–19,45]. Despite a growing consensus that such standards for light microscopy are desirable, previous efforts to develop shared microscopy data models and application programming interfaces[4,5,47] have not yet succeeded in the establishment of a universal set of norms. In this manuscript, a framework to extend the OME Data Model is put forth to help address this challenge. In addition to aligning the OME Data Model to current technological developments, the specifications advanced here focus on the maximization of usability via the introduction of a tiered system of documentation requirements, on an expandable suite of model extensions, including the first available data model for quality-control metadata for light microscopy imaging and flexible use of required, and recommended, fields.

Microscopy is not the only field in which recent technological advances have resulted in increasingly rich datasets. Recent examples are genomic DNA and transcriptomics RNA sequencing, which are, in fact, much younger fields than microscopy. While protocols varied substantially in their early days (the original images from the sequencer were kept with the determined sequence), it took only about a decade to establish metadata requirements. One factor that helped establish such metadata criteria was the NIH Encyclopedia of DNA Elements (ENCODE) consortium[61]. The development of Standard Operating Procedures (SOPs) and shared benchmarks (i.e., gold-standards) within this group was pivotal for the establishment of agreeable standards for practical day-to-day use. In the interest of scientific progress and making data FAIR, data and metadata standards should not be dictated by individual laboratories or microscope manufacturers. Instead, they should emerge organically from discussions involving all members of the community who can benefit from standardization and be subjected to evaluation before adoption.

In this spirit, the initial draft Microscopy Metadata specifications put forth by the 4DN[1] IWG were evaluated and revised by the BINA QC-DM-WG[2], resulting in the current proposal. Because it is inherently impossible to predict all future changes the light microscopy field might undergo and in order to ensure rigor and reproducibility for image data now and in the future, it is clear that more work is needed to ensure that the 4DN-BINA (as well as future) extensions of the OME Data Model for bioimaging metadata proposed here continue to evolve as a result of regular exchange of information and views across the community. This is required to capture any future technical development in a manner consistent with current specifications while supporting FAIR data principles[10,24]. This is particularly important in the face of the establishment of a growing number of public image data resources[51] such as the European IDR[48], EMPIAR[52], Movincell[62] and Bioimage Archive[49]; the Japanese SSBD hosted by RIKEN[55]; and, in the US the Allen Cell Explorer[63], the Human Cell Atlas[64], and the NIH-funded Cell Image Library[54], Human BioMolecular Atlas Program (HuBMAP)[65], and BRAIN initiative's imaging resources[66].

These resources offer the opportunity to emulate for light microscopy the successful path that has led to community standards in the field of genomics[33,67–69].

Because of the community-nature of this effort, the 4DN-BINA-OME specifications must evolve first and foremost in alliance with the QUAREP-LiMi initiative[7,8] to ensure that all participating imaging community stakeholders, importantly including microscope and software tool manufacturers, who are ultimately responsible for providing the information to be recorded in microscopy metadata, are involved from the ground up and provide timely feedback. In addition, the further development of the Microscopy Metadata Specifications is being coordinated with other parallel initiatives including:

1.   The development of strategies and pipelines to integrate images and their metadata with -omics data from the same experiment, such as is underway as part of 4DN.

2.   The OME community development of general criteria and procedures to capture and store metadata in OME-NGFF (Text Box I). The OME NGFF effort[31] is implementing storage approaches to hold the binary pixel data and the metadata described herein in standardized, shareable, long-lived, efficient, and performant containers (e.g. files).

3.   The EMBL-EBI development of the REMBI recommendations for metadata to be included with imaging datasets deposited to BioImage Archive[49,50].

4.   The development of the International Standards Organization (ISO) 23494-1 standard that will include the 4D**N**-**B**INA-**O**ME (NBO namespace) Microscopy Metadata specifications as part of a Provenance information model for biological material and data[26,27].

5.   The development of online educational material, workshops and in-person courses in the context of BINA and in collaboration with Global Bioimaging[70] and other community partners[71,72]

The specific purpose of this multi-pronged community effort will be to: 1) Educate microscope manufacturers, custodians, and users about the importance of metadata standards and documentation to ensure image data quality, reproducibility and re-use value. 2) Increase awareness about the 4DN-BINA-OME Microscopy Metadata specifications proposed here and the complementary software tools for implementation developed in parallel efforts)[10–15]. 3) Engage all major stakeholders (including commercial, government and academic worlds) in our effort towards community-driven metadata standards for light microscopy. Initially, this mechanism is going to be employed to generate a wider consensus around the current framework and lead towards the development of true community standards. A similar mechanism will be employed to engage representatives of different domains to generate Microscopy Metadata extensions and tier-systems that best suit their research areas and avoid splintering off in multiple incompatible directions. As an example, more extensions will have to be defined to capture sources of image data that our model does not fully define, both in experimental (e.g., light-sheet and airy scan confocal microscopy) and synthetic image frameworks (e.g., predictive multi-channel image synthesis, and super resolution level image restoration).

In conclusion, we are confident that because of its strong roots in the community, and because it is closely linked with the parallel development of easy-to-use interactive tools to facilitate metadata collection)[10–15], the flexible model framework presented here will provide a significant step forward towards the establishment of robust and future-proof light microscopy metadata standards. With the key partnership and increasing support from institutions and funding agencies, this work will continue to expand and help increase rigor and reproducibility in imaging data, rewarding everyone involved with improved trust in published results.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data and Code Availability Statement

This manuscript describes the community-driven 4DN-BINA-OME Microscopy Metadata specifications. The intention of this proposal is therefore to encourage participation, constructive feedback and contributions from the entire imaging community and all stakeholders, including research and imaging scientists, facility personnel, instrument manufacturers, software developers, standards organizations, scientific publishers, and funders. To achieve this goal documents describing the 4DN-BINA-OME Microscopy Metadata specifications are publicly available as follows:

1. *Project name:* 4DN-BINA-OME (NBO) Microscopy Metadata Specifications

2. *Project main page*: https://github.com/WU-BIMAC/ NBOMicroscopyMetadataSpecs

3. *Metadata Model schema documents:*

    a. GitHub repository: https://github.com/WU-BIMAC/ NBOMicroscopyMetadataSpecs/tree/master/Model (DOI: https:// zenodo.org/record/4710731)

    b. Excel spreadsheet:

        i. **4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XLS Spreadsheet.** *Available at:* https://zenodo.org/record/4711229

    c. Entity Relationship graphical representation:

        i. **4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - Entity Relationship schemas.** *Available at:* https://zenodo.org/record/4711229

    d. XSD Schema:

        i. **4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XSD schema.** *Available at:* https:// doi.org/10.5281/zenodo.4711426

4. *Tiers-system documents:*

    a. *a* GitHub repository: https://github.com/WU-BIMAC/ NBOMicroscopyMetadataSpecs/tree/master/Tier%20System

5. *Provide feedback about the model:*

    a. Model issues page → https://github.com/WU-BIMAC/ NBOMicroscopyMetadataSpecs/issues

    b. Video tutorial → If you intend to provide feedback about the 4DN-BINA-OME Microscopy Metadata Specifications please start from this series of 4 video tutorials

        i. Video 1 of 4 → https://vimeo.com/624971871

          ii.   Video 2 of 4 → https://vimeo.com/624971915

          iii.   Video 3 of 4 → https://vimeo.com/624971980

          iv.   Video 4 of 4 → https://vimeo.com/624995861

6. *Schema language:* XML Schema Definition (XSD), Entity-Relationship schema

7. License: GNU LESSER GENERAL PUBLIC LICENSE, Version 3

Abbreviation list

BINA	BioImaging North America
4DN	4D Nucleome
FAIR	Findable Accessible Interoperable and Reproducible
OME	Open Microscopy Environment
QUAREP-LiMi	QUality Assessment and REProducibility for Instrument and Images in light-microscopy

References

1. Dekker J et al. The 4D nucleome project. Nature 549, 219–226 (2017). [PubMed: 28905911]

2. Strambio-De-Castillia C et al. Quality Control and Data Management | Bioimaging North America (BINA). Bioimaging North America https://www.bioimagingna.org/qc-dm-wg (2019).

3. Hammer M et al. Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model - preprint. BioRxiv (2021) doi:10.1101/2021.04.25.441198v1.

4. Goldberg IG 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). [PubMed: 15892875]

5. Linkert M et al. Metadata matters: access to image data in the real world. J. Cell Biol 189, 777–782 (2010). [PubMed: 20513764]

6. Ram S & Liu J A Semantic Foundation for Provenance Management. J. Data Semant 1, 11–17 (2012).

7. Boehm U et al. QUAREP-LiMi: A community endeavour to advance Quality Assessment and Reproducibility in Light Microscopy. Nature methods vol. ePub (online ahead of print) (2021).

8. Nelson G et al. QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy. J Microsc (10.1111/jmi.13041) ePub (online ahead of print), 1–18 (2021).

9. Rigano A et al. 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01. (https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs, 2021). doi:10.5281/zenodo.4710731.

10. Kunis S et al. MDEmic in a use case for microscopy metadata harmonization: Facilitating FAIR principles in practical application with metadata annotation tools. arXiv [q-bio.QM]: http://arxiv.org/abs/2103.02942 (2021).

11. Kunis S et al. MDEmic: a metadata annotation tool to facilitate FAIR image data management in the bioimaging community. Nat. Methods (in press) in press, (2021).

12. Rigano A et al. Micro-Meta App: an interactive software tool to facilitate the collection of microscopy metadata based on community-driven specifications. BioRxiv (2021) doi:10.1101/2021.05.31.446382.

13. Rigano A et al. Micro-Meta App: an interactive tool for collecting microscopy metadata based on community specifications. Nature Methods (in press) in press, (2021).

14. Ryan J et al. MethodsJ2: A Software Tool to Improve Microscopy Methods Reporting - preprint. BioRxiv (2021) doi:10.1101/2021.06.23.449674.

15. Ryan J et al. MethodsJ2: A Software Tool to Capture and Generate Comprehensive Microscopy Methods Text and Improve Reproducibility. Nat. Methods in press, (2021).

16. Baker M & Penny D 1,500 scientists lift the lid on reproducibility. Nature vol. 533 452–454 (2016). [PubMed: 27225100]

17. Marqués G, Pengo T & Sanders MA Imaging methods are vastly underreported in biomedical research. Elife 9, (2020).

18. Sheen MR et al. Replication Study: Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. eLife vol. 8 (2019).

19. Viana MP et al. Robust integrated intracellular organization of the human iPS cell: where, how much, and how variable. BioRxiv (2021) doi:10.1101/2020.12.08.415562.

20. Botvinik-Nezer R et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582, 84–88 (2020). [PubMed: 32483374]

21. Nature Editorial Staff. Better research through metrology. Nat. Methods 15, 395 (2018). [PubMed: 29855569]

22. Pines J Image integrity and standards. Open biology vol. 10 200165 (2020). [PubMed: 32574547]

23. Huisman M et al. A perspective on Microscopy Metadata: data provenance and quality control. arXiv [q-bio.QM]: https://arxiv.org/abs/1910.11370 (2021).

24. Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). [PubMed: 26978244]

25. Allan C et al. OMERO: flexible, model-driven data management for experimental biology. Nat. Methods 9, 245–253 (2012). [PubMed: 22373911]

26. Wittner R et al. ISO 23494: Biotechnology - provenance information model for biological specimen and data. in Provenance and Annotation of Data and Processes. International Provenance and Annotation Workshop (IPAW) 2020, IPAW 2021 (eds. Glavic B, Braganholo V & Koop D) vol. 12839 222–225 (Springer, Cham, 2021).

27. Holub P et al. Towards a Common Standard for Data and Specimen Provenance in Life Sciences. Zenodo (2021) doi:10.5281/zenodo.5093125.

28. Kobayashi N, Moore J, Onami S & Swedlow JR OME Core Ontology: An OWL-based Life Science Imaging Data Model. in Proceedings of the 12th SWAT4(HC)LS (Semantic Web Applications and Tools for Healthcare and Life Sciences) Conference (eds. Cornet R & Waagmeester A) vol. 2849 149–150 (http://ceur-ws.org/Vol-2849/paper-25.pdf, Dec 10–11 2019).

29. Moore J et al. On Bringing Bioimaging Data into the Open (World). in Proceedings of the 12th SWAT4(HC)LS (Semantic Web Applications and Tools for Healthcare and Life Sciences) Conference (eds. Cornet R & Waagmeester A) vol. 2849 44–53 (CEUR Workshop Proceedings (CEUR-WS.org), 2019).

30. Hammer M et al. The 4DN-OME ontology: an OME-OWL extension with emphasis on usability, minimum information guidelines and quality control for super-resolution fluorescence microscopy. in Proceedings of the 12th Annual SWAT4(HC)LS (Semantic Web Applications and Tools for Healthcare and Life Sciences) Conference (eds. Cornet R & Waagmeester A) vol. 2849 153–154 (CEUR Workshop Proceedings (CEUR-WS.org), December, 9–12 2019).

31. Moore J et al. OME-NGFF: scalable format strategies for interoperable bioimaging data. BioRxiv (2021) doi:10.1101/2021.03.31.437929.

32. Deutsch EW et al. Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE). Nat. Biotechnol 26, 305–312 (2008). [PubMed: 18327244]

33. Brazma A Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges. ScientificWorldJournal 9, 420–423 (2009). [PubMed: 19484163]

34. Rigano A & Strambio-De-Castillia C Proposal for minimum information guidelines to report and reproduce results of particle tracking and motion analysis. bioRxiv 155036 (2017) doi:10.1101/155036.

35. Miura K & Nørrelykke SF Reproducible image handling and analysis. EMBO J. e105889 (2021) doi:10.15252/embj.2020105889. [PubMed: 33480052]

36. Masuzzo P Minimum Information About Cell Migration Experiments (MIACME 0.2). https://cmso.science/MIACME/v0.2/(2021).

37. North AJ Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. J. Cell Biol 172, 9–18 (2006). [PubMed: 16390995]

38. Theer P, Mongis C & Knop M PSFj: know your fluorescence microscope. Nat. Methods 11, 981–982 (2014). [PubMed: 25264772]

39. Lambert TJ & Waters JC Assessing camera performance for quantitative microscopy. Methods Cell Biol. 123, 35–53 (2014). [PubMed: 24974021]

40. Deagle RC, Wee T-LE & Brown CM Reproducibility in light microscopy: Maintenance, standards and SOPs. Int. J. Biochem. Cell Biol 89, 120–124 (2017). [PubMed: 28606390]

41. Mubaid F et al. Fluorescence microscope light source stability. Histochem. Cell Biol 151, 357–366 (2019). [PubMed: 30767050]

42. Kiepas A, Voorand E, Mubaid F, Siegel PM & Brown CM Optimizing live-cell fluorescence imaging conditions to minimize phototoxicity. J. Cell Sci 133, (2020).

43. Hill E Announcing the JCB DataViewer, a browser-based application for viewing original image files. J. Cell Biol 183, 969–970 (2008).

44. Montero Llopis P et al. Best practices and tools for reporting reproducible fluorescence microscopy methods. Nat. Methods (10.1038/s41592-021-01156-w) ePub (online ahead of print), (2021).

45. Heddleston JM, Aaron JS, Khuon S & Chew T-L A guide to accurate reporting in digital image processing: can anyone reproduce your quantitative analysis? J. Cell Sci 134: jcs254151, (2021). [PubMed: 33785609]

46. Aaron JS & Chew T-L A guide to accurate reporting in digital image acquisition: can anyone replicate your microscopy data? J. Cell Sci 134: jcs254144, (2021). [PubMed: 33785608]

47. Swedlow JR, Goldberg IG, Brauner E & Sorger PK Informatics and Quantitative Analysis in Biological Imaging. Science 300, 100–102 (2003). [PubMed: 12677061]

48. Williams E et al. Image Data Resource: a bioimage data integration and publication platform. Nat. Methods 14, 775 (2017). [PubMed: 28775673]

49. Ellenberg J et al. A call for public archives for biological image data. Nat. Methods 15, 849–854 (2018). [PubMed: 30377375]

50. Sarkans U et al. REMBI: Recommended Metadata for Biological Images - realizing the full potential of the bioimaging revolution by enabling data reuse. Nat. Methods (10.1038/s41592-021-01166-8) ePub (ahead of print), (2021).

51. Swedlow JR et al. A Global View of Standards for Open Image Data Formats and Repositories. Nat. Methods (10.1038/s41592-021-01113-7) ePub (online ahead of print), (2021).

52. Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ & Patwardhan A EMPIAR: a public archive for raw electron microscopy image data. Nat. Methods 13, 387–388 (2016). [PubMed: 27067018]

53. Guilbert T Cochin Image Database. Cochin Image Database https://imagerie.cochin.inserm.fr/sis4web/login.php (2019).

54. Orloff DN, Iwasa JH, Martone ME, Ellisman MH & Kane CM The cell: an image library-CCDB: a curated repository of microscopy data. Nucleic Acids Res. 41, D1241–50 (2013). [PubMed: 23203874]

55. Tohsato Y, Ho KHL, Kyoda K & Onami S SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena. Bioinformatics 32, 3471–3479 (2016). [PubMed: 27412095]

56. Hammer M et al. 4DN-BINA-OME (NBO)-Microscopy Metadata Specifications - Tiers System_v2.01. (https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Tier%20System/stable%20version/v02-01, 2021). doi: 10.5281/zenodo.4710731.

57. Chen PP-S The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS) 1, 9–36 (1976).

58. Rigano A & Strambio-De-Castillia C 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XSD schema. (https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs, 2021). doi: 10.5281/zenodo.4711426.

59. Rigano A et al. 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XLS Spreadsheet and Entity Relationship schemas. (https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs, 2021). doi:10.5281/zenodo.4711229.

60. Bradner S [RFC2119] Key words for use in RFCs to Indicate Requirement Levels. http://www.ietf.org/rfc/rfc2119.txt (1997).

61. de Souza N The ENCODE project. Nature methods vol. 9 1046 (2012). [PubMed: 23281567]

62. Movincell Consortium. Multi-dimensional marine organism dataview. Movincell http://movincell.org/ (2015).

63. Allen Institute for Cell Science. Allen Institute for Cell Science. Allen Cell Explorer. alleninstitute.org https://www.allencell.org/(2017).

64. Rozenblatt-Rosen O et al. Building a high-quality Human Cell Atlas. Nat. Biotechnol 39, 149–153 (2021). [PubMed: 33500565]

65. HuBMAP Consortium. The Human BioMolecular Atlas Program - HuBMAP. https://hubmapconsortium.org/ https://hubmapconsortium.org/ (2017).

66. DORY Working Group. Defining Our Research Methodology. https://doryworkspace.org/ https://doryworkspace.org/ (2019).

67. Sansone S-A et al. The First RSBI (ISA-TAB) Workshop: 'Can a Simple Format Work for Complex Studies?' OMICS 12, 143–149 (2008). [PubMed: 18447634]

68. Ioannidis JPA et al. Repeatability of published microarray gene expression analyses. Nat. Genet 41, 149–155 (2009). [PubMed: 19174838]

69. Rung J & Brazma A Reuse of public genome-wide gene expression data. Nat. Rev. Genet 14, 89–99 (2013). [PubMed: 23269463]

70. Grebnev G & BioImaging G Global BioImaging Training Resource. https://globalbioimaging.org https://globalbioimaging.org/international-training-courses/repository (2021).

71. Waters J, Jug F, Elliott H & Lambert T Cold Spring Harbor Laboratory Courses: Quantitative imaging: From acquisition to analysis. https://meetings.cshl.edu https://meetings.cshl.edu/courses.aspx?course=C-QICM&year=21 (2021).

72. North A & Bewersdorf J Marine Biology Laboratory Courses: Optical microscopy & imaging in the biomedical sciences (OMIBS). https://www.mbl.edu https://www.mbl.edu/education/courses/optical-microscopy-imaging/ (2021).

73. Allen L, O'Connell A & Kiermer V How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. Learn. Publ 32, 71–74 (2019).

Text Box I

Charting a solid path towards next-generation storage mechanisms for community-driven, OME-based Microscopy Image Data Standards

Microscopy Metadata, stored (Figure 1, **HOW** *yellow* and **WHERE** *blue bubbles*) following the Open Microscopy Environment (OME) Data Model[4,5] is represented in the form of OME-Extensible Markup Language (OME-XML), which is typically stored in the header of OME-TIFF files. Consequently, the XML Schema Definition (XSD) formalism is used to represent the model schema in a machine-readable manner. However, despite its advantages, XSD is not ideally suited to allow the OME Data Model to serve as the foundation for the community-development and maintenance of globally accepted light microscopy standards (Figure 1). Because XSD does not support the storage of novel types of information within the core of the model, the capture of ever-evolving microscopy technologies and modalities requires the periodical release of new versions of the OME XSD schema (https://docs.openmicroscopy.org/ome-model/ 6.2.2/) accompanied by XML Stylesheet Language (XSL) based templates for making sure legacy documents could be kept up to date. This burdensome process is ultimately unsustainable. Consequently, it is necessary to develop new strategies with a more open paradigm.

Under this new paradigm, one would assume that no single authority exists to decide which information must be recorded in metadata models making it necessary for commonly used concepts to be incorporated over time into community-driven standards. In this context, agreement has to be reached on **WHAT** concepts have to be recorded for the documentation of imaging experiments (Figure 1, *magenta bubble*), and in particular on the development of shared mechanisms defining **HOW** new types of (meta)data have to be recorded (Figure 1, *yellow bubble*) and associated with the Image data file format (Figure 1, **WHERE** *blue bubble*)[31].

In this context, the OME consortium, in collaboration with RIKEN, has started experimenting with the idea of utilizing Resource Description Framework (RDF; https:// www.w3.org/RDF/) triples conforming to the W3C Web Ontology Language (OWL; https://www.w3.org/OWL) to describe OME-compatible image metadata[28–30] to be incorporated in the Next-Generation File Formats (NGFF) currently being developed by the OME consortium[31]. By employing this method, it would be possible for users to produce, find and access quality-controlled image data for re-analysis and integration. Specifically, the depicted method will provide two major advantages:

1. Individual groups specializing in different aspects of the imaging world will have equal status and a shared path to develop new areas of the model[10,11,29]. In turn, this will provide a method for different communities to collectively develop a complete picture (Figure 2) of all the information required to ensure rigor and reproducibility for modern imaging experiments.

2. At the same time, community-driven standards could evolve gradually over time by incorporating novel concepts into the core as they are developed peripherally from the core, vetted by the community, and commonly adopted.

As a *proof-of-concept*, an implementation of the OME Data Model was built in RDF/OWL[28], and applied to the modeling of specifications proposed here for the exchange of image data and integration with genomics datasets[30]. This demonstrated the potential utility of this approach, laying the foundation for ongoing community discussions to identify the path of choice for modern Light Microscopy Image Data Standards (Figure 1).
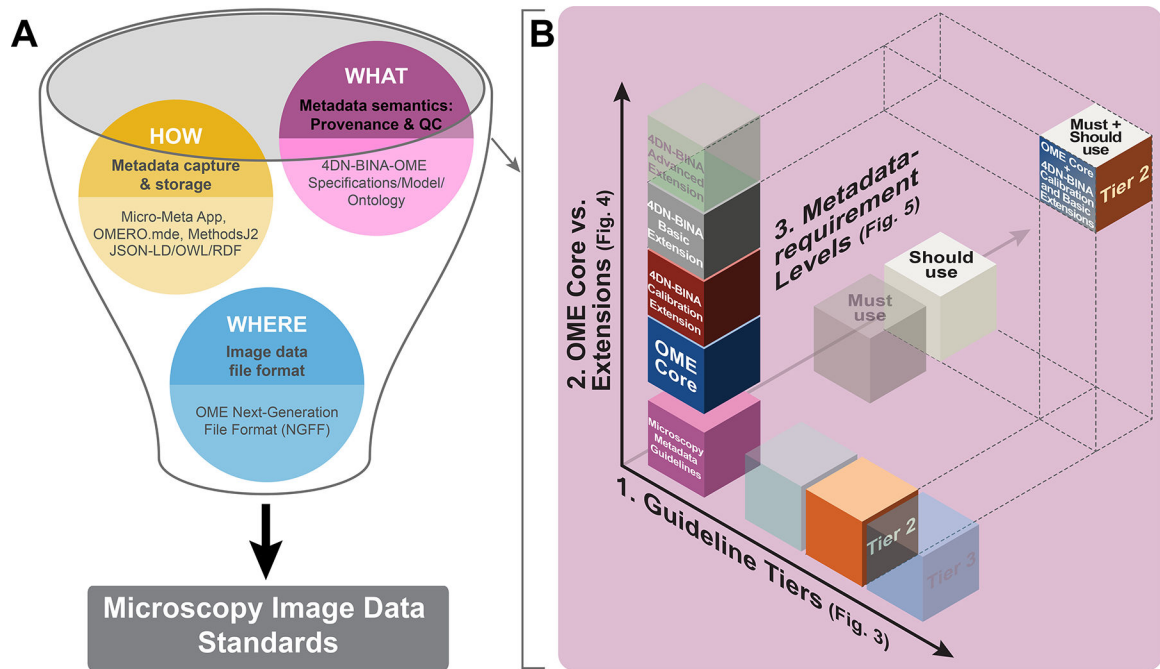
Figure 1 |. The definition of community-driven Microscopy Image Data Standards requires three complementary components and needs a flexible framework to manage complexity.
A) The establishment of community-driven Microscopy Image Data Standards requires development on three interrelated fronts: 1) Community-driven specifications for **WHAT** Microscopy Metadata information about an imaging experiment are essential for rigor, reproducibility, and reuse and should therefore be captured in Microscopy Metadata (*magenta bubble*); 2) Shared rules for **HOW** the (ideally) automated capture, representation and storage of Microscopy Metadata should be implemented in practice (*yellow bubble*). Last but not least, 3) Next-Generation File Formats (NGFF) **WHERE** the ever-increasing scale and complexity of image data and metadata would be contained for exchange[31]; *blue bubble*). **B)** The 4DN-BINA-OME specifications for *WHAT* hardware specifications, image acquisition settings, and quality control metrics should be reported articulate along three complexity axes: 1) **Guideline Tiers**: The three guideline Tiers are employed to scale reporting requirements with experimental complexity. 2) **Model Core vs. Extensions**: The use of the Core of the OME Data Model vs. one or more of the 4DN-BINA extensions allows capturing different microscopy modalities. 3) **Metadata-requirement Levels**: The distinction between *Must use* vs. *Should use* metadata fields, is used to define what information is needed for different reporting purposes (i.e., quality, reproducibility, sharing value). Depicted is the intersection between the three dimensions (*OME Core + 4DN-BINA Basic and Calibration extensions* ∩ *Tier 2* ∩ *All available fields*) that would be appropriate to describe an experiment in which a wide-field microscope is used to capture the dynamics of viral particle trafficking within infected cells. ∩, intersection.
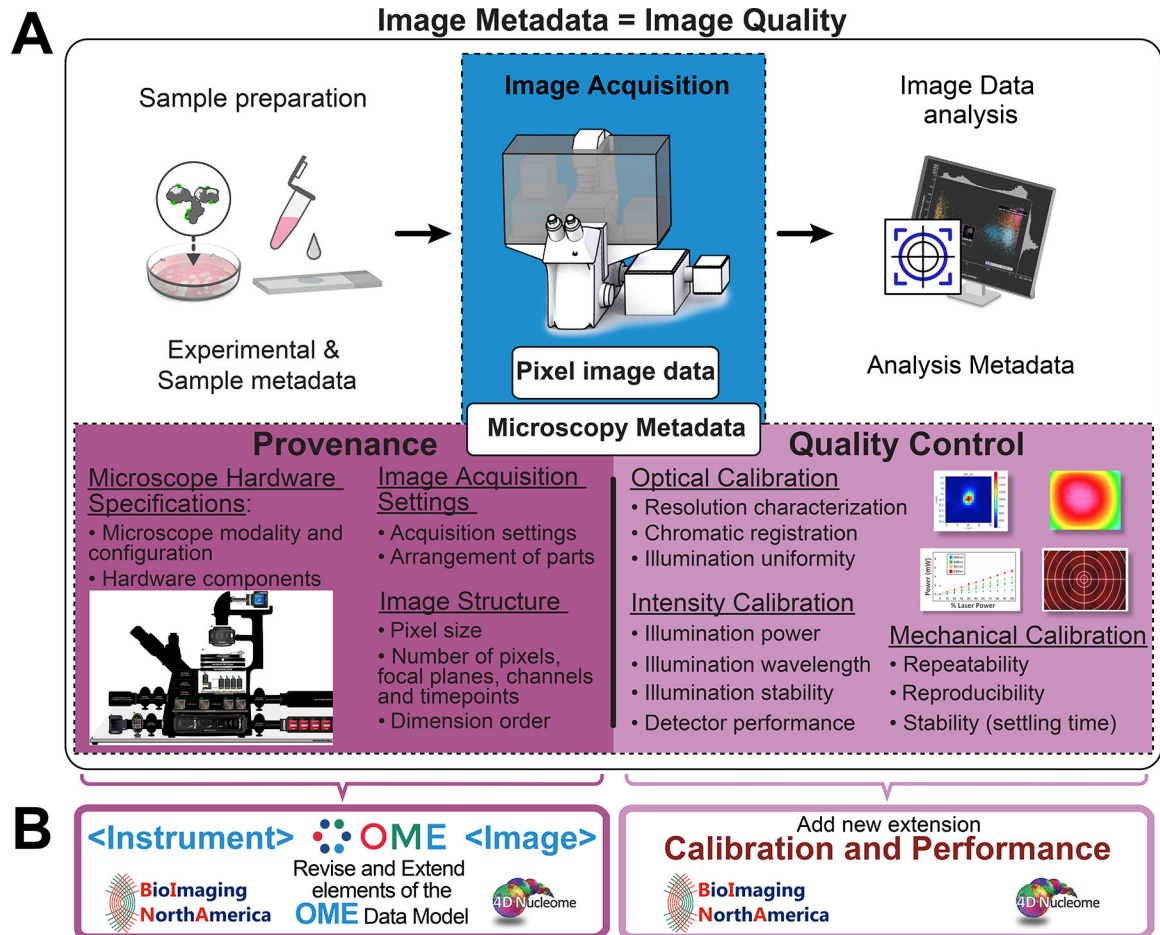
A

Image Metadata = Image Quality

Sample preparation

Experimental &
Sample metadata

Image Acquisition

Pixel image data

Microscopy Metadata

Image Data
analysis

Analysis Metadata

Provenance

Microscope Hardware
Specifications:
• Microscope modality and
configuration
• Hardware components

Image Acquisition
Settings
• Acquisition settings
• Arrangement of parts

Image Structure
• Pixel size
• Number of pixels,
focal planes, channels
and timepoints
• Dimension order

Quality Control

Optical Calibration
• Resolution characterization
• Chromatic registration
• Illumination uniformity

Intensity Calibration
• Illumination power
• Illumination wavelength
• Illumination stability
• Detector performance

Mechanical Calibration
• Repeatability
• Reproducibility
• Stability (settling time)

B

<Instrument> **OME** **<Image>**

Revise and Extend
elements of the
OME Data Model

BioImaging NorthAmerica

4D Nucleome

Add new extension
Calibration and Performance

BioImaging NorthAmerica

4D Nucleome

Figure 2 |. Light Microscopy Metadata is essential for the assessment, interpretation, reproducibility, comparison, and re-use of the results of microscopy experiments.
A) Depicted is a schematic representation of a typical bio-imaging experiment and of the *Image Metadata* that has to be collected to ensure the quality reproducibility and scientific value of the resulting *Pixel Image Data* (*blue box*). Specifically, imaging experiments and the associated metadata can be subdivided as follows: 1) *Sample preparation* documented by *Experimental & Sample Metadata*; 2) *Image Data Acquisition* documented by *Microscopy Metadata;* 3) *Image Analysis* documented by *Analysis Metadata*. In turn, *Microscopy Metadata* can be subdivided in two categories as indicated (*magenta boxes*): 1) *Provenance* metadata includes information that documents *Microscope Hardware Specifications*, *Image Acquisition Settings,* and *Image Structure*; 2) *Quality Control* metadata includes metrics that quantitatively assess the performance of the microscope and the quality of image data and are obtained through the execution of specifically designed *Optical*, *Intensity* and *Mechanical calibration* procedures. **B)** In order to capture and store Microscopy Metadata, the 4DN-BINA-OME specifications presented here take advantage of the structure of the OME Data Model[4,5], which serves as the *de facto* specification for the exchange of image data and metadata. Specifically, *Provenance* metadata is stored into revised and extended versions of the **<Instrument>** and **<Image>** elements of the

OME Data Model. On the other hand, ***Quality Control*** metadata is stored utilizing a newly designed ***Calibration and Performance extension*** of the same model.
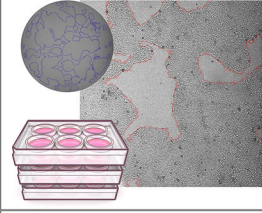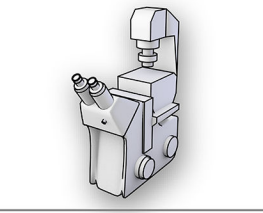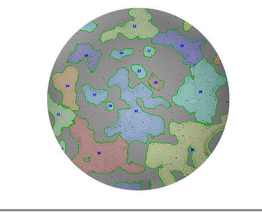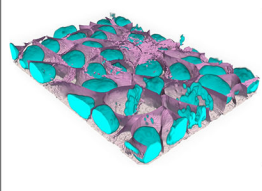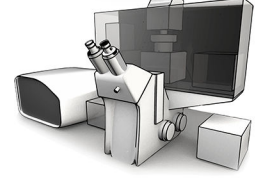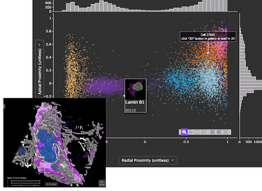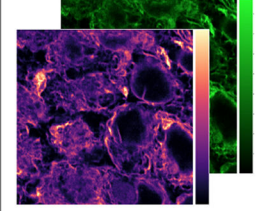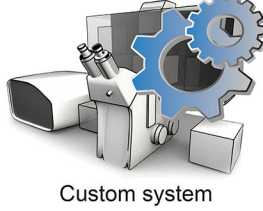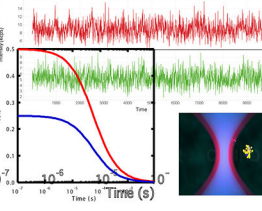
Figure 3 |. Scaling light microscopy metadata guidelines with experimental, technical, and analytical intent and complexity minimize recordkeeping burden while maximizing value, quality, and reproducibility of image data.

Shown is a schematic representation of the graded system for metadata specifications proposed by 4DN and BINA to tailor reporting requirements to experimental complexity. In this system microscope hardware and imaging experiments are classified based on the following *a priori* criteria: 1) ***Experiment and Image*** complexity, 2) ***Microscope*** technology and imaging ***Modality,*** and 3) ***Results and Analysis*** requirements. For each of the criteria, the schema provides graphical illustrations of increasing complexity along the Tier-axis that can be used as an initial guide for microscope users to map reporting requirements to their experimental needs.
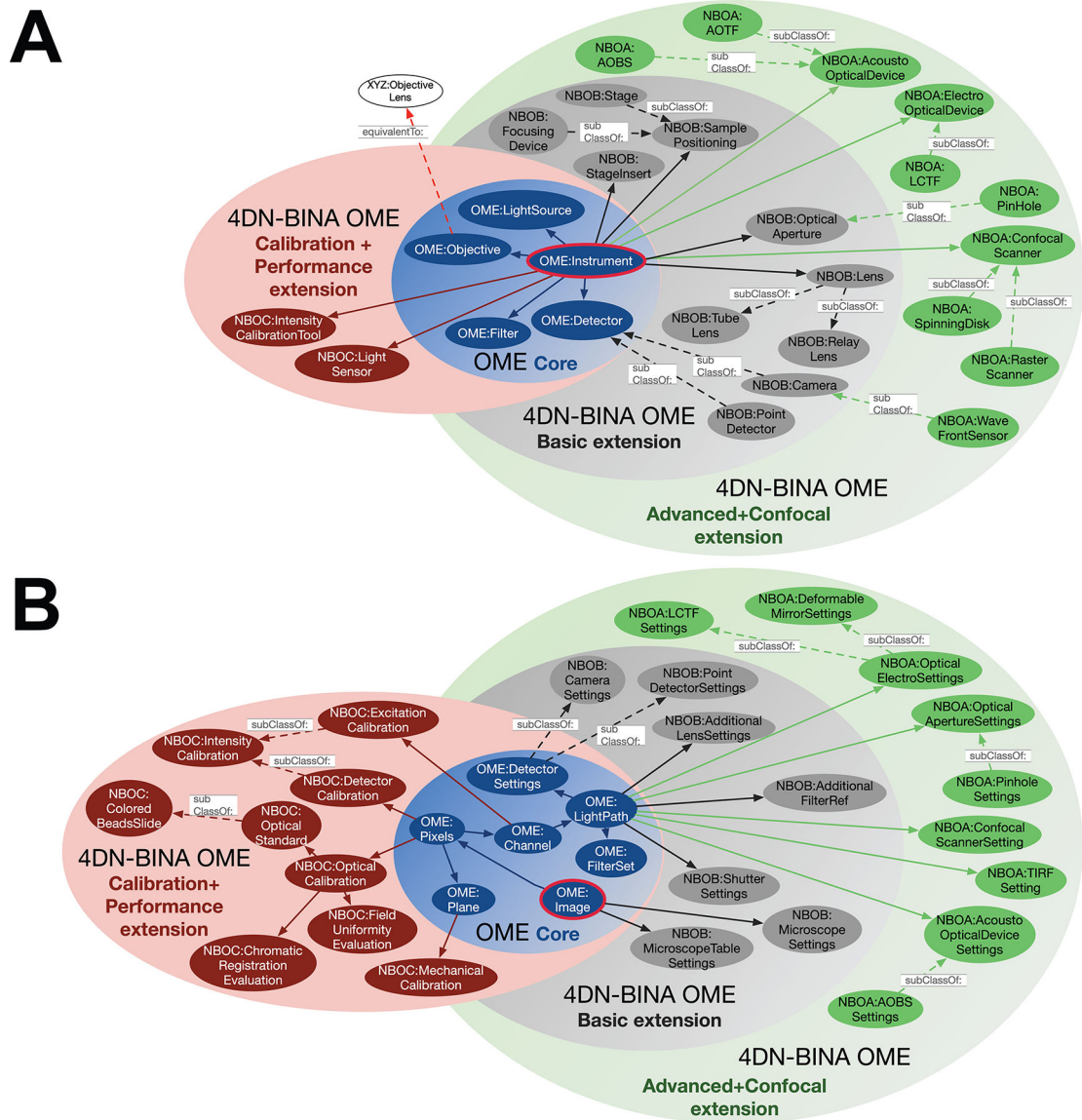
Figure 4 |. The 4DN-BINA Light Microscopy Metadata Guidelines extends the core of the OME Data Model.

Portrayed are Venn diagrams containing a Linked-Open-Data (LOD; Figure 6B) representation of the core vs. extension relation between metadata model elements that belong to the core of the OME Data Model (*OME:* namespace; blue ovals) and those that belong to the three proposed extensions specified by 4D**N**-**B**INA-**O**ME (*NBO:* namespace; maroon, grey and green ovals). Specifically, the schema illustrates the relationship between the <Instrument> (**A**) and <Image> (**B**) elements (*OME:Instrument; OME:Image*; red-bordered blue oval) and their sub-elements belonging to the Core of the OME Data Model (*light blue set* containing *blue ovals*; e.g., *OME:Filter*, *OME:Channel*, etc.,), with sub-elements specified by the *Calibration + Performance* (*pink set* containing *maroon ovals*; e.g., *NBOC:LightSensor*, *NBOC:OpticalCalibration*, etc.,), *Basic* (*light grey* set containing *dark grey ovals*; e.g., *NBOB:OpticalAperture*, *NBOB:CameraSettings*, etc.,), and *Advanced + Confocal* (*green set* containing *green ovals*; e.g., *NBOA:SpinningDisk*,

NBOA:PinHoleSettings, etc.,) 4DN-BINA-OME extensions. The schema is not intended to be comprehensive and only includes a small subset of the classes that compose the model. Abbreviations: *ABOS*, Acousto-Optical Beam Splitter; *AOTF*, Acousto-Optical Tunable Filter; *LCTF*, Liquid Crystal Tunable Filter.
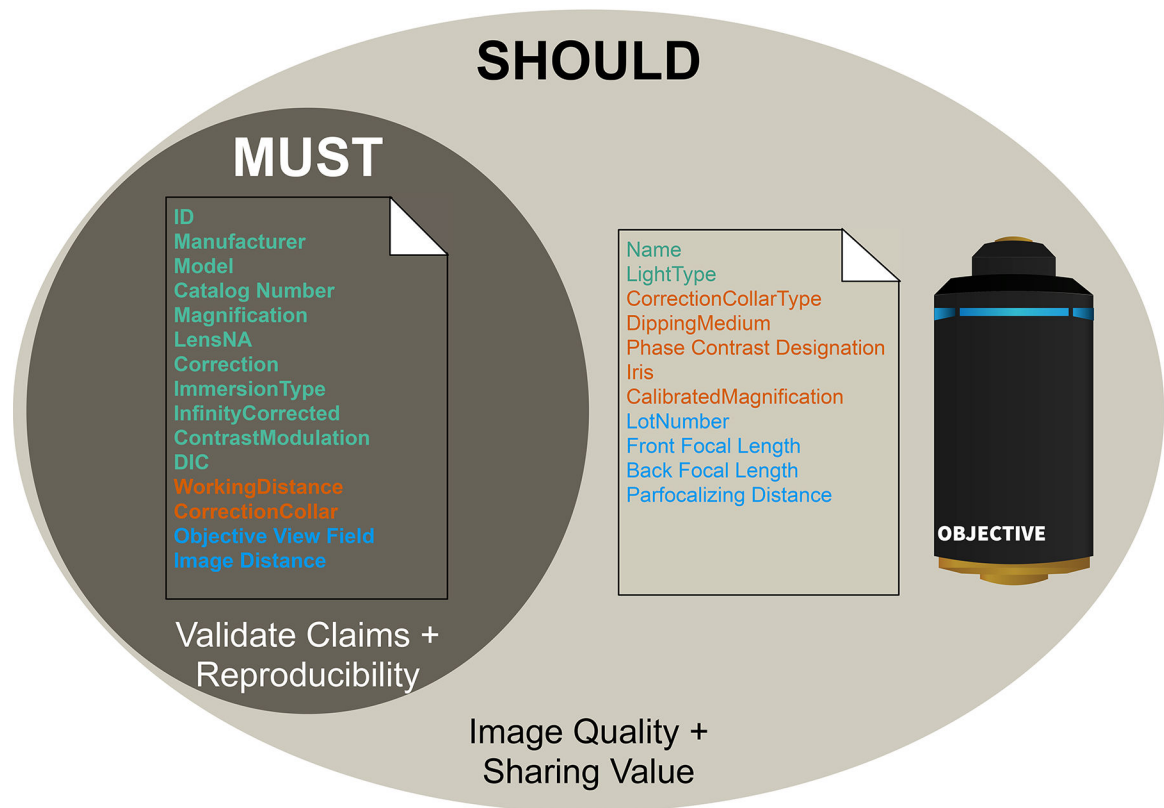
Figure 5 |. The third axis of the 4DN-BINA-OME light Microscopy Metadata Guidelines adds further flexibility to minimize imaging experiment documentation burden.
Depicted is an example Venn diagram representing attributes that are required to document the characteristics of an objective lens, and that are stored in the `<Objective>` element of 4DN-BINA-OME Core and Basic extension. In the schema, Objective attributes are color-coded based on their Tier-level and are subdivided requirement-level categories based on the following criteria: 1) *required (MUST)* fields are necessary to *Validate Claims + Reproducibility*; 2) *recommended (SHOULD)* field are prescribed to ensure maximal *Image Quality + Sharing value*. Color-coding is consistent with the one utilized throughout the manuscript: *Green*, Tier 1, *Orange*, Tier 2; *Blue*, Tier 3.
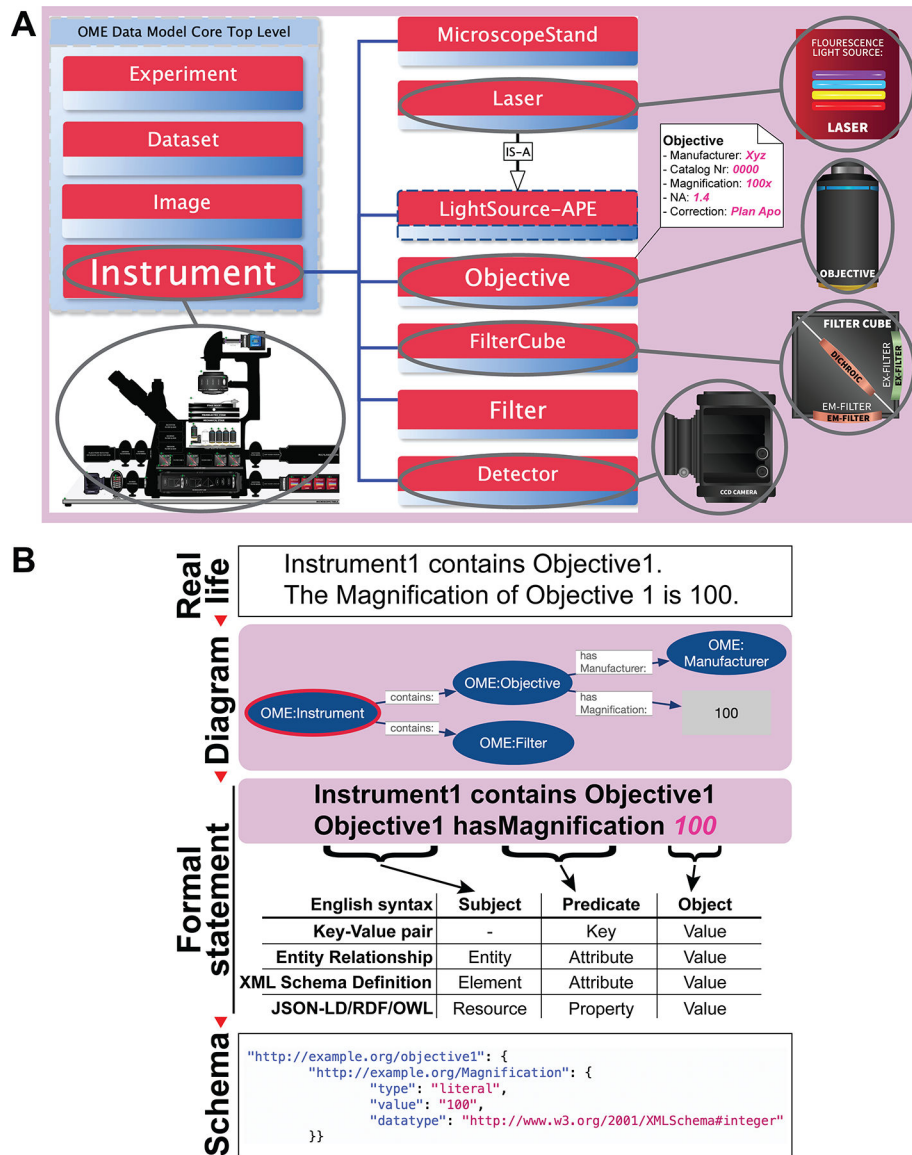
Figure 6 |. A data model is a schematic representation of reality that can be utilized to organize metadata and produce tools.

A) Building visually compelling conceptual metadata models captures not only individual attributes and their values, but also the often-complex relationships between different real-world's entities. Presented is a simplified *Entity Relationship* (ER)[57]-depiction of the 4DN-BINA-OME data model that portrays the hardware configuration of a microscope. In this formal representation: 1) *Solid-lines boxes* are used to symbolize individual hardware components (e.g., <Light Source>, <Objective>, etc.,). In addition, dashed-line boxes denote generalized element-families to which specific "children" elements belong (i.e., a Laser belongs to the Light Source family). 2) Lists of attributes (*key-value pairs* enclosed *in a white box*) represent metadata that needs to be recorded about each hardware component (e.g., Magnification, Numerical Aperture etc.). 3) *Lines* are utilized to model relationships between components. Specifically, *Blue lines* represent "*HAS-A*" relationships (i.e., "An Instrument HAS-A Light Source"). *Black arrows* represent "*IS-A*" relationships connecting

generalized to specific concepts (i.e., "A Laser IS-A Light Source"). Based on the rules indicated above, the depicted schema can be read to signify: "This Instrument has a Laser, which is a specific type of Light Source, and has an Objective built by Nikon, with 100x Magnification and 1.4 Numerical Aperture". **B)** Starting from human readable statements describing the data (***Real life***, top panel), depicted is the process that is often employed to produce actionable code (***Schema***, bottom panel) used to build essential metadata capture and management tools. Statements are first rendered into graphical illustrations that provide a bird's eye view of the entire system, such as the Linked-Open-Data (LOD) graph depicted here (***Diagram***; second panel from the top). Subsequently, diagrams are parsed to produce structured statements (***Formal statement***, second panel from the bottom) using one or more available methods (e.g., *English syntax*, *Key-value pair*, *Entity Relationship*, *XML Schema Definition*, *JSON-LD/RDF/OWL*, etc.,). Finally, statements are encoded using formal schema languages. In the depicted example (bottom panel), JSON-Linked Data (JSON-LD; https://www.w3.org/TR/json-ld11/) is used to serialize Resource Description Framework (RDF; https://www.w3.org/RDF/) triples to build extensible LOD information graphs.

Table I:

Tiers for light microscopy metadata reporting as proposed by the Imaging Standards Working Group of the 4D Nucleome initiative and by the Quality Control and Data Management Working Group of Bioimaging North America.

Category	Nr.	Name	Description	Example
Descriptive	1	Minimum Information/ Qualitative or Basic Quantification/ Material & Methods	Developmental and stem-biology experiments in which qualitative analysis of image data is used to support major findings, transfection control, viability assay, counting of cells and nuclei, expression level measurements, localization of markers in cellular sub-compartments	Histochemistry, Immuno-Histochemistry, Fluorescent In Situ Hybridization (FISH), Immuno Fluorescence (IF), Fluorescent Protein (FP) labelling
Analytical	2	Advanced Quantification	Reporting quantitative effects that require advanced quantification including the localization of single molecules and tracking of intracellular dynamics	Diffraction-limited spot localization, measurement of distances, co-localization studies, detection of low-signal features, advanced processing, cell tracking and single-particle tracking, dynamic expression level quantification
	3	Manufacturing/ Technical Development/Full Documentation	Full documentation of microscopic setup, image acquisition and quality control	Microscopy hardware manufacturing; development of novel and yet to be validated technology in both commercial and academic settings; full reproducibility of microscopy set-up and image acquisition settings

Each tier accommodates increasingly complex images, experiments, instrumentation, and analytical needs and therefore requires progressively more metadata. This tiered system is not intended to meet the needs of all imaging communities. Rather it is proposed as a framework that might need to be adapted and modified depending on the needs of individual data collection consortia, disciplines, or institutions.

Table II:

Example utilization of the three axes matrix of the 4DN-BINA-OME Microscopy Metadata Specifications

Nr.	Description	Sample	Resolution	Image structure	Analysis	1st Axis Tier	2nd Axis OME extension	3rd Axis Requirement level
1	Investigation of PTC299 effect on SARS-CoV2 replication in Vero cells using widefield commercial microscope	Fixed cells	Low	Dual-channel; multi-point	Simple segmentation and intensity quantification	1	C + B	Required only
2	High-resolution OligoPaint investigation of chromatin structure using commercial widefield microscope	Fixed cells	High	Multi-channel; 3D Z-stack	Spot detection and subpixel localization	2	C + B + CP	Required only
3	High-resolution, real-time tracking of export of HIV-1 genomic RNA from infected T-cells using custom made microscope	Live cells	Super	Multi-channel; 3D Z-stack; time-series	Spot detection, subpixel localization and tracking	3	ALL	ALL
4	Widefield microscope manufacturers	Any	Any	Any	Any	3	C + B + CP	ALL

C, OME Core; **B**, 4DN-BINA-OME Basic Extension; **AC**, 4DN-BINA-OME Advanced & Confocal Extension; **CP**, 4DN-BINA-OME Calibration & Performance Extension