REMBI: Recommended Metadata for Biological Images

- realizing the full potential of the bioimaging revolution by enabling data reuse

Ugis Sarkans[1], Wah Chiu[2], Lucy Collinson[3], Michele C. Darrow[4], Jan Ellenberg[5], David Grunwald[6], Jean-Karim Hériché[5], Andrii Iudin[1], Gabriel G. Martins[7], Terry Meehan[1], Kedar Narayan[8], Ardan Patwardhan[1], Matthew Robert Geoffrey Russell[3], Helen R Saibil[9], Caterina Strambio-De-Castillia[10], Jason R Swedlow[11], Christian Tischer[5], Virginie Uhlmann[1], Paul Verkade[12], Mary Barlow[1], Omer Bayraktar[13], Ewan Birney[1], Cesare Catavitello[1], Christopher Cawthorne[14], Stephan Wagner-Conrad[15], Elizabeth Duke[16], Perrine Paul-Gilloteaux[17], Emmanuel Gustin[18], Maria Harkiolaki[16], Pasi Kankaanpää[19], Thomas Lemberger[20], Jo McEntyre[1], Josh Moore[11], Andrew W Nicholls[21], Shuichi Onami[22], Helen Parkinson[1], Maddy Parsons[23], Marina Romanchikova[24], Nicholas Sofroniew[25], Jim Swoger[26], Nadine Utz[27], Lenard Voortman[28], Frances Wong[9], Peijun Zhang[29], Gerard J. Kleywegt[1] & Alvis Brazma[1]

Affiliations

1. European Molecular Biology Laboratory, European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, CB10 1SD, UK
2. Stanford University, Stanford, CA
3. Francis Crick Institute, London, NW1 1AT
4. SPT Labtech, Ltd
5. EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
6. RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA, 01605, USA
7. Instituto Gulbenkian de Ciencia Oeiras 2780-156 - Portugal and Faculdade de Ciências, Universidade de Lisboa 1740-016 Lisboa, Portugal
8. Center for Molecular Microscopy, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA and Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA.
9. Dept of Biological Sciences and Institute of Structural and Molecular Biology, Birkbeck, University of London, Malet St, London WC1E 7HX, UK
10. Program in Molecular Medicine, University of Massachusetts Medical School, 373 Plantation Street, Suite 114, Worcester, MA01605.
11. Division for Computational Biology, Centre for Gene Regulation and Expression, University of Dundee, UK.
12. School of Biochemistry, University of Bristol, UK
13. Wellcome Sanger Institute, Cambridge, CB10 1SA, UK
14. Nuclear Medicine and Molecular Imaging, Department of Imaging and Pathology, KU Leuven, 3000 Leuven, Belgium
15. Carl Zeiss Microscopy GmbH
16. Diamond Light Source, Harwell Science and Innovation, Campus, Harwell. Oxfordshire, OX111 0DE
17. Université de Nantes, CNRS, INSERM, l'institut du thorax, SFR Santé, Inserm UMS 016, CNRS UMS 3556, F-44000 Nantes, France.
18. Janssen Pharmaceutica N.V.
19. Turku BioImaging, University of Turku and Åbo Akademi University, Turku, Finland; and Euro-BioImaging ERIC, Turku, Finland

20. EMBO, Meyerhofstrasse 1, 69117 Heidelberg, Germany
21. GlaxoSmithKline R&D, Gunnels Wood Road, Stevenage, Herts, SG1 2NY, UK
22. RIKEN Center for Biosystems Dynamics Research, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan
23. Randall Centre for Cell and Molecular Biophysics, King's College London
24. National Physical Laboratory, Hampton Road, Teddington, United Kingdom
25. Chan Zuckerberg Initiative
26. European Molecular Biology Laboratory (EMBL), 08003 Barcelona, Spain
27. German BioImaging e.V., University of Konstanz, Universitätsstrasse 10, D-78464 Konstanz
28. Department of Cell and Chemical Biology, Leiden University Medical Center LUMC, Leiden, The Netherlands
29. Electron Bio-Imaging Centre, Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK and Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

Spectacular advances in the light and electron microscopy[1, 2] are rapidly transforming the life sciences. For instance, scientists are now able to image molecular complexes at atomic resolution[3], follow the fates of individual molecules in a living cell, and image the development of an organism starting from a single fertilised cell[4, 5]. These imaging technologies are generating large amounts of complex data, the interpretation of which often requires sophisticated analyses, as in other "omics" technologies. Moreover, most advanced imaging technologies are expensive, while the biological samples used in the experiments may be unique. To maximise the use of the generated data and to realise the full potential of the advances in biological imaging, these datasets need to be made available to other researchers in a timely manner, consistent with the FAIR principles (Findable, Accessible, Interoperable and Reusable)[6], and thus amenable for reuse.

Experience from other omics domains has taught us that to make data reusable, some standardisation is necessary, and in particular, in reporting the metadata describing the experiments and the samples. To achieve this, appropriate "minimal" or recommended information guidelines or standards have been adopted by various life-science communities. One of the first such initiatives was MIAME (Minimum Information About a Microarray Experiment), which was published in 2001[7] and has had a major impact on how functional genomics data are collected and reported via public repositories and on the reusability of these data[8, 9]. As the biological imaging field is maturing, this community is now recognizing that it faces similar challenges. In fact, the metadata challenge in the bioimaging domain has been discussed in the ELMI community (https://elmi.embl.org) since 2001 and an attempt to address it was undertaken by the OME Consortium[10]. In the domain of medical imaging the challenge is partially addressed by the Digital Information and Communications in Medicine (DICOM) standard[11]. Nevertheless, it has been reported recently that metadata on imaging methods are vastly underreported in biomedical research[12]. One might argue that microscopy experiments are too complex and heterogeneous to be amenable to a standardised description, yet 20 years ago the same was often said about microarray data, but the best practices for collecting and representing metadata

for various biomedical domains have considerably evolved since then. Arguably, the biological imaging field is ready for some initial data standardisation and would benefit from it.

A workshop held in Hinxton (UK) in 2017 unanimously supported the establishment of a public bioimage archive to store data associated with peer-reviewed publications or systematic imaging projects[13]. The workshop recommended the adoption of initially flexible data standards, which could be gradually tightened, as different imaging communities reach consensus. The EMBL-EBI BioImage Archive (https://www.ebi.ac.uk/bioimage-archive) was established in July 2019 and it provides the community with the means to share different types of imaging data. The BioImage Archive is a deposition database for all microscopy images associated with peer-reviewed scientific publications for which a more specialised resource is not available. It is part of a larger and developing bioimaging "ecosystem", which also includes more specialised and structured image resources, such as EMPIAR for all Electron Microscopy images[14], Cell IDR[15] for curated images of cells and Tissue-IDR for curated images of biological tissues. The BioImage Archive is built on a high-performance, high-volume data-storage system, which can be used as a platform by other existing or emerging biological imaging resources.

A follow-up workshop to discuss minimal metadata recommendations in several biological imaging fields was held in Hinxton in October 2019. Representatives from light and electron microscopy communities exchanged their experiences and ideas and began the process of developing the *Recommended Metadata for Biological Images* (REMBI) guidelines presented here, to address the needs of these communities. A common theme in community efforts such as this is that standardised dataset annotation becomes more complex and time-consuming with every additional metadata element. Thus, attempts to impose requirements that are not yet sufficiently adopted by a given community or supported by relevant data-annotation tools may be counterproductive. However, this challenge is not dissimilar to the one the microarray community faced at the beginning of this century and the arguments presented for and against inclusion of a greater or lower level of detail in the minimum standard are quite similar in both domains. In addition, the amount of information required for reuse may differ depending on the imaging technology, the scientific application and the needs of different user groups (see **Figure 1**). We are thus convinced that there is a need to strike the right balance between minimising the barriers to data submission and maximising opportunities for data reuse.

It has to be taken into account that microscopy technology development is highly dynamic, that there are many existing file formats (with new formats appearing regularly), and that data sets are becoming larger and more complex. Given the enormous heterogeneity of biological imaging methods and the wide range of scales (from sub-nanometre to centimetre scale), the workshop established three working groups to address metadata recommendations for different sub-domains, namely: (1) the Electron cryo-Microscopy (cryo-EM) and Electron cryo-Tomography (cryo-ET) working group; (2) the Volume EM and Correlative Imaging working group; and (3) the Light Microscopy (LM) working group, which covered cell, tissue and organism-level imaging (**see Boxes 1-3**). While these types of imaging each require specific types of metadata, they are all applied to study biological systems, and therefore commonalities are to be expected. The working groups converged on a common high-level structure of the recommended metadata guidelines, as shown in **Figure 2** (for more detail see **Supplementary material**).

The purpose of the proposed guidelines is to provide a framework for discussing different aspects of useful sharing of imaging data with the goal of reaching community-wide consensus on the level of detail that is optimal. The workshop participants agreed that it is important to distinguish minimal metadata requirements from particular data models: the former concern the semantic requirements of what annotation is needed to understand and reuse image data, whereas the latter concern the syntactic representation of these metadata elements by computer software. There is also a third layer, specifying the implementation of a data model in a deposition system for a particular archive, and the user-interface of such a system.

As an agreement on recommended metadata is emerging, data-deposition tools that facilitate collection of these metadata (including submission tools for the BioImage Archive) will be developed, testing these standards in practice. Intelligent software strategies, such as auto-filling common fields and automatic "data-harvesting" of information from log files, should be used to lower the barriers for data upload and to increase the quality of the captured information. Use of some metric(s) to assess the completeness and correctness of uploads may encourage better deposition practices, resulting in wider use of and greater trust in the shared data from the community. Based on the experience gained and the community feedback regarding the practicalities of data submissions and reuse, the standards will need to be kept up-to-date and evolving with the science, technology and practices of bioimaging. However, the ultimate test of this effort will be the extent to which biological imaging data deposited in relevant archives will be reused (see **Figure 3**). The lessons from microarray data show that the earliest mode of data reuse may be related to testing new data-analysis tools, rather than providing novel biological insights, which happened later[9].

The recommended imaging metadata standard described here will be adopted by the BioImage Archive, EMPIAR, Cell-IDR and Tissue-IDR. We hope that other existing and future archives will also adopt REMBI and engage with us to help shape the future development of the standard, in the spirit of the worldwide drive towards FAIR data sharing. To facilitate this, we encourage interested parties to contact us on rembi@ebi.ac.uk. We encourage scientific journals to support (or make mandatory) the deposition of bioimaging data in such FAIR resources, and funders to make data deposition a condition of grant funding. We also hope that instrument manufacturers and software developers will increasingly support recording of the recommended metadata automatically (in agreed formats), thereby minimising any burden on the data submitters and minimising data-entry errors. Finally, we call upon all scientists who use imaging methods in their published work to consider depositing their data and the associated rich metadata in the appropriate archives.

Box 1. Cryo-EM and cryo-ET

In recent years cryo-EM and cryo-ET have proven to be powerful tools for determining high-resolution structures of biological "matter" and examining the functional cellular context of macromolecular complexes. This has been possible due to technical advancements in microscope optics and detectors, sample-preparation techniques such as micropatterning grids and focused-ion-beam milling, and data-analysis pipelines including reliable automation of data-

acquisition and processing workflows. Advances in the field of cryo-EM were recognised in 2017 with the Nobel Prize in Chemistry and the method has continued to advance, now reaching truly atomic resolution[3]. In concert, cryo-ET has matured into a method that is capable of probing 3D cellular context from micrometer to sub-nanometer scales, providing insight into biological processes such as viral infection and disease states. There is wide agreement in the cryo-EM community that detailed metadata has to be recorded and deposited to the public archives, and that metadata standards must be reviewed over time to ensure they are fit-for-purpose and continue to address evolving community needs. In this sense, the cryo-EM community is leading the way for other imaging communities.

Cryo-EM volumes (maps and tomograms) are commonly deposited in the Electron Microscopy Data Bank (EMDB)[16], and any fitted atomic models in the Protein Data Bank (PDB)[17]. The Electron Microscopy Public Image Archive (EMPIAR) was established at EMBL-EBI in 2013[14] and has been the public resource for raw cryo-EM images that underpin the structures in EMDB. EMPIAR provides easy access to state-of-the-art raw data to facilitate methods development and validation, which will ultimately lead to better methods, better structures and a better understanding of biological questions. The EMPIAR metadata schema therefore defines the *de facto* standard for that community (ftp://ftp.ebi.ac.uk/pub/databases/emtest/empiar/schema). It has evolved over time based on feedback from depositors and workshops with community experts. It now supports data from a range of imaging modalities, including soft X-ray tomography and various Volume EM methodologies (**Box 2**). Initially, EMPIAR only accepted raw datasets belonging to maps and tomograms in EMDB, which contains extensive metadata about the experiment (e.g., specimen preparation, microscopy, image processing and validation). Therefore, the EMPIAR data model was designed to be lightweight and capture only information directly pertaining to the image sets (e.g., number of images, image width and height). However, with the growing use of EMPIAR by the cryo-EM community, there have been increasing calls to expand its data model to incorporate more information, e.g., about processing workflows and particle-picking files. As EMPIAR has expanded its remit to include many new imaging modalities, it has been consulting with the relevant communities to expand the EMPIAR data model to capture essential information about the experiment as for Volume EM entries, contrary to cryo-EM ones, such information is not captured in any other archive. This work is on-going and will be further informed by the minimum metadata guidelines presented here.

Box 2. Volume EM and correlative imaging

Volume Electron Microscopy (VEM) is a collective term for techniques that are used to acquire serial electron images through volumes of resin-embedded, heavy-metal-stained cells and tissues. VEM covers semi-automated blockface and array modalities in the scanning electron microscope. In blockface techniques, the exposed surface of the sample, either resin-embedded or frozen-hydrated, is imaged using the electron beam, and then a slice of material is removed from the surface using a diamond knife, an ion beam, plasma milling, or a

femtosecond laser. The imaging and cutting process is repeated sequentially to collect hundreds or thousands of images. In array tomography, serial sections are cut using a diamond knife and collected onto a silicon wafer or conductive glass, and each section is imaged sequentially. VEM also encompasses techniques enabling the collection of volume data from serial sections in a transmission electron microscope, a process that is now becoming more automated. Typical image volumes range from 10^3 μm^3 (a sub-volume of a single cell) to 10^9 μm^3 (small regions of tissues), with connectomics researchers aiming to reach mm^3 volumes at 10 nm voxel resolution in the near future. Even though this combination of volume and resolution was unthinkable just ten years ago, the volumes are minute compared to the size of most organisms, and so a targeted approach is critical to locate the region of interest (ROI) to be imaged, both spatially and temporally. Correlative imaging is often used to target the ROI (correlation for identification) and/or to add functional information to the structural data using macromolecular probes (correlation for colocalisation). Correlative workflows are complex, and are usually custom-designed for each research question, taking into account contrast mechanisms, probes, sample preparation, tissue structure, target structure, field of view and resolution, and may incorporate fluorescence, X-ray, elemental or chemical imaging modalities.

The aim of the VEM and Correlative Imaging working group was to create a sufficiently flexible framework to accommodate these heterogeneous image datasets while retaining enough structure to enable FAIR practices for broader data mining in the future. For example, biosamples entering correlative experimental pipelines necessarily diverge at one or more of the sample preparation, image acquisition, or image processing steps. The accompanying metadata must accurately and sufficiently describe each step and, where required, include transformation information to correctly connect disparate image data in 2D or 3D, in appropriate contexts.

Box 3. Light microscopy

Light microscopy covers a wide range of imaging modalities spanning several temporal and spatial scales generated by a wide range of technologies, including single-molecule localisation microscopy, widefield or confocal microscopy, optical projection tomography, and light-sheet microscopy. Considering the vast number of analysis methods and types of experiments in this subdomain, defining an overarching standard at a high level of detail turned out to be challenging. It was noted that several groups are already actively working on defining best practices and standards for these types of imaging data (e.g., Euro-BioImaging, Global-Bioimaging, BioImaging North America, the 4D Nucleome Program, and the recently launched Quality Assessment and Reproducibility in Light Microscopy initiative). The current plethora of experimental set-ups (e.g., high-content screening, light-sheet microscopy, and digital pathology), file formats, compression methods, and the increasing complexity of data sets, are all complicating factors. Acknowledging the fact that this subdomain produces datasets to address an extremely wide range of research questions, the working group concluded that it is currently difficult to expand the obligatory metadata required for archival deposition beyond the

basic information needed to open a dataset and access the pixel data such that visualisation and/or re-analysis is possible. While it does not immediately ensure full experimental reproducibility or provide a biological understanding of the sample, imaging conditions or other contextual information, this approach can serve as a starting point. The standard will of course evolve gradually and be subject to refinement by the community, as standardisation progresses in the field. To some extent the drawbacks can be mitigated by implementing minimal criteria in a way that encourages submission of additional structured metadata in the archive-submission systems. Better documented datasets will therefore benefit from enhanced reusability and gain broader visibility. We also hope that discussions in the light-microscopy community will continue, eventually producing consensus on minimal metadata that allows for experimental reproducibility and is fully consistent with the FAIR principles.

Supplementary material

The current version of REMBI, including examples from the fields covered by the three working groups, is available from http://bit.ly/rembi-v1

Acknowledgments

The workshop was hosted and funded by EMBL-EBI. We are grateful to Jill Christiaens, Rebecca Sherry and Christina Karikides for logistical and administrative support. We thank the workshop participants Clement Lore, Olaf Selchow, Sebastian Tille and Kristian Wadel for their valuable contributions to the discussion. Figures 1 and 2 were created by Spencer Phillips (EMBL-EBI). Finally, we would like to thank Anna Reed for help with the preparation of the manuscript. Travel was funded by the individual participants. D.G. was supported by NIH 8U01DA047733-05 and NSF 1917206. L.C. was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999). P.P.-G. was supported by CROCOVAL (ANR-18-CE45-0015) and is part of the national infrastructure “France BioImaging” supported by the ANR PIA1 (ANR-10-INBS-04). CSDC was supported by the Chan Zuckerberg Initiative (Imaging Scientist award nr. 2019-198155) K.N. was supported by Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Figures

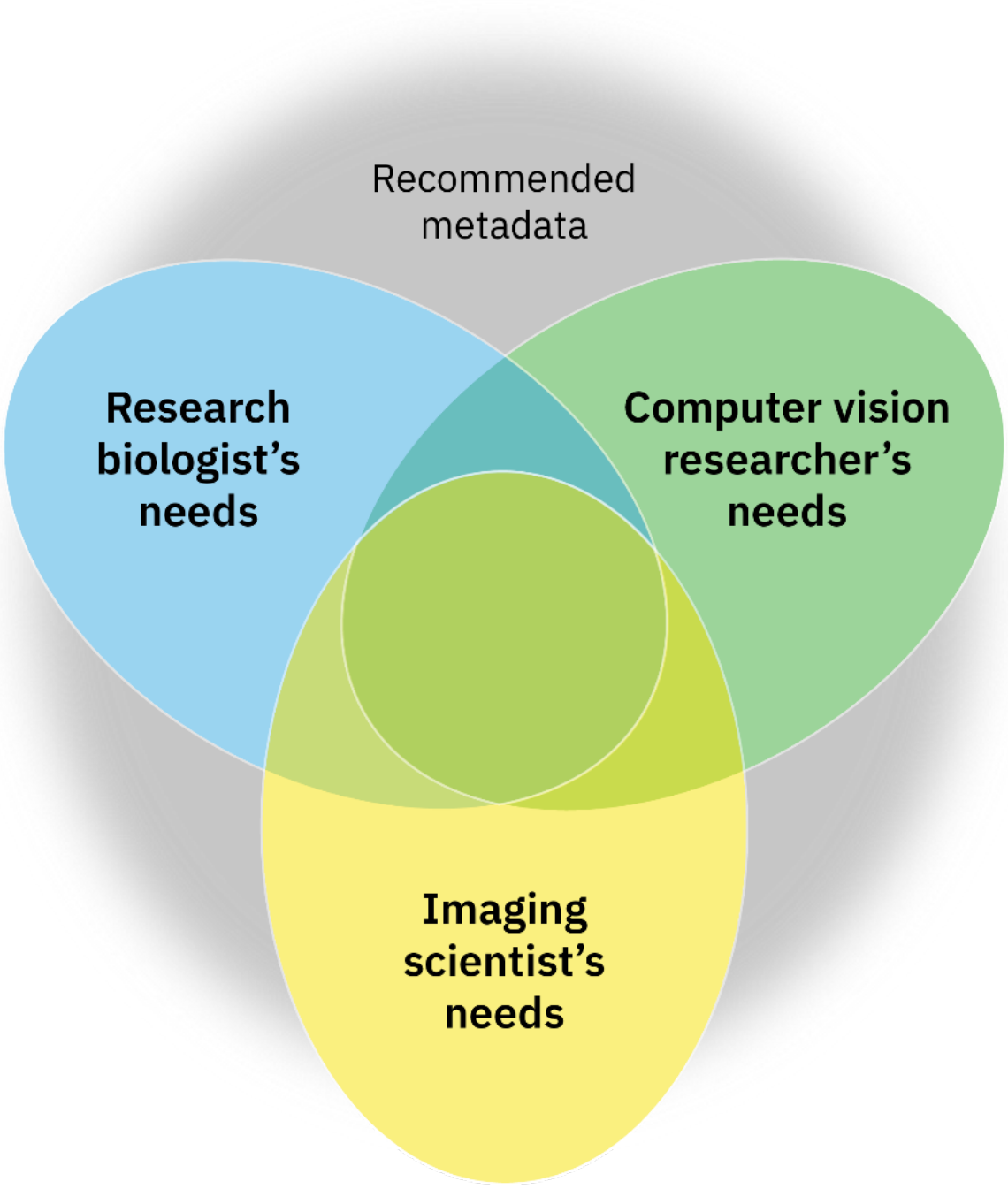

Figure 1. There are at least three different categories of users of archived images, each with different needs with respect to metadata. (1) Biologists and life scientists who are interested in repeating experiments, (re-)analysing or comparing bioimage data and understanding results. For this, they need detailed information on the experimental context such as the composition of biological samples, molecular entities, experimental interventions (e.g. control vs treatment) and

how these relate to the image data. (2) Imaging scientists (microscopists and technology developers) who are interested in developing new imaging technologies. For this, they need detailed information on the image-acquisition process such as physical properties of the image-acquisition set-up, and may benefit from some high-level information on the biological problem at hand. (3) Computer-vision researchers who develop novel algorithms (not limited to biological applications). Depending on the objective, they may need any of the information listed above. For example, to train a machine-learning algorithm, they would need "ground-truth" information such as adequately labelled images with categories (e.g. control vs treatments/phenotypes) or object outlines (i.e., segmentations).

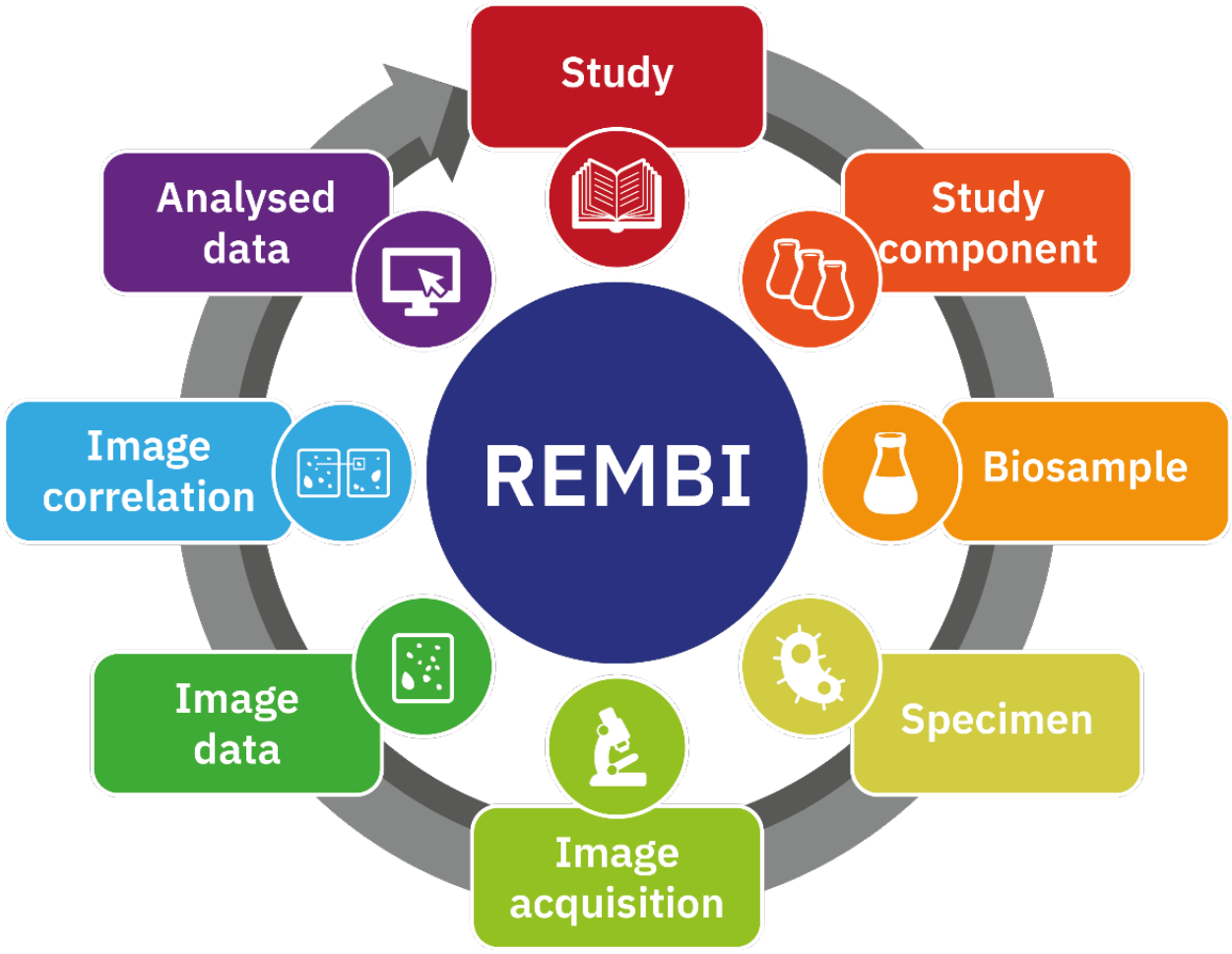

Figure 2. Different categories of metadata that need to be covered. Briefly, the metadata elements can be grouped into eight modules capturing information about the study, study component, biological sample, specimen, image acquisition, image data, analysed data, and information about the correlations between the underlying images). The "study component"

module represents a set of images and the associated metadata in a larger, multi-modal study. For example, in a correlative study comprising SBF-SEM and confocal images one of the study components would contain all information on the EM image stack, the other study component would correspond to the confocal stack, and a transformation description would allow an overlay of both types of images. Data that retain spatial fidelity to underlying images (e.g., label maps, volume renderings) are described in the “image data” module, whereas “analysed data” (e.g. volumetric analyses, image segment features, counts) contains image-derived measurements, typically presented in tabular form. For more detail see **supplementary material)**.

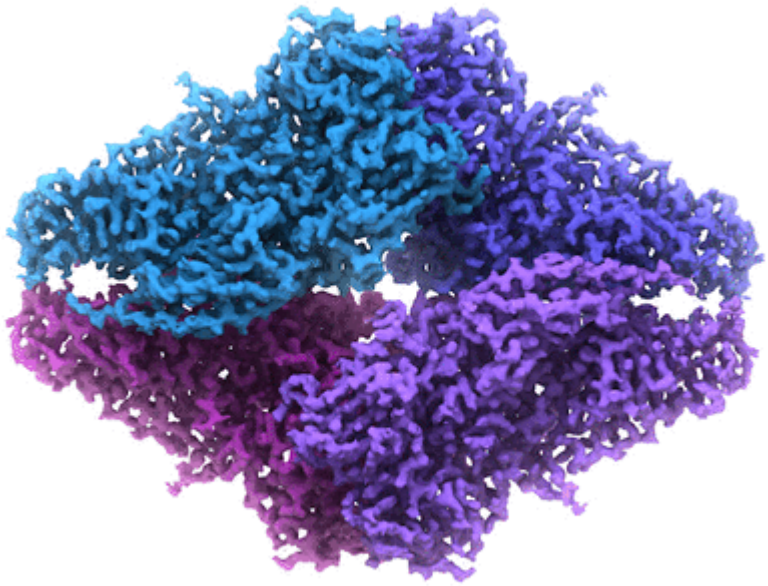

Figure 3. Imaging data is already being reused. An example of a widely reused dataset is EMPIAR entry EMPIAR-10061 (https://empiar.org/10061) which contains the raw cryo-EM data (12.4 TB in size) underpinning what was a breakthrough structure and at the time the highest resolution cryo-EM structure available, the 2.2Å resolution structure of β-galactosidase[18]. Several groups have re-processed the data to even higher resolution and published and deposited the resulting EM maps. The dataset has been used by several developers of cryo-EM processing software to improve and test their algorithms, it was used in the development of two deep-learning methods for automated particle picking, and to demonstrate cloud-based data processing. More details (and literature references) can be found online (https://www.ebi.ac.uk/pdbe/emdb/empiar/reuse). The raw data for this structure was deposited in EMPIAR and has found wide reuse.

References

1. Schermelleh, L., et al., *Super-resolution microscopy demystified.* Nat Cell Biol, 2019. **21**(1): p. 72-84.
2. Fernandez-Leiro, R. and S.H. Scheres, *Unravelling biological macromolecules with cryo-electron microscopy.* Nature, 2016. **537**(7620): p. 339-46.
3. Nakane, T., et al., *Single-particle cryo-EM at atomic resolution.* BioRxiv, 2020.
4. Huisken, J., et al., *Optical sectioning deep inside live embryos by selective plane illumination microscopy.* Science, 2004. **305**(5686): p. 1007-9.
5. Duncan, L.H., et al., *Isotropic Light-Sheet Microscopy and Automated Cell Lineage Analyses to Catalogue Caenorhabditis elegans Embryogenesis with Subcellular Resolution.* J Vis Exp, 2019(148).
6. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* Sci Data, 2016. **3**: p. 160018.
7. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, 2001. **29**(4): p. 365-71.
8. Ioannidis, J.P., et al., *Repeatability of published microarray gene expression analyses.* Nat Genet, 2009. **41**(2): p. 149-55.
9. Rung, J. and A. Brazma, *Reuse of public genome-wide gene expression data.* Nat Rev Genet, 2013. **14**(2): p. 89-99.
10. Linkert, M., et al., *Metadata matters: access to image data in the real world.* J Cell Biol, 2010. **189**(5): p. 777-82.
11. Mildenberger, P., M. Eichelberg, and E. Martin, *Introduction to the DICOM standard.* Eur Radiol, 2002. **12**(4): p. 920-7.
12. Marques, G., T. Pengo, and M.A. Sanders, *Imaging methods are vastly underreported in biomedical research.* Elife, 2020. **9**.
13. Ellenberg, J., et al., *A call for public archives for biological image data.* Nat Methods, 2018. **15**(11): p. 849-854.
14. Iudin, A., et al., *EMPIAR: a public archive for raw electron microscopy image data.* Nat Methods, 2016. **13**(5): p. 387-8.
15. Williams, E., et al., *The Image Data Resource: A Bioimage Data Integration and Publication Platform.* Nat Methods, 2017. **14**(8): p. 775-781.
16. Lawson, C.L., et al., *EMDataBank unified data resource for 3DEM.* Nucleic Acids Res, 2016. **44**(D1): p. D396-403.
17. ww, P.D.B.c., *Protein Data Bank: the single global archive for 3D macromolecular structure data.* Nucleic Acids Res, 2019. **47**(D1): p. D520-D528.
18. Bartesaghi, A., et al., *2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor.* Science, 2015. **348**(6239): p. 1147-51.

REMBI overview

Module	Attribute	Comments	Data entry method	WG1 - example 1	WG1 - example 2	WG2 - example 1	WG2 - example 2	WG3 - example 1	WG3 - example 2	WG3 - example 3	WG3 - example 4
Study											
(contains 1 or more Study components)	Study type	Type of the overall study, which may include other imaging and/ or non-imaging data	text, ontology	Single-particle cryo-EM	cryo-ET	CLEM/FIB-SEM of activated CD4+ T cells	FRAP/FIB SEM of parasitophorous vacuoles containing Toxoplasma gondii ΔCAP parasites	N/A	N/A	High-throughput/high-content screening	Atlas
	Study description	Study description, e.g., title of published paper	text	Single-particle cryo-EM at atomic resolution	Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.7 Å inside cells	Plasma Membrane LAT Activation Precedes Vesicular Recruitment Defining Two Phases of Early T-cell Activation	Differential requirements for cyclase-associated protein (CAP) in actin-dependent processes of Toxoplasma gondii	N/A	Ex vivo live cell tracking in kidney organoids using light sheet fluorescence microscopy		Data to build a mitotic cell atlas
	General dataset info	Authors, publications, licenses etc	misc.	https://doi.org/10.1101/2020.05.22.110189	https://doi.org/10.1101/2020.06.05.136341	PMID: 29789604	Hunt A , Russell MRG , Wagener J , Kent R, Carmeille R, Peddie CJ, Collinson L, Heaslip A, Ward GE, Treeck M Elife 8 (2019) PMID: 31577230 DOI: 10.7554/elife.50598	Images acquired by Jeffrey Skerker and annotated by Tom Morgan	PMID: 31527839	Hériché JK et al.. Mol Biol Cell. 2014 Aug 15;25(16):2522-2536. doi: 10.1091/mbc.E13-04-0221. PubMed PMID: 24943848	Cai Y, et al. An experimental and computational framework to build a dynamic protein atlas of human cell division. Nature. 2018 Sep 10. doi: 10.1038/s41586-018-0518-z. PubMed PMID: 30202089.
Study component											
(contains Image data and Analysed data)	Imaging method	Technique used to acquire image data	ontology	Single-particle cryo-EM	cryo-ET	FIB-SEM	FIB SEM	Differential interference contrast	Light sheet fluorescence microscopy	Epifluorescence microscopy	Confocal fluorescence microscopy + Fluorescence correlation spectroscopy
	Study component description	Description specific to this image dataset component	text	N/A	N/A	StudyComponentDescription.rtf	Processed FIB SEM images of a parasitophorous vacuole containing Toxoplasma gondii ΔCAP parasites	N/A	N/A	Image-based RNAi screen of 100 candidate genes predicted to be involved in mitotic chromosome condensation.	Image data used to build the mitotic cell atlas. This consists of three subsets: primary images, segmentation masks for the landmarks and protein concentration maps.
Biosample											
	Identity	Internal unique ID				pLAT_8O-2_FIBSEM	EM04226_2_U19	BBBC009			
	Biological entity	What is being imaged	text and/or ontology entry (multiple possible)	apo-ferritin	ribosome-antibiotic complex	Jurkat E6.1 T cells; http://purl.obolibrary.org/obo/CL_0000084	Parasitophorous vacuole, in a human foreskin fibroblast, containing Toxoplasma gondii ΔCAP parasites	Red blood cells	Kidney tissue	HeLa cells	HeLa cells
	Organism	Species	NCBI taxonomy (multiple possible)	Mouse	Homo sapiens	Homo Sapien; http://purl.obolibrary.org/obo/NCBITaxon_9606	Homo Sapiens, Toxoplasma gondii	Homo sapiens	CD1 Mus musculus embryonic, day 13.5		
	Intrinsic variable	Intrinsic (e.g. genetic) alteration if applicable	text and/or ontology entry (multiple possible)	N/A	N/A	Jurkat E6.1 transfected with emerald-VAMP7	N/A	N/A	Wild type	stable overexpression of HIST1H2BJ-mCherry and LMNA-e	Homozygous GFP integration into mitotic genes + fluorescent dextran in medium + either stable overexpression of HIST1H2BJ-mCherry or SiR-Hoechst staining
	Extrinsic variable	External biosample treatment (e.g. reagent) if applicable	text and/or ontology entry (multiple possible) or associated file	N/A	N/A	Plate-bound anti-CD3 activation	N/A	N/A	N/A	Library of 200 siRNAs, the sequences of which should be listed here as they are required for target gene inference.	
	Experimental variables	What is intentionally varied (e.g. time) between multiple entries in this study component	text	N/A	N/A	Time	Complementation of ΔCAP mutant with wild-type CAP (not done for this sample, hence mutant phenotype)	N/A	Genotype	time	time
Specimen											
(linked to Biosample)	Experimental status	Test/ control		N/A	N/A	N/A	Test (ΔCAP)	N/A	N/A		
	Location within Biosample	Plate/dish coordinate or tissue location	text or associated file	N/A	N/A	Plate2_8O	Experiment EM04226, dish 2, grid square U19	N/A	N/A	list of plate/wells coordinates	
	Preparation method	Sample preparation protocol	text, file, or widget for specific method types	A frozen aliquot of 7mg/ml mouse apo-ferritin in 20mM HEPES pH 7.5, 150mM NaCl, 1mM dithiothreitol (DTT) and 5% trehalose, which we received from the Kikawa Lab at Tokyo University, was thawed at room temperature and cleared by centrifugation at 10,000g for 10 min. The supernatant was diluted to 5mg/ml with 20mM HEPES pH 7.5 150mM NaCl, and 3 µl of the diluted sample was applied onto glow-discharged R1.2/1.3 300 mesh UltrAuFoil gold grids (Quantifoil) for 30 s and then blotted for 5 s before plunge-freezing the grids into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot Mark IV (Thermo Fisher Scientific) at 100% humidity and 4 °C.	Mycoplasma pneu- moniae strain M129 (ATCC 29342)cells were grown on 200 mesh gold grids coated with a holey carbon support (R 2/1, Quantifoil). Cells were cultivated at 37 °C in modified Hayflick medium: 14.7 g/L Difco PPLO (Becton Dickinson, USA), 20% (v/v) Gibco horse serum (New Zealand origin, Life Technologies, USA), 100 mM HEPES-Na (pH 7.4), 1% (w/w) glucose, 0.002% (w/w) phenol red and 1,000 U/mL freshly dissolved penicillin G. Chloramphenicol (Cm; Sigma-Aldrich, USA) was added 15 minutes prior to vitrification, at a final concentration of 0.5 mg/ml. Grids were quickly washed with PBS buffer containing 10 nm protein A-conjugated gold beads (Aurion, Netherlands), blotted from the back side for 2 seconds, and plunged into mixed liquid ethane/propane at liquid N2 temperature with a manual plunger (Max Planck Institute of Biochemistry, Germany). The cryo-EM grids were stored in a sealed box in liquid N2 before usage.	EM_SampProcProt_Revised_01.xls	RT fixation and megametal sample preparation for FIB SEM	N/A	CLARITY cleared		
	Signal/contrast mechanism	How is the signal generated by this sample		3D classification	Multiplying the FT of a particle image by the corresponding real-valued CTF	Heavy metal staining; m/z contrast detected by back-scatter electron detector	Osmium, uranium, lead, as contrast agents, m/z contrast detected by back-scatter electron detector	N/A	Antibodies	fluorescent proteins	fluorescent proteins and fluorescent dyes
	Channel - content	Specific specimen staining (e.g. IEM, DAB)		N/A	N/A	N/A	N/A	N/A	Megalin 1:200 overnight, Laminin 1:1000 overnight	red: HIST1H2BJ-mCherry, green: LMNA-eGFP	490-552:protein of interest-GFP; 587-621:HIST1H2BJ-mCherry | 622-695:SiR-DNA; 622-695:Dy-481XL-labelled 500 kDa dextran
	Channel - biological entity	What molecule is stained		N/A	N/A	N/A	N/A	N/A	Megalin: tubular lumen, green; Laminin: basement membrane, red	red: HIST1H2BJ-mCherry, green: LMNA-eGFP	490-552:protein of interest-GFP; 587-621:HIST1H2BJ-mCherry | 622-695:SiR-DNA; 622-695:Dy-481XL-labelled 500 kDa dextran
Image acquisition											
(linked to Specimen)	Instrument attributes	Details about instruments used	text, file, or widget for specific instrument types								
	Image acquisition parameters	Image acquisition details	text, file, or widget for specific acquisition method types	All cryo-EM data were collected on Falcon cameras in electron counting mode using Titan Krios microscopes (Thermo Fisher Scientific) operating at 300 kV. Before data acquisition, two-fold astigmatism was corrected and beam tilt was adjusted to the coma-free axis using the autoCTF program. All datasets were acquired automatically using EPU software (Thermo Fisher Scientific). Detailed data acquisition parameters for all data sets are given in Extended Data Table 1.	Tilt series data were collected on a Titan Krios TEM operated at 300 kV (Thermo Fisher Scientific) equipped with a field-emis- sion gun, a Gatan K2 Summit direct detector and a Quantum post-column energy filter (Gatan). Images were recorded in exposure-fractionation, counting mode using SerialEM 3.7.2. Tilt-series were acquired with a dose-symmetric scheme us- ing dedicated scripts51 with the following settings: TEM in na- no-probe mode, magnification 81,000 with a calibrated pixel size of 1.7 Å, energy filter in zero loss mode, defocus range 1.5 to 3.5 µm, tilt range -60° to 60° with 3° tilt increment and constant ex-posure per tilt, total exposure of 120 e-/Å2. In total, 65 tilt series were collected from Cm-treated cells.	FIB-SEM_Acquisition_Parameters_Samleson_pLAT_8O-2_correct.xls	5 nm isotropic, 10 µs dwell time. SEM at 1.5 kV with 1 nA current. EsB detector with grid voltage of 1.2 kV. Ion beam milling at accelerating voltage of 30 kV and current of 700 pA	N/A	N/A	Automated Olympus IX-81 epifluorescence microscope with 20X objective, images acquired every 8.5 min for 44h.	
Image data											
(result of Image acquisition, or processing of Image data)	Type	Primary image/processed image/segmentation	pull-down	micrographs	tilt series	Processed Image	Processed image	Primary image and segmentation outlines	Primary image	Primary image	Primary images, segmentation masks for the landmarks and protein concentration maps
	Format & compression	File type	extract	EER	MRC	.mrc	.mrc	.tif	.czi	TIFF	
	Dimension extents	Volume in pixels: x, y, z, tilts	extract	4096, 4096, 434	4096, 4096, 434	1000, 800, 546		800 x 600	x: 1920, y: 1920, z: 600, c: 2, v: 5, t: 1	x: 1344, y: 1024, t: variable, ~330	x: 256, y: 256, z: 31
	Size description	Physical size of image volume in x,y,z & units (pull down), OR magnification	extract	N/A (can be calculated from voxel size)	N/A (can be calculated from voxel size)	15 x 12 x 8.719 um	15.62 x 16.70 x 3.61 µm	N/A	N/A (can be calculated from voxel size)		
	Pixel/voxel size description	Physical size of pixels in x, y, z & units (pull down)	extract	14 14 14 nm	10 x 10 x 10 nm	15 x 15 x 15.97 nm	10 x 10 x 10 nm	N/A	0.57 um x 0.57 um x 1.30 um	0.32 um x 0.32 um	0.25 um x 0.25 um x 0.75 um
	Image processing method	Image registration, other processing applied to this dataset	text, file, or widget for specific method types that may facilitate ontology use, e.g., EDAM Bioimaging ontology	A total of 3370 movies in EER format were motion corrected with RELION's implementation of the MotionCor2 algorithm53. For this purpose, original hardware movie frames were dose fractionated into groups of 14 frames, corresponding to an accumulated dose of 1.3 e-/Å2 per fraction. CTF estimation was performed with CTFFIND-4.1.1354 using the sums of power spectra from combined fractions corresponding to an accumulated dose of 4 e-/Å2. Micrographs whose estimated resolution from CTFFIND was worse than 5 Å were removed, leaving 3080 micrographs for further processing.	Raw tilt movies were processed in Warp. De novo tilt series alignment was performed in IMOD using gold fiducials picked automatically with Warp's BoxNet, and the results were imported in Warp, where the tilt series CTFs were estimated. Using full tomograms reconstructed at 10 Å/px, two tomograms were de- noised using Warp's Noise2Map tool to pick the ribosome par- ticles manually. Using these coordinates, sub-tomograms were exported from Warp to RELION to obtain an initial reference.	cross-correlation based stack alignment + binning + inversion	Zeiss Atlas 5.2.1.25; Gradient align Export of cropped region of interest exported as 8 bit TIF files Fiji; Invert Gaussian blur: Sigma 0.75 Unsharp mask: Radius (Sigma) 1 pixel, Mask weight 0.6 Rotate to straighten XY (D19: no rotation) Reslice: YZ assigned to XY plane Scale to 10 nm isotropic voxels IMOD; tif2mrc	N/A	N/A		Segmentation, FCS-based intensity calibration
	Contrast inversion to TEM	Y/N; N if stained features result in brighter (whiter) signal; Y if it looks like a TEM image	pull-down	N/A	N/A	Y	Y	N/A	N/A		
	QC info	QC score for uploaded image quality if applicable		N/A	N/A	N/A	N/A	N/A	N/A		
Image Correlation											
(linked to 1 or more Image data)	Spatial and temporal alignment	Method used to correlate images from different modalities (e.g. manual overlay, alignment algorithm etc)		N/A	Warp's BoxNet	Manual	Manual	N/A	N/A		
	Fiducials used	Features from correlated datasets used for colocalization		N/A	gold fiducials	Plasma membrane and nuclear membrane	Paraseite fluorescent reporter used for FRAP	N/A	N/A		
	Transformation matrix/ other info	Correlation transforms	text, or related project files (e.g. .hx Amira files)	N/A	N/A	N/A	N/A	N/A	N/A		
	Related images and relationship	Correlated dataset or images	link	N/A	N/A	pLAT_8O-2_Fluo.tif	FRAP, fluorescence after fixation, same area imaged (not deposited). EMPIAR-10326; raw data from which this entry was processed	N/A	N/A		
Analysed data											
	Analysis result type	Numerical analyses, segmention (non-image), categorical features/phenotypes		N/A	N/A	N/A	Decentralised residual body central axis skeletonisation (green), basal pole (orange), coarse segmenation of part of parasite surface membrane (yellow), putative ER in a region of the decentralised residual body (blue).	Segmentation outlines	N/A	Phenotypes	protein concentrations in time and space
	Data used for analysis	Specific feature set used for analysis (e.g. volume measurements, locations of features)	text or file(s)	N/A	N/A	N/A	N/A	Primary image	N/A	associated HDF5 file	associated files: cell features, mitotic cell model, sequences used to edit the genomic loci (to identify the genes), protein concentrations (in nM) at each voxel of the mitotic cell model
	Analysis method and details	Analysis method	text, file, or pointer to Methods section	N/A	N/A	N/A	3dmod manual segmentation	Manual annotation	N/A	see Hériché JK et al.. Mol Biol Cell. 2014 Aug 15;25(16):2522-2536. doi: 10.1091/mbc.E13-04-0221. PubMed PMID: 24943848	Cai Y, et al. An experimental and computational framework to build a dynamic protein atlas of human cell division. Nature. 2018 Sep 10. doi: 10.1038/s41586-018-0518-z. PubMed PMID: 30202089.