

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

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.

Mathias Hammer, Maximiliaan Huisman, Alessandro Rigano, Ulrike Boehm, James J. Chambers, Nathalie Gaudreault, Alison J. North, Jaime A. Pimentel, Damir Sudar, Peter Bajcsy, Claire M. Brown, Alexander D. Corbett, Orestis Faklaris, Judith Lacoste, Alex Laude, Glyn Nelson, Roland Nitschke, Farzin Farzam, Carlos S. Smith, David Grunwald and Caterina Strambio-De-Castillia

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 that 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)<sup>1,2</sup> Imaging Standards Working Group (IWG), working in conjunction with the BioImaging North America (BINA) Quality Control and Data Management Working Group (QC-DM-WG)<sup>3,4</sup>, here propose flexible microscopy metadata specifications for light microscopy<sup>5–7</sup> 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<sup>8,9</sup>, which forms the basis for the ubiquitous Bio-Formats library<sup>9</sup>. Because of their tiered nature, the proposed specifications clearly define which provenance<sup>10</sup> and quality-control metadata should be recorded for a given experiment. This endeavor is closely aligned with the recently established QUALity Assessment and REProducibility for Instruments and Images in Light

Microscopy (QUAREP-LiMi) global community initiative<sup>11–13</sup>. As a result, the ensuing 4DN-BINA-OME (NBO) framework<sup>5,14</sup>, alongside three interoperable metadata collection tools being developed in parallel (OMERO.mde, Micro-Meta App and MethodsJ2)<sup>15–20</sup>, represents a major turning point toward 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.

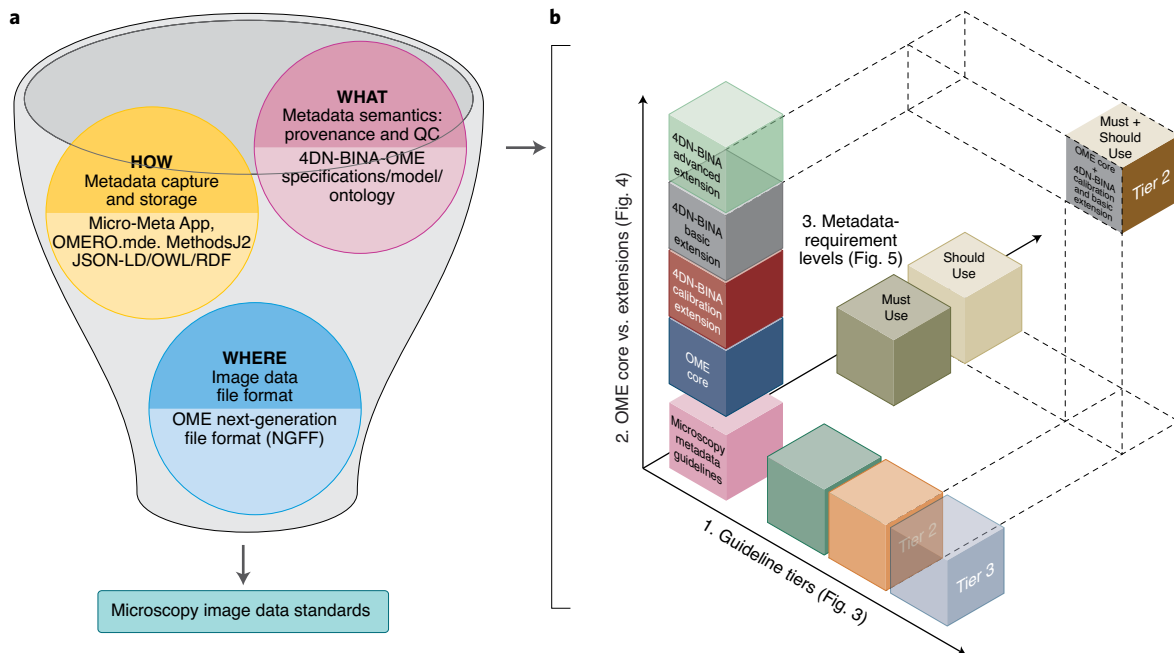
## Introduction

The reproducibility crisis affecting the biological sciences is well documented<sup>21–25</sup>. 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<sup>9,22,26,27</sup>. 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 reusability, and should be recorded in scientific publications and alongside the actual image data in the file header or in supplementary files<sup>6,7</sup>.

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<sup>6,28</sup> (see also Text Box 1 in ref. <sup>6</sup>). 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 for recording this quantitative data and metadata faithfully and reproducibly have not kept up.

The OME consortium<sup>29,30</sup> has made significant advances with the development of the OME Data Model<sup>8,9</sup>, which, together with the ubiquitous Bio-Formats image file format conversion library<sup>9</sup>, 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; Fig. 1)<sup>8,9</sup>, 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 microscopy metadata developed jointly by the 4DN<sup>1,2</sup> IWG and by the BINA QC-DM-WG<sup>3</sup> to extend the OME Data Model<sup>8,9</sup> (Figs. 1 and 2). In order to foster widespread adoption of the 4DN-BINA-OME<sup>5</sup> framework (Fig. 1a, pink bubble), key components of this effort



**Fig. 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 is essential for rigor, reproducibility and reuse and should therefore be captured in microscopy metadata (pink bubble); (2) shared rules for *HOW* the (ideally) automated capture, representation and storage of microscopy metadata should be implemented in practice (yellow bubble); and, last but not least, (3) next-generation file formats (NGFFs) *WHERE* the ever-increasing scale and complexity of image data and metadata would be contained for exchange<sup>36,37</sup>; 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* and *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  $\cap$  Tier 2  $\cap$  All available fields, where  $\cap$  signifies intersection) 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.

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<sup>15–20</sup> and are coupled with standards for metadata representation and storage (Fig. 1a, yellow bubble)<sup>31–35</sup>; and (2) sustainable roadmaps for the switch from proprietary image data formats to common, cloud-ready OME Next-Generation File Formats (NGFF, Fig. 1a, blue bubble)<sup>36,37</sup>. Importantly, all of these activities are expected to be carried out in the context of QUAREP-LiMi<sup>11–13</sup> and involve key members of the community, including microscope users, custodians and manufacturers, imaging scientists, national and global bioimaging organizations, bioimage informaticians, standards organizations, funders and scientific publishers.

The proposed 4DN-BINA-OME Microscopy Metadata (Fig. 2) Specifications articulate along three mutually independent axes (Fig. 1b).

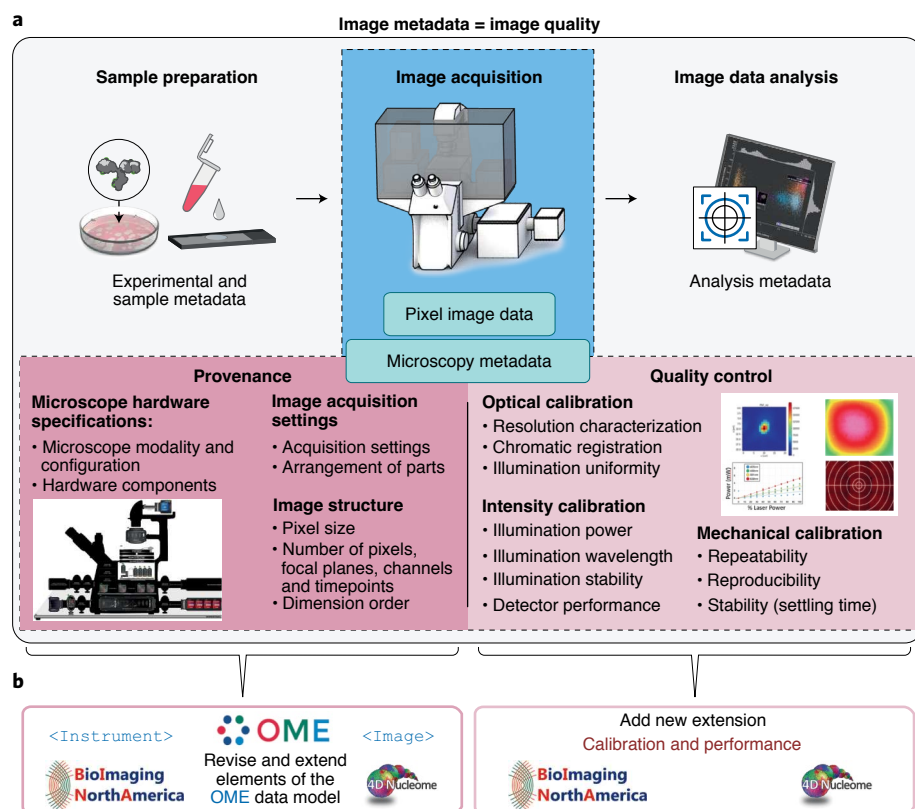
1. **Guideline tiers**—metadata specification (Fig. 3): a system of adaptable tiers that spells out which specific subset of metadata information should be included depending on experimental context and intent, technical complexity and image analysis needs.
2. **Core model and extensions**—metadata extension (Fig. 4): a suite of extensions that expand the core of the OME Data Model<sup>8,9</sup> to comprehensively capture state-of-the-art transmitted light and wide-field 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<sup>11–13</sup>.
3. **Metadata-requirement levels**—metadata inclusion (Fig. 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; Fig. 1b, *Must Use*) or “recommended” (useful to improve image quality and to maximize scientific and sharing value; Fig. 1b, *Should Use*).

Although 4DN-BINA-OME is inherently adaptable (Fig. 1b), 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.

### 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



**Fig. 2 | Light microscopy metadata are essential for the assessment, interpretation, reproducibility, comparison and reuse of the results of microscopy experiments.** **a**, A schematic representation of a typical bioimaging experiment and the *image metadata* that must 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 and sample metadata*; (2) *image data acquisition* documented by *microscopy metadata*; and (3) *image analysis* documented by *analysis metadata*. In turn, microscopy metadata (pink boxes) can be subdivided in two categories as indicated: (1) *provenance* metadata include information that documents *microscope hardware specifications*, *image acquisition settings* and *image structure* and (2) *quality-control* metadata include 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<sup>8,9</sup>, which serves as the de facto specification for the exchange of image data and metadata. Specifically, *provenance* metadata are stored into revised and extended versions of the <Instrument> and <Image> elements of the OME Data Model. On the other hand, *quality-control* metadata are stored using a newly designed *Calibration and Performance* extension of the same model.

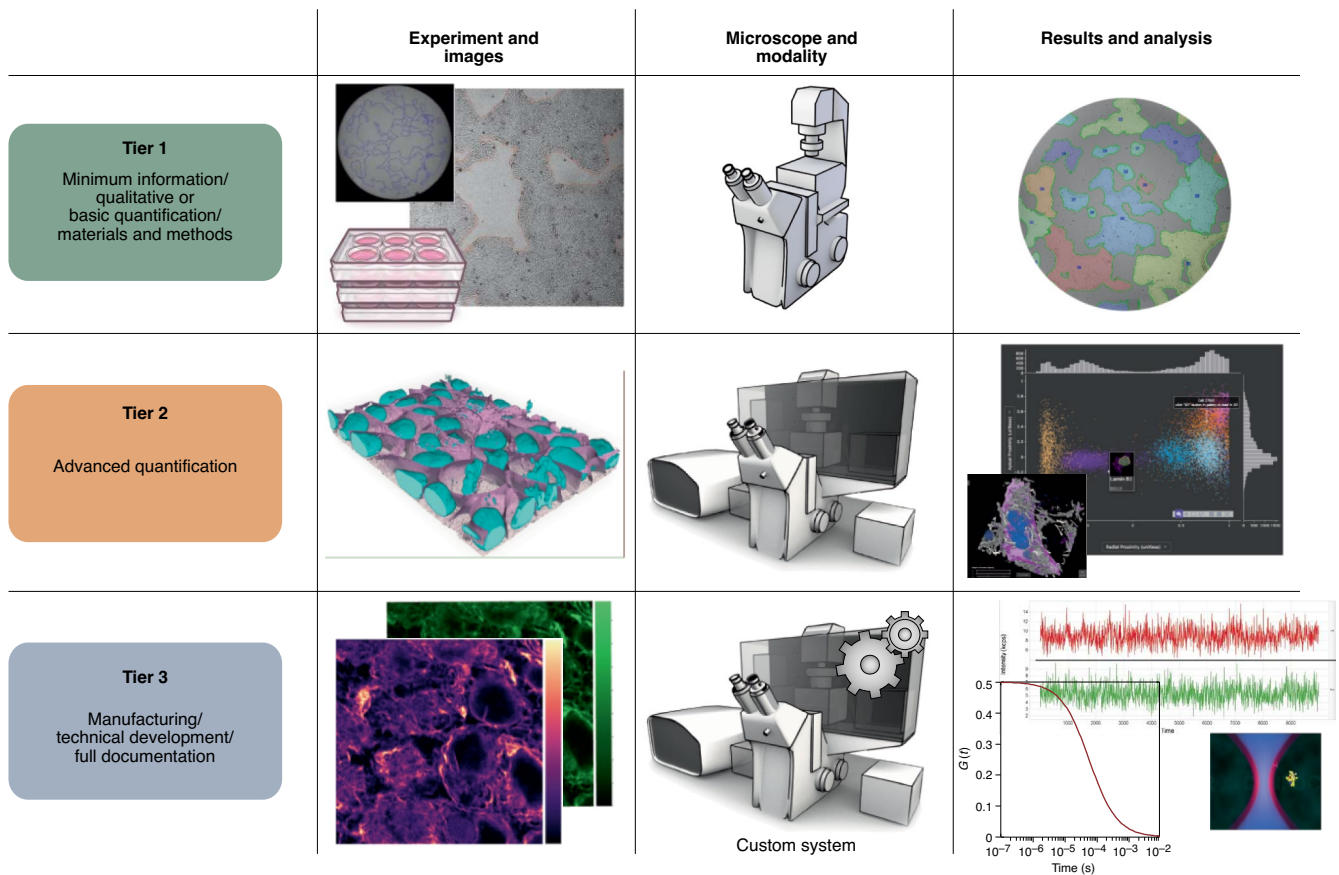
profound ways. First, it has led to digital image formation, signal processing and computational methods that enable the extraction of quantitative information from images and that have transformed light (and, in particular, fluorescence) microscopy into a key quantification tool for biomedical research. Second, it has allowed the increasingly accurate recording of progressively lower amounts of light signal, enabling the visualization and quantitative measurement of subcellular 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 use 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 they can be used to answer the questions it was intended to address, but also on the extent to which they can be trusted and reused by others. It follows that in performing imaging experiments, scientific rigor is inextricably tied to image quality, the reproducibility of experimental results and the degree to 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”) <sup>10</sup>, as well as metrics that quantitatively assess the quality of the microscope and of the images (i.e., “quality control”) <sup>6,7,11,12</sup>. A typical light microscopy experiment includes three (sometimes integrated) major steps centered around the production of image data (Fig. 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<sup>6</sup> describing any and all information that allows the actual image data (i.e., quantitative values associated with the image pixels; Fig. 2, *pixel image data*) and imaging results to be evaluated, interpreted, reproduced, found, cited, compared and reused as established by measurable data quality criteria (i.e., FAIR principles)<sup>6,9,34</sup>. Fundamentally, image metadata can be defined as metadata that document all phases of a typical microscopy experiment (Fig. 2) from (1) experimental treatment, sample preparation and labeling (Fig. 2a, *experimental and sample metadata*)<sup>38,39</sup> to (2) microscope hardware specifications, image acquisition settings, microscope performance metrics and image data structure (Fig. 2a, *microscopy metadata*)<sup>6</sup>; to (3) details about any image analysis procedure employed to extract quantitative information from the images (Fig. 2a, *analysis metadata*)<sup>40–42</sup>. As such, *microscopy metadata* consist of a subset of image metadata and, in turn, can be subdivided into two subcategories<sup>6,7</sup>: (1) *microscopy data provenance metadata* (MPM) describing the origin of the data microscope hardware specifications, image acquisition settings and image structure (Fig. 2a, *provenance*); and (2) *microscopy quality-control metadata* (MQM) including



**Fig. 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 criteria: (1) *experiment and image* complexity, (2) *microscope* technology and imaging *modality* and (3) *results and analysis* requirements. For each criterion, the schema provides graphical illustrations of increasing complexity along the three-tier axis that can be used as an initial guide for microscope users in mapping reporting requirements to their experimental needs.

calibration metrics that quantitatively assess the performance of the microscope (Fig. 2a, *quality control*). In addition to capturing MPM and MQM (Figs. 1 and 2), microscopy metadata standards should also address the following:

1. *Light microscopy utilizes a vast array of adaptable modalities*, each requiring the reporting of different metadata 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 the file formats* and the consequent need for raw data files to be converted into other formats before interpretation and comparison often yield a significant loss of metadata, or, worse still, inadvertently compromises the data during the conversion process.

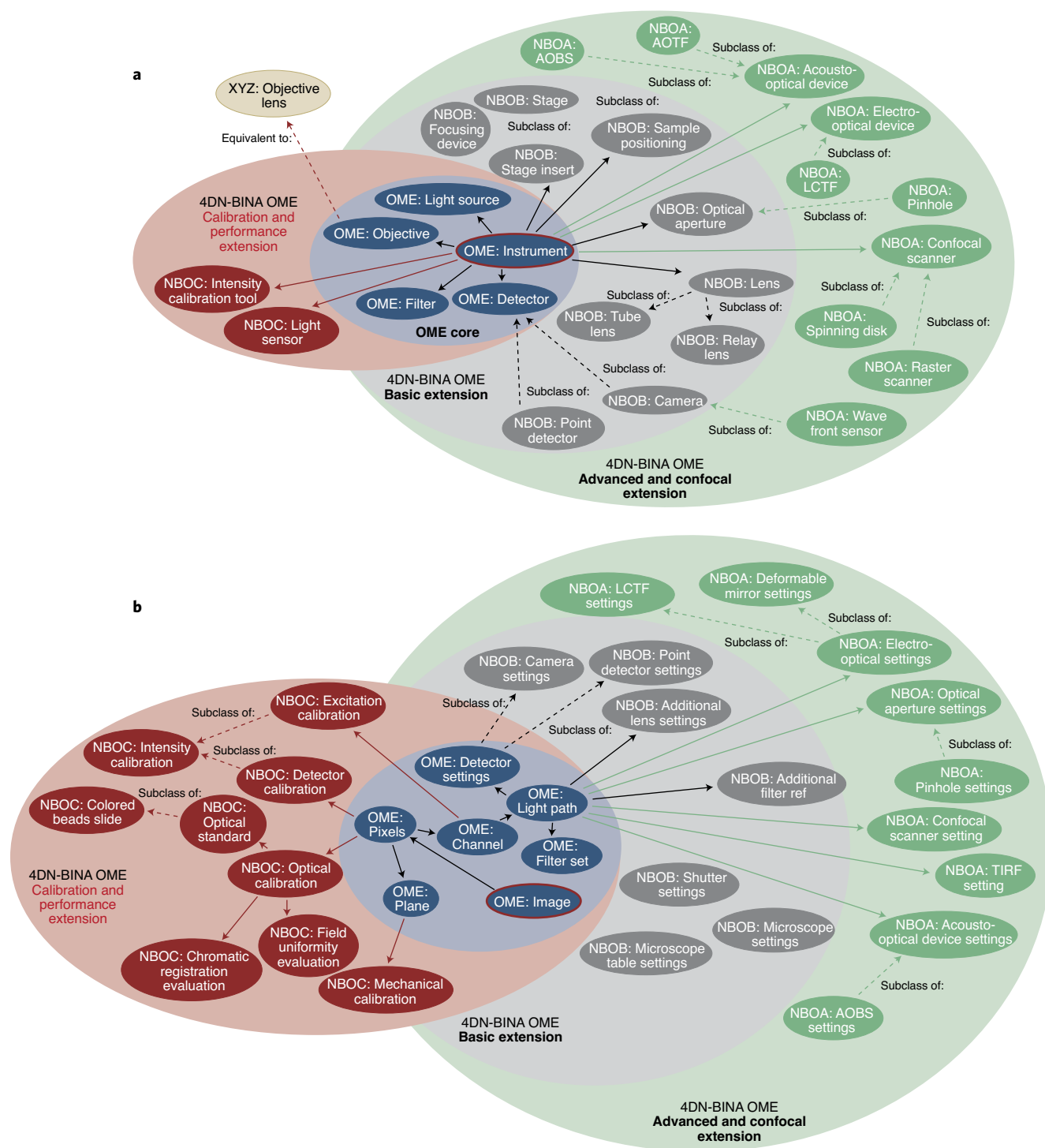
Despite this apparent complexity, it is worth noting that the image acquisition step of an imaging experiment (Fig. 2) is eminently manageable and quantifiable, as long as the microscope and imaging system are properly documented, maintained and operated. Consequently, the development of community-sanctioned specifications for microscopy metadata that encompass MPM and MQM not only is essential for image data quality, reproducibility, and sharing value, but also should be easy to obtain, as described in more detail in an accompanying manuscript<sup>6</sup>.

### Importance and potential pitfalls of standardization

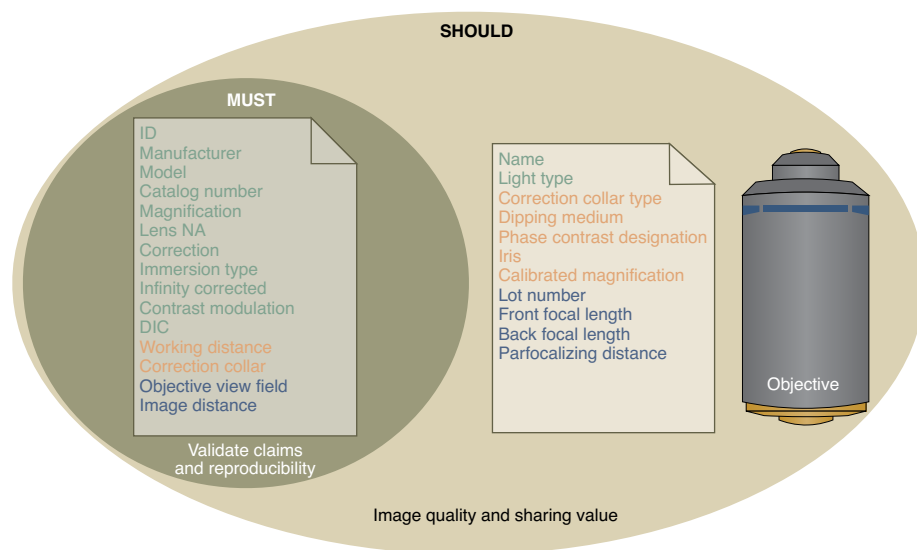
The value of microscopy image data standards (Fig. 1) has been widely recognized, resulting in important efforts to establish best performance testing and instrument calibration practices<sup>43–48</sup>, to unify data-submission requirements from journals<sup>49–52</sup> and to produce the exchange format between image data and metadata that forms the basis for this work<sup>8,9,30,36,37,53</sup>.

Nonetheless, existing efforts have not yet reached normative value, primarily because of the insufficiency of essential elements that are key components of this endeavor, 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<sup>3,11–13</sup>; (2) *software tools* that make prescribed microscopy metadata models actionable by microscope manufacturers, custodians and users





**Fig. 4 | The 4DN-BINA Light Microscopy Metadata Specifications extend the core of the OME Data Model.** Portrayed are Venn diagrams containing a linked-open-data (LOD; Fig. 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 the 4DN-BINA-OME (NBO) namespace (maroon, gray and green ovals). Specifically, the schema illustrates the relationship between the <Instrument> (a) and <Image> (b) elements (OME: Instrument, OME: Image; red-bordered blue ovals) and their subelements belonging to the core of the OME Data Model (light blue set containing blue ovals; for example, OME: Filter, OME: Channel, etc.), with subelements specified by the 4DN-BINA-OME extensions Calibration and Performance (light red set containing maroon ovals; for example, NBOC: LightSensor, NBOC: OpticalCalibration, etc.), Basic (light gray set containing dark gray ovals; for example, NBOB: OpticalAperture, NBOB: CameraSettings, etc.), and Advanced and Confocal (light green set containing green ovals; for example, NBOA: SpinningDisk, NBOA: PinHoleSettings, etc.). The schema is not intended to be comprehensive and includes only a small subset of the elements that comprise the model. AOBS, acousto-optical beam splitter; AOTF, acousto-optical tunable filter; LCTF, liquid crystal tunable filter; TIRF, total internal reflection.



**Fig. 5 | The third axis of the 4DN-BINA-OME Microscopy Metadata Specifications adds further flexibility to minimize the imaging experiment documentation burden.** Depicted is an example Venn diagram representing attributes that are required to document the characteristics of an objective lens and are stored in the <Objective> element of the 4DN-BINA-OME Core and the Basic extension. In the schema, objective attributes are color-coded based on their tier level and are subdivided into requirement-level categories based on the following criteria: (1) *required (MUST)* fields are necessary to validate claims and reproducibility; (2) *recommended (SHOULD)* fields are prescribed to ensure maximal image quality and sharing value. Color-coding is consistent with that used throughout the manuscript: green, Tier 1; orange, Tier 2; dark blue, Tier 3.

faced with the challenge of producing well-documented, high-quality, reproducible and reusable datasets, such as the ones being developed in complementary efforts (OMERO.mde, Micro-Meta App, Methods)<sup>2</sup><sup>15–20</sup>, and (3) *available endpoints* (i.e., deposition to image data repositories; data reuse pipelines) making the purpose and worth of good documentation clear to all members of the community<sup>53–56</sup>. 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 because they are perceived as too expensive, complicated and cumbersome.

There is thus much to be gained from harmonizing the reporting standards in light microscopy. First, this would facilitate the documentation of any microscopy-based protocol, minimize error and quantify residual uncertainty associated with each step of the procedure (Fig. 2). This, in turn, would provide a wealth of valuable contextual information—collectively referred to as data provenance<sup>10</sup>—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 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 (for example, the Image Data Resource, IDR<sup>53</sup>; Electron Microscopy Public Image Archive, EMPIAR<sup>57</sup>; Bioimage Archive<sup>54</sup>; the Cochin Image Database<sup>58</sup>; the NIH (National Institutes of Health) Cell Image Library<sup>59</sup>; the RIKEN Systems Science of Biological Dynamics database, SSBD<sup>60</sup>), will undoubtedly increase the need for community-wide documentation and quality-control standards that can adapt to new technologies. As a first step in this direction<sup>56</sup>, the Recommended Metadata for Biological Images (REMBI)<sup>55</sup> guidelines were recently proposed; these would maximize the possibility of making bioimaging datasets available to other researchers in a timely manner, consistent with FAIR principles<sup>15,20,28</sup> and thus amenable to 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 have 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 its modernization (see Box 1), the OME Data Model<sup>8,9</sup> provides a robust foundation for microscopy metadata (Fig. 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<sup>20,33–35,40,42</sup> become more commonly used, they can be incorporated into the core model through community announcements and related vetting processes to ensure that they meet expanding community needs.

### A three-dimensional matrix of 4DN-BINA-OME Microscopy Metadata Specifications

Given that a one-size-fits-all solution for microscopy metadata requirements is clearly not tenable, here we propose the 4DN-BINA-OME framework (available as described in Table 1), in which microscopy documentation and quality-control requirements are organized along three *orthogonal axes* that are largely independent from each other (Fig. 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 (Fig. 1b, guideline tiers; Fig. 3, Table 2 and Supplementary Table 1). The second axis starts with the OME Data Model<sup>8,9</sup> and extends it with additional metadata components that are introduced on the basis of microscopic modality (for example, epifluorescence vs. confocal microscopy) and accommodate expansion as new technologies are developed that are covered neither by the core nor by the currently proposed extensions (Fig. 1b, OME core vs. extensions; Fig. 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*; Fig. 1b,

### Box 1 | Charting a solid path toward next-generation storage mechanisms for community-driven, OME-based microscopy image data standards

Microscopy metadata stored (Fig. 1; *HOW*, yellow bubble, and *WHERE*, blue bubble) according to the OME Data Model<sup>8,9</sup> are currently 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, although it has other advantages, XSD is not ideally suited to enable the OME Data Model to serve as the foundation for the community development and maintenance of globally accepted light microscopy standards (Fig. 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 to ensure that legacy documents can 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 (Fig. 1, pink bubble) and, in particular, on the development of shared mechanisms defining *HOW* new types of (meta)data have to be recorded (Fig. 1, yellow bubble) and associated with a given Image data file format (Fig. 1, *WHERE*, blue bubble)<sup>36</sup>.

In this context, the OME consortium, in collaboration with RIKEN, has started experimenting with the use of 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<sup>33–35</sup> to be incorporated in the Next-Generation File Formats (NGFF) currently being developed by the OME consortium<sup>36,37</sup>. Employing this RDF/OWL based method would make it possible for users to produce, find and access quality-controlled image data for re-analysis and integration and would 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<sup>15,20,34</sup>. In turn, this will provide a method for different communities to collectively develop a complete picture (Fig. 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<sup>33</sup> and applied to the modeling of specifications proposed here for the exchange of image data and integration with genomics datasets<sup>35</sup>. 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 (Fig. 1).

metadata-requirement levels; Fig. 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 3 lists where some representative experiments

would fall within the microscopy metadata guideline matrix (Fig. 1b).

It should be noted that the 4DN-BINA-OME and REMBI<sup>55</sup> 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<sup>55</sup>. Hence, the adoption of 4DN-BINA-OME Specifications (especially through the use of the complementary software tools being simultaneously presented in related manuscripts)<sup>15–20</sup> would greatly facilitate the work of microscopists who want to deposit imaging data on BioImage Archive<sup>54</sup>.

#### 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 not only is appropriate but also minimizes the burden of collecting metadata for each experiment while maximizing the opportunities for rigor, reproducibility, evaluation, analysis and comparison. We envision a flexible system in which different imaging communities (that is, 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 (Fig. 3, Table 2, Supplementary Table 1 and Supplementary Information, Supplementary 4DN-BINA-OME Tier-system description - Axis 1)<sup>61</sup> 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<sup>11–13</sup> should help to ensure that new metadata specifications are agreed upon by the community and consistent with existing standards.

**Table 1 | Public availability of the NBO Microscopy Metadata Specifications**

General information		
Project main page	<a href="https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs">https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs</a> ( <a href="https://doi.org/10.5281/zenodo.4710731">https://doi.org/10.5281/zenodo.4710731</a> )	
Schema language	XML Schema Definition (XSD), entity-relationship diagram	
License	GNU Lesser General Public License, version 3	
Metadata model schema documents		
GitHub repository	<a href="https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Model">https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Model</a>	
Excel spreadsheet	<a href="https://doi.org/10.5281/zenodo.4711229">https://doi.org/10.5281/zenodo.4711229</a>	
Entity-relationship diagrams	<a href="https://zenodo.org/record/4711229">https://zenodo.org/record/4711229</a>	
XSD Schema	<a href="https://zenodo.org/record/4711426">https://zenodo.org/record/4711426</a>	
Tier-system documents		
GitHub repository	<a href="https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Tier%20System">https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Tier%20System</a>	
Instructions for providing feedback		
GitHub repository issue page	<a href="https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/issues">https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/issues</a>	
Video tutorials	Video 1: background	<a href="https://vimeo.com/624971871">https://vimeo.com/624971871</a>
	Video 2: overview of the model framework and introduction to Micro-Meta App	<a href="https://vimeo.com/624971915">https://vimeo.com/624971915</a>
	Video 3: details of the model structure	<a href="https://vimeo.com/624971980">https://vimeo.com/624971980</a>
	Video 4: description of the GitHub repository and instructions on how to provide feedback	<a href="https://vimeo.com/624995861">https://vimeo.com/624995861</a>

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. Toward this goal, documents describing the 4DN-BINA-OME Microscopy Metadata Specifications are publicly available, as described in this table.

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 (Tier 1) and two analytical tiers (Tiers 2 and 3; Fig. 3, Table 2, Supplementary Table 1 and Supplementary Information, Supplementary 4DN-BINA-OME Tier-system description - Axis 1)<sup>61</sup>, 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 (for example, subpixel 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. Are 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 (although more comprehensive information is always allowed and encouraged). As an example, the proposed specifications encompass information about the sample that directly affects the imaging conditions (for example, labeling method, mounting medium). However, because of the complexity of fully describing experimental and sample preparation procedures, such endeavors pertain more directly to the communities involved in the different research areas that utilize microscopy as an investigation method (cell biology, developmental biology, etc.) and are clearly beyond the scope of this effort. Although 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<sup>61</sup> is available in Supplementary Table 1 and Supplementary Information, Supplementary 4DN-BINA-OME Tier-system description - Axis 1.

**The second axis: a suite of three 4DN-BINA-sponsored community-driven OME extensions.** In its simplest form, metadata can be easily represented as lists

of key-value pairs, in which 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 make it possible to capture the often complex relationships between different real-world components and situations. A better approach 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 (Fig. 6a). A useful formalism for developing, describing and viewing an appropriate data model is the entity-relationship (ER) diagram<sup>62</sup>, which must subsequently be translated into formalized schemas and file formats (Fig. 6b) to facilitate the implementation of metadata capture and management tools.

Because of its status as the only existing exchange format for imaging experiments, the robustness of its design and its solid path forward toward modernization (Box 1; Fig. 2b; details in Supplementary Information, Supplementary OME Model Description), the OME Data Model (i.e., OME Core) represents the ideal starting point for the suite of 4DN-BINA extensions presented here (Fig. 4 and Supplementary



**Table 2 | Tiers for light microscopy metadata reporting as proposed by the IWG of the 4DN initiative and by the QC-DM-WG of BINA**

Category	Nr.	Name	Description	Examples
<b>Descriptive</b>	<b>1</b>	Minimum Information/ Qualitative or Basic Quantification/ Material & Methods	Developmental and stem-cell 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 and localization of markers in cellular subcompartments	Histochemistry, immunohistochemistry, fluorescence in situ hybridization (FISH), immunofluorescence, fluorescent protein labeling
<b>Analytical</b>	<b>2</b>	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, colocalization studies, detection of low-signal features, advanced processing, cell tracking and single-particle tracking, dynamic expression level quantification
	<b>3</b>	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 setup 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.

Figs. 1–4). As such, the 4DN-BINA-OME specifications proposal consists of three extensions of the OME Core<sup>8,9</sup>, each incorporating the concept of graded documentation requirements based on a tiered system of guidelines (Fig. 3, Table 2, Supplementary Table 1 and Supplementary 4DN-BINA-OME Tier-system description - Axis 1). A first step toward this goal, the 4DN-BINA-OME Microscopy Metadata Specifications<sup>14,63,64</sup>, extend the core OME elements <Instrument> and <Image> (Figs. 2b and 4 and Supplementary Figs. 1–4) to reflect the technological advances and the quality control requirements associated with state-of-the-art transmitted light, wide-field and confocal-fluorescence microscopy. A detailed description of the system of three proposed 4DN-BINA-OME extensions is available in the Supplementary Information, Supplementary 4DN-BINA-OME Extension system description - Axis 2. In summary:

1. The *Basic extension* is designed to better capture the technical complexity of transmitted light microscopy and wide-field fluorescence, including subpixel single-particle localization and single-particle tracking experiments (Fig. 4, blue and gray elements; Supplementary Figs. 1 and 2).
2. The *Advanced and Confocal extension* is designed to better capture experiments requiring tunable optics and confocal microscopy (Fig. 4, green elements; Supplementary Figs. 1 and 3).
3. The *Calibration and Performance extension* introduces specifications for the capture of metrics required for microscope calibration and quantitative

instrument performance assessment (Fig. 4, maroon elements; Supplementary Figs. 1 and 4).

Although 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 (such as light-sheet and Airy scan confocal microscopy).

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<sup>14</sup> in three formats (Table 1): (1) a set of graphical ER schemas to facilitate an overall understanding of the model structure<sup>64</sup>; (2) an Excel spreadsheet to express the details of the model in a human-readable form<sup>64</sup>; and finally (3) an XML Schema Definition (XSD) to represent the model schema in a machine-readable manner<sup>63</sup>.

#### The third axis: metadata requirement levels.

Along the third axis (Fig. 5), individual metadata fields are classified based on requirement level as described by Internet Engineering Task Force Request for Comment (RFC) document 2119, produced by the Network Working Group<sup>65</sup>. 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* mean that although 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*<sup>28</sup>. Two examples of the use of the third dimension to add flexibility to the proposed 4DN-BINA-OME Microscopy Metadata Specifications are presented below.

1. Example 1: OME Core and 4DN-BINA Basic extension element <Objective> (Fig. 5)

Whereas the Manufacturer, Model, Magnification and Numerical Aperture (Lens NA) of an objective are required to make it possible 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.

2. 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 bead preparation alongside the Diameter of the beads are essential for the interpretation of the

**Table 3 | Examples of utilization of the three-axis matrix of the 4DN-BINA-OME Microscopy Metadata Specifications**

No.	Description	Sample	Res.	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 wide-field 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 wide-field 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	Wide-field microscope manufacturers	Any	Any	Any	Any	3	C + B + CP	ALL

Res., resolution; C, OME Core; B, 4DN-BINA-OME Basic extension; AC, 4DN-BINA-OME Advanced and Confocal extension; CP, 4DN-BINA-OME Calibration and Performance extension.

calibration results and for reproducibility. However, the beads' Type and Material may be omitted because it can be argued that although that information improves the completeness of the data, it is not absolutely required for the correct interpretation of the results of the optical calibration procedure in which the beads are used.

#### Model implementation: recommendations for Materials and Methods descriptions.

A recent exploration of the quality of published Methods 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 lacking important basic details<sup>22</sup>. Yet 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<sup>23–25,50–52</sup>. In this context, the 4DN-BINA-OME Microscopy Metadata Specifications presented are intended to provide a major contribution toward 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 parallel proposals<sup>50,52</sup>, we propose that microscopy metadata appropriate for Tier 1 (both Must and Should fields) should always be included in the Materials and Methods section of any journal publication to meet minimal rigor and reproducibility criteria<sup>22</sup>. As such, the generalized and

automated availability of Tier 1 metadata, such as provided by the MethodsJ2 app<sup>19,17</sup>, 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.

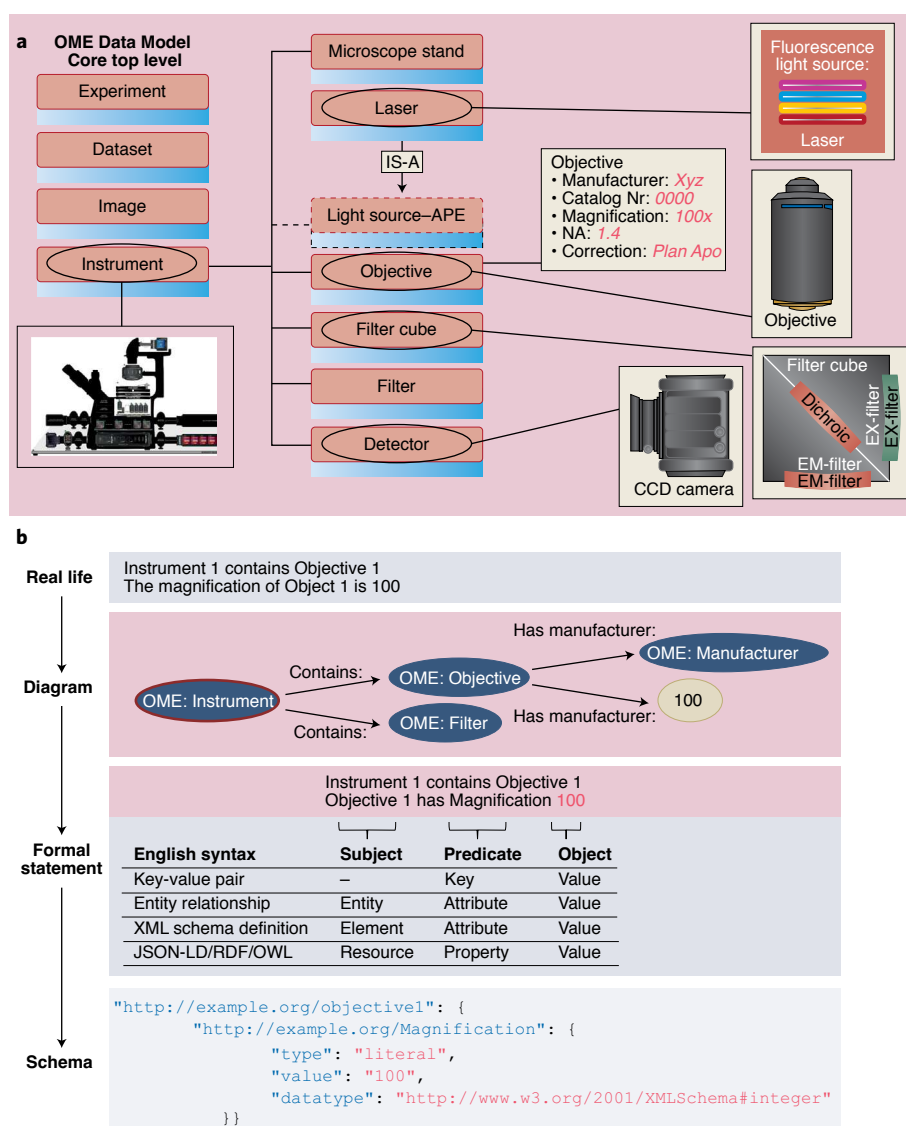
#### Model implementation: facilitated metadata collection.

The importance of rich metadata to ensuring the quality, reproducibility, and scientific and sharing value of image data cannot be overstated. However, collecting 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<sup>15,20</sup> 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 their testing and validation before incorporation in community-accepted standards. (2) The Micro-Meta App<sup>16,18</sup> 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, the Micro-Meta App is particularly suited to enabling imaging scientists to enter all microscope metadata and use the tool to teach trainees about microscopy and the importance of microscopy metadata and to train microscope users in imaging facilities. (3) Finally, MethodsJ2<sup>17,19</sup> focuses on automating the process of writing Methods and Acknowledgements sections that are compliant with microscopy metadata guidelines for scientific publications involving microscopy experiments. MethodsJ2, by design, operates in concert to automatically import microscopy metadata from image data files using BioFormats and from the Micro-Meta App<sup>16,18</sup>.

#### 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



**Fig. 6 | A data model is a schematic representation of reality that can be used 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 entities. Presented is a simplified *entity-relationship* (ER)<sup>62</sup> 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 (for example, <Light Source>, <Objective>, etc.); dashed-line boxes denote generalized element-families to which specific ‘child’ elements belong (i.e., a Laser belongs to the Light Source family). (2) Lists of attributes (*key-value pairs*, enclosed in a beige box) represent metadata that need to be recorded about each hardware component (for example, Magnification, Numerical Aperture, etc.). (3) Lines are used 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,** Depiction of the process, starting from human-readable statements describing the data (*Real life*, top panel), that is often used 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 (for example, *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 example depicted (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.

baseline metadata requirements for any light microscopy experiment to be interpretable, used and shared for scientific purposes<sup>54,55</sup>. 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 timepoints, pixel size, fluorophore name, emission and excitation wavelength, etc.).

## Conclusions

Light microscopy images need to be accompanied by thorough documentation of the microscope hardware and imaging settings used in creating them to ensure correct interpretation of the results. A significant challenge to the reproducibility of microscopy results and their integration with other data types, such as chromatin folding maps generated by the 4DN consortium<sup>1,2,66</sup>, lies in the lack of standardized reporting guidelines for microscopy experiments, as well as instrument performance and calibration benchmarks<sup>22–24,51</sup>. Despite a growing consensus that such standards for light microscopy are desirable, previous efforts to develop shared microscopy data models and application programming interfaces<sup>8,9,30</sup> have not yet succeeded in establishing universal sets 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 maximizing usability via the introduction of a tiered system of documentation requirements; on an expandable suite of model extensions, which includes the first available data model for light-microscopy quality-control metadata; and on the 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. Although 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<sup>67</sup>. 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<sup>1,2</sup> IWG were evaluated and revised by

the BINA QC-DM-WG<sup>3</sup>, resulting in the current proposal. Because it is inherently impossible to predict all future changes that might occur in the field of light microscopy field 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<sup>15,20,28</sup>. This is particularly important in the face of the establishment of a growing number of public image data resources<sup>56</sup> such as the European IDR<sup>53</sup>, EMPIAR<sup>57</sup>, Movincell<sup>68</sup> and Bioimage Archive<sup>54</sup>; the Japanese SSBD hosted by RIKEN<sup>69</sup>; and, in the US, the Allen Cell Explorer<sup>69</sup>, the Human Cell Atlas<sup>70</sup> and the NIH-funded Cell Image Library<sup>59</sup>, Human BioMolecular Atlas Program (HuBMAP)<sup>71</sup> and BRAIN initiative imaging resources<sup>72</sup>. These resources offer the opportunity to emulate, for light microscopy, the successful path that has led to community standards in the field of genomics<sup>39,73–75</sup>.


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<sup>11–13</sup> 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 4DN-BINA-OME Microscopy Metadata Specifications is being coordinated with other parallel initiatives, including the following.

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<sup>1,2</sup>.
2. The OME community development of general criteria and procedures to capture and store metadata in OME-NGFF (Box 1). The OME NGFF effort<sup>36,37</sup> is implementing storage approaches to hold the binary pixel data and the metadata described herein in standardized, shareable, long-lived, efficient and performant containers (for example, files).
3. The EMBL-EBI development of the REMBI recommendations for metadata to be included with imaging datasets deposited to BioImage Archive<sup>54,55</sup>.
4. The development of the International Standards Organization (ISO) 23494-1 standard that will include the 4DN-BINA-OME (NBO namespace) Microscopy Metadata Specifications as part of a provenance information model for biological material and data<sup>31,32</sup>.
5. The development of online educational material, workshops and in-person courses in the context of BINA and in collaboration with Global BioImaging<sup>76</sup> and other community partners<sup>77,78</sup>.

The specific purpose of this multipronged 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 reuse 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<sup>15–20</sup>; and (3) engage all major stakeholders (including those in the commercial, government and academic worlds) in our effort toward community-driven metadata standards for light microscopy. Initially, this mechanism will be used to generate a wider consensus around the current framework and lead toward the development of true community standards. A similar approach 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 (for example, light-sheet and Airy scan confocal microscopy) and synthetic image frameworks (for example, predictive multichannel 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<sup>15–20</sup>, the flexible model framework presented here will provide a significant step forward toward the establishment of robust and future-proof metadata standards for light microscopy. With its key partnerships 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. 

Mathias Hammer<sup>1,2,18</sup>, Maximiliaan Huisman<sup>1,18</sup>, Alessandro Rigano<sup>3,18</sup>, Ulrike Boehm<sup>1,4</sup>, James J. Chambers<sup>5</sup>, Nathalie Gaudreault<sup>1,6</sup>, Alison J. North<sup>7</sup>, Jaime A. Pimentel<sup>1,8</sup>, Damir Sudar<sup>1,9</sup>, Peter Bajcsy<sup>10</sup>, Claire M. Brown<sup>1,11</sup>, Alexander D. Corbett<sup>1,12</sup>, Orestis Faklaris<sup>1,13</sup>, Judith Lacoste<sup>1,14</sup>, Alex Laude<sup>15</sup>, Glyn Nelson<sup>1,15</sup>, Roland Nitschke<sup>1,16</sup>, Farzin Farzam<sup>1</sup>, Carlos S. Smith<sup>1,17</sup>, David Grunwald<sup>1,18</sup> and Caterina Strambio-De-Castillia<sup>1,18</sup> 

<sup>1</sup>RNA Therapeutics Institute, UMass Chan Medical School, Worcester, MA, USA. <sup>2</sup>Department of Biology, Technical University of Darmstadt, Darmstadt, Germany. <sup>3</sup>Program in Molecular Medicine, UMass Chan Medical School, Worcester, MA, USA. <sup>4</sup>Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. <sup>5</sup>Institute for Applied Life Sciences, University of Massachusetts, Amherst, MA, USA. <sup>6</sup>Allen Institute for Cell Science, Seattle, WA, USA. <sup>7</sup>The Rockefeller University, New York, NY, USA. <sup>8</sup>Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México. <sup>9</sup>Quantitative Imaging Systems LLC, Portland, OR, USA. <sup>10</sup>National Institute of Standards and Technology, Gaithersburg, MD, USA. <sup>11</sup>Advanced BioImaging Facility (ABIF), McGill University, Montreal, Quebec, Canada. <sup>12</sup>Department of Physics and Astronomy, University of Exeter, Exeter, UK. <sup>13</sup>MRI, BCM, University of Montpellier, CNRS, INSERM, Montpellier, France. <sup>14</sup>MIA Cellavie Inc., Montreal, Quebec, Canada. <sup>15</sup>Bioimaging Unit, Newcastle University, Newcastle upon Tyne, UK. <sup>16</sup>Life Imaging Center and Signalling Research Centres CIBSS and BIOS, University of Freiburg, Freiburg, Germany. <sup>17</sup>Delft Center for Systems and Control and Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands. <sup>18</sup>These authors contributed equally: Mathias Hammer, Maximiliaan Huisman, Alessandro Rigano.

✉e-mail: [caterina.strambio@umassmed.edu](mailto:caterina.strambio@umassmed.edu)

Published online: 3 December 2021  
<https://doi.org/10.1038/s41592-021-01327-9>

## References

- Dekker, J. et al. *Nature* **549**, 219–226 (2017).
- 4D Nucleome Consortium. The 4D Nucleome Web Portal. <https://www.4dnucleome.org/> (2017).
- 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).
- BioImaging North America. <https://www.bioimagingna.org> (2018).
- Hammer, M. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.31.446382> (2021).
- Huisman, M. et al. Preprint at <https://arxiv.org/abs/1910.11370> (2021).
- Huisman, M. et al. Preprint at <https://arxiv.org/abs/1910.11370v3> (2019).
- Goldberg, I. G. et al. *Genome Biol.* **6**, R47 (2005).
- Linkert, M. et al. *J. Cell Biol.* **189**, 777–782 (2010).
- Ram, S. & Liu, J. *J. Data Semant.* **1**, 11–17 (2012).
- Boehm, U. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01162-y> (2021).
- Nelson, G. et al. *J. Microsc.* <https://doi.org/10.1111/jmi.13041> (2021).
- Nelson, G. et al. Preprint at <http://arxiv.org/abs/2101.09153> (2021).
- Rigano, A. et al. 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01. <https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs> (2021); <https://doi.org/10.5281/zenodo.4710731>
- Kunis, S. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01288-z> (2021).
- Rigano, A. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01315-z> (2021).
- Ryan, J. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01290-5> (2021).
- Rigano, A. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.31.446382> (2021).
- Ryan, J. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.06.23.449674> (2021).
- Kunis, S. et al. Preprint at <http://arxiv.org/abs/2103.02942> (2021).
- Baker, M. *Nature* **533**, 452–454 (2016).
- Marqués, G., Pengo, T. & Sanders, M. A. *eLife* **9**, e55133 (2020).
- Sheen, M. R. et al. *eLife* **8**, e45120 (2019).
- Viana, M. P. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.08.415562> (2021).
- Botvinnik-Nezer, R. et al. *Nature* **582**, 84–88 (2020).
- Nature Editorial Staff. *Nat. Methods* **15**, 395 (2018).
- Pines, J. *Open Biol.* **10**, 200165 (2020).
- Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
- Allan, C. et al. *Nat. Methods* **9**, 245–253 (2012).
- Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K. *Science* **300**, 100–102 (2003).
- Wittner, R. et al. in *Provenance and Annotation of Data and Processes. International Provenance and Annotation Workshop (IPAW) 2020–2021* (eds. Glavic, B., Braganholo, V. & Koop, D.) 222–225 (Springer, 2021).
- Holub, P. et al. Preprint at *Zenodo* <https://doi.org/10.5281/zenodo.5093125> (2021).
- Kobayashi, N. et al. in *Proceedings of the 12th Annual SWAT4(HC) LS (Semantic Web Applications and Tools for Healthcare and Life Sciences) Conference* (CEUR Workshop Proceedings, vol. 2849) (eds. Cornet, R. & Waagmeester, A.) 149–150 (2019); <http://ceur-ws.org/Vol-2849/paper-25.pdf>
- Moore, J. et al. in *Proceedings of the 12th SWAT4(HC)LS Conference* (CEUR Workshop Proceedings, vol. 2849) (eds. Cornet, R. & Waagmeester, A.) 44–53 (2019); <http://ceur-ws.org/Vol-2849/paper-06.pdf>
- Hammer, M. et al. in *Proceedings of the 12th Annual SWAT4(HC) LS Conference* (CEUR Workshop Proceedings, vol. 2849) (eds. Cornet, R. & Waagmeester, A.) 153–154 (2019); <http://ceur-ws.org/Vol-2849/paper-27.pdf>
- Moore, J. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.31.437929> (2021).
- Moore, J. et al. OME-NGFF: a next-generation file format for expanding bioimaging data access strategies. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01326-w> (in the press).
- Deutscher, E. W. et al. *Nat. Biotechnol.* **26**, 305–312 (2008).
- Brazma, A. *ScientificWorldJournal* **9**, 420–423 (2009).
- Rigano, A. & Strambio-De-Castillia, C. Preprint at *bioRxiv* <https://doi.org/10.1101/155036> (2017).
- Miura, K. & Nørrelykke, S. F. *EMBO J.* **40**, e105889 (2021).
- Masuzzo, P. Minimum Information About Cell Migration Experiments (MIACME 0.2) (2021); <https://cmso.science/MIACME/v0.2/>
- North, A. J. *J. Cell Biol.* **172**, 9–18 (2006).
- Theer, P., Mongis, C. & Knop, M. *Nat. Methods* **11**, 981–982 (2014).
- Lambert, T. J. & Waters, J. C. *Methods Cell Biol.* **123**, 35–53 (2014).
- Deagle, R. C., Wee, T. E. & Brown, C. M. *Int. J. Biochem. Cell Biol.* **89**, 120–124 (2017).
- Mubaid, F. et al. *Histochem. Cell Biol.* **151**, 357–366 (2019).
- Kiepas, A., Voorand, E., Mubaid, F., Siegel, P. M. & Brown, C. M. *J. Cell Sci.* **133**, jcs242834 (2020).
- Hill, E. J. *J. Cell Biol.* **183**, 969–970 (2008).
- Montero Llopis, P. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01156-w> (2021).
- Heddeston, J. M., Aaron, J. S., Khuon, S. & Chew, T.-L. *J. Cell Sci.* **134**, jcs254144 (2021).
- Aaron, J. S. & Chew, T.-L. *J. Cell Sci.* **134**, jcs254144 (2021).
- Williams, E. et al. *Nat. Methods* **14**, 775–781 (2017).
- Ellenberg, J. et al. *Nat. Methods* **15**, 849–854 (2018).
- Sarkans, U. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01166-8> (2021).
- Swedlow, J. R. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01113-7> (2021).
- Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. *Nat. Methods* **13**, 387–388 (2016).
- Guilbert, T. Cochin Image Database (2019); <https://imagerie.cochin.inserm.fr/sis4web/login.php>
- Orloff, D. N., Iwasa, J. H., Martone, M. E., Ellisman, M. H. & Kane, C. M. *Nucleic Acids Res.* **41**, D1241–D1250 (2013).
- Tohsato, Y., Ho, K. H. L., Kyoda, K. & Onami, S. *Bioinformatics* **32**, 3471–3479 (2016).
- Hammer, M. et al. 4DN-BINA-OME (NBO) Microscopy Metadata Specs - Tiers System - v2.01. <https://github.com/WU-BIMAC/NBOMicroscopyMetadataSpecs/tree/master/Tier%20System/stable%20version/v02-01> (2021); <https://doi.org/10.5281/zenodo.4710731>
- Chen, P. P.-S. *ACM Trans. Database Syst.* **1**, 9–36 (1976).
- Rigano, A. & Strambio-De-Castillia, C. 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XSD schema. *Zenodo* <https://doi.org/10.5281/zenodo.4711426> (2021).
- Rigano, A. et al. 4DN-BINA-OME (NBO) Tiered Microscopy Metadata Specifications - v2.01 - XLS Spreadsheet and Entity Relationship schemas. *Zenodo* <https://doi.org/10.5281/zenodo.4711229> (2021).
- Bradner, S. Key words for use in RFCs to indicate requirement levels [RFC2119] (1997); <http://www.ietf.org/rfc/rfc2119.txt>
- Reiff, S. B. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.10.14.464435> (2021).
- de Souza, N. *Nat. Methods* **9**, 1046 (2012).
- Movincell Consortium. Multi-dimensional marine organism dataview. *Movincell* (2015); <http://movincell.org/>
- Allen Institute for Cell Science. Allen Cell Explorer (2017); <https://www.allencell.org/>
- Rozenblatt-Rosen, O. et al. *Nat. Biotechnol.* **39**, 149–153 (2021).
- HuBMAP Consortium. The Human BioMolecular Atlas Program - HuBMAP (2017); <https://hubmapconsortium.org/>
- DORY Working Group. Defining Our Research Methodology (2019); <https://doryworkspace.org/>
- Sansone, S.-A. et al. *OMICS* **12**, 143–149 (2008).
- Ioannidis, J. P. A. et al. *Nat. Genet.* **41**, 149–155 (2009).
- Rung, J. & Brazma, A. *Nat. Rev. Genet.* **14**, 89–99 (2013).
- Grebnev, G. & Global BioImaging. Global BioImaging training resource (2021); <https://globalbioimaging.org/international-training-courses/repository>
- Waters, J., Jug, F., Elliott, H. & Lambert, T. Cold Spring Harbor Laboratory courses: Quantitative imaging: from acquisition to analysis. *Cold Spring Harbor Laboratory* (2021); <https://meetings.csh.edu/courses.aspx?course=C-QICM&year=21>
- North, A. & Bewersdorff, J. Marine Biology Laboratory courses: optical microscopy & imaging in the biomedical sciences (OMIBS). *University of Chicago Marine Biological Laboratory* (2021); <https://www.mbl.edu/education/courses/optical-microscopy-imaging/>
- Allen, L., O'Connell, A. & Kiermer, V. *Learn. Publ.* **32**, 71–74 (2019).

## Acknowledgements

We are deeply indebted to Thao P. Do, cell visualization and web specialist at the Allen Institute for Cell Science, for her expert advice and skilled work, which were invaluable for the production of Figs. 1, 2, 3 and 6. We would like to thank Darryl Conte for the critical reading of the manuscript. We would like to thank Kevin Fogarty, Lawrence Lifshitz and Karl Bellve at the Biomedical Imaging Group of the Program in Molecular Medicine at the University of Massachusetts Medical School for invaluable intellectual input and countless fruitful discussions and for their friendship, advice and steadfast support throughout the development of this project. This project could never have been carried out without the leadership, insightful discussions, support and friendship of all OME consortium members, with particular reference to Jason Swedlow, Josh Moore, Chris Allan, Jean Marie Burel and Will Moore. We are massively indebted to the RIKEN community for their fantastic work to bring open science into biology. We would like to particularly acknowledge Norio Kobayashi and Shuichi Onami for their friendship and support.

We thank all members of the 4D Nucleome Imaging Standard Working Group, Data Coordination and Integration Center and NIH leadership team for the vision and encouragement that allowed us to get this project

started. We are particularly grateful to Burak Alver, Joan Ritland, Rob Singer, Warren Zipfel, Ian Fingerman and John Satterlee. We thank all members of BioImaging North America (in particular Lisa Cameron, Michelle Itano and Paula Montero-Llopis), German Bioimaging (in particular Susanne Kunis and Stephanie Wiedkamp-Peters), Euro-Bioimaging (in particular Antje Keppler and Federica Paina) and QUAREP-LiMi (in particular all members of the Working Group 7 - Metadata; quarep.org) for invaluable intellectual input, fruitful discussions, advice and moral support. The authors wish to thank all personnel at their home core facilities and their institutions for unwavering support and material help in the absence of which this work could not have been successfully concluded. In particular, we are profoundly indebted to the Allen Institute's founder, P. G. Allen, for his vision, encouragement and support. We are also indebted to the following individuals for their continued and steadfast support: Jeremy Luban, Michael Czech, Roger Davis and Thoru Pederson at the University of Massachusetts Medical School.

This work was supported by grant no. 2019-198155 (5022) awarded to C.S.D.C. by the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation, as part of their Imaging Scientist Program. Major funding also came from (1) NIH grant no. U01CA200059 to C.S.D.C. and D.G. and (2) NIH grants no. 1U01EB021238 and no. 8U01DA047733-05 and NSF grant no. 1917206 to D.G. A.D.C. was supported by a Royal Society Short Industry Fellowship (no. SIF\R2\202044). C.M.B. was funded in part by grant no. 2020-225398 from the Chan Zuckerberg Initiative DAF. C.S.S. was supported by the Netherlands Organisation for Scientific Research (NWO), under NWO START-UP project no. 740.018.015 and NWO Veni project no. 16761. D.S. was funded in part by NIH/National Cancer Institute grants U54CA209988 and U2CCA23380. O.F. is a member of the RTmf network, and the MRI core facility is supported by National Infrastructure France BioImaging (grant ANR-10-INBS-04) and IBISA consortium. P.B. was supported by the US National Institute of Standards and Technology (NIST) Scientific

and Technical Research and Services funds. R.N. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant number Ni 451/9-1 MIAP-Freiburg. Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

#### Author contributions

Author contributions categories used here were devised by the CRediT initiative<sup>79</sup>. M. Hammer: conceptualization, methodology, investigation, resources, data curation, writing—review and editing; M. Huisman: conceptualization, methodology, investigation, writing—original draft, writing—review and editing, visualization; A.R.: conceptualization, methodology, software; U.B.: validation, data curation, writing—review and editing; J.J.C.: validation, data curation, writing—review and editing; N.G.: validation, data curation, writing—review and editing, visualization; A.J.N.: validation, writing—review and editing; J.A.P.: validation, data curation, writing—review and editing; D.S.: validation, data curation, writing—review and editing; P.B.: validation, writing—review and editing; C.M.B.: validation, writing—review and editing; A.D.C.: validation, writing—review and editing; O.F.: validation, writing—review and editing; J.L.: validation, writing—review and editing; A.L.: validation, writing—review and editing; G.N.: validation, writing—review and editing; R.N.: validation, writing—review and editing; F.F.: investigation, data curation; C.S.S.: conceptualization, methodology; D.G.: conceptualization, methodology, investigation, resources, writing—original draft, supervision, project administration, funding acquisition; C.S.-D.-C.: conceptualization, methodology, software, validation, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration, funding acquisition.

#### Members of the Bioimaging North America Quality Control and Data Management Working

**Group (BINA-DM-QC-WG):** Ulrike Boehm, James J. Chambers, Nathalie Gaudreault, Alison J. North, Jaime A. Pimentel, Damir Sudar, Peter Bajcsy, Claire M. Brown, Alexander D. Corbett, Orestis Faklaris, Judith Lacoste, Alex Laude, Glyn Nelson, Roland Nitschke and Caterina Strambio-De-Castillia.

#### Members of Quality Assessment and Reproducibility for Instrument and Images in Light Microscopy (QUAREP-LiMi)

Mathias Hammer, Alessandro Rigano, Ulrike Boehm, James J. Chambers, Nathalie Gaudreault, Alison J. North, Jaime A. Pimentel, Damir Sudar, Peter Bajcsy, Claire M. Brown, Alexander D. Corbett, Orestis Faklaris, Judith Lacoste, Alex Laude, Glyn Nelson, Roland Nitschke, David Grunwald and Caterina Strambio-De-Castillia.

#### Members of the 4D Nucleome (4DN) Network:

Mathias Hammer, Maximilian Huisman, Alessandro Rigano, Nathalie Gaudreault, David Grunwald and Caterina Strambio-De-Castillia.

#### Competing interests

A.D.C. provides consultancy for the University of Exeter Consulting Ltd., which commercializes the PSFcheck calibration slide. D.S. is employed by Quantitative Imaging Systems LLC, a commercial developer of imaging software. J.L. is employed by MIA Cellavie Inc., which offers consulting expert services in microscopy, imaging and analysis. The remaining authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

#### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-021-01327-9>.

**Peer review information** *Nature Methods* thanks Kelly Rogers, Rafael Carazo Salas and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.