

Preparing for the NIH Data Management and Sharing Policy: An Overview and A Case Study on Image Data

Valeria Mezzano
Research Assistant Professor
Department of Pathology
NYU Langone Health

Nicole Contaxis
Data Librarian
NYU Health Sciences Library
NYU Langone Health

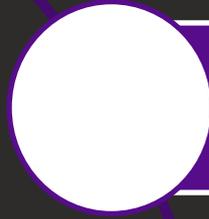


Agenda

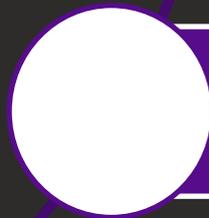
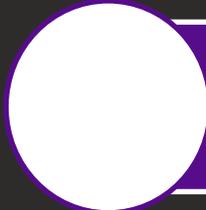
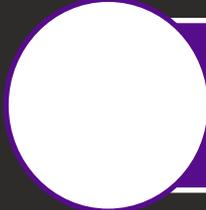
- Four steps for preparing for the NIH Data Management and Sharing Policy at NYU Langone
- From the Cores' perspective
- Conclusions

Step #1: Understand the policy

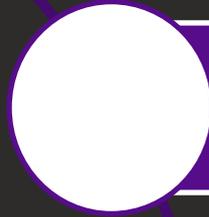
2023 NIH Policy for Data Management and Sharing



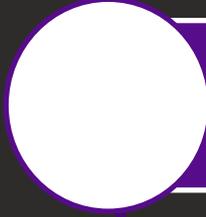
Requires researchers to submit **Data Management and Sharing (DMS) Plans** for all research conducted at or funded by the NIH that generates scientific data.



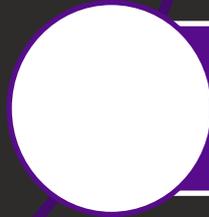
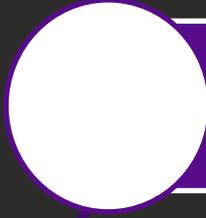
2023 NIH Policy for Data Management and Sharing



Requires researchers to submit **Data Management and Sharing (DMS) Plans** for all research conducted at or funded by the NIH that generates scientific data.



Requires researchers to follow through with their DMS Plan when they publish or by the time their award period ends.



2023 NIH Policy for Data Management and Sharing

Requires researchers to submit **Data Management and Sharing (DMS) Plans** for all research conducted at or funded by the NIH that generates scientific data.

Requires researchers to follow through with their DMS Plan when they publish or by the time their award period ends.

Effective January 25, 2023

2023 NIH Policy for Data Management and Sharing

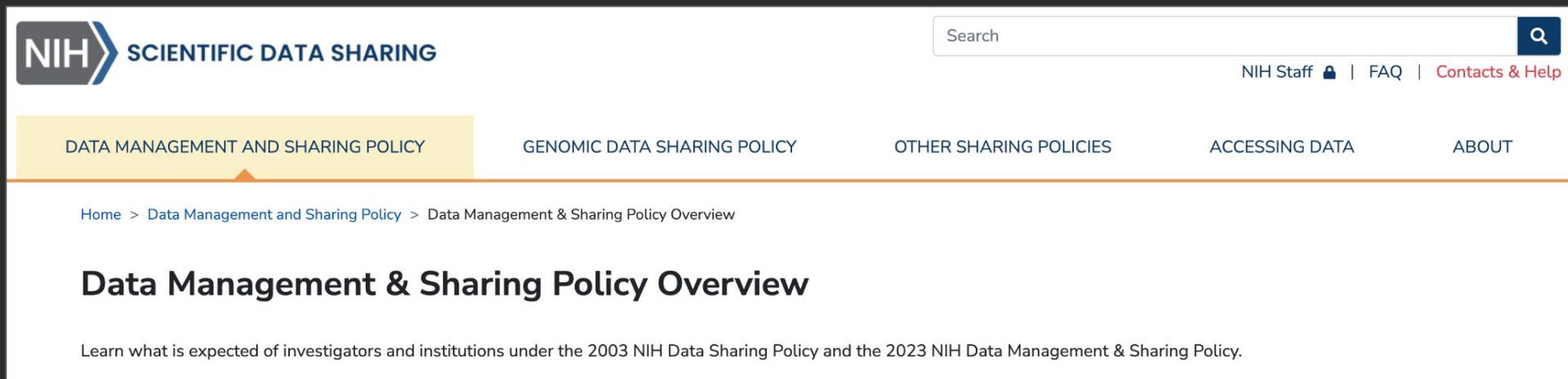
Requires researchers to submit **Data Management and Sharing (DMS) Plans** for all research conducted at or funded by the NIH that generates scientific data.

Requires researchers to follow through with their DMS Plan when they publish or by the time their award period ends.

Effective January 25, 2023

Non-compliance may be factored into future NIH funding decisions.

NIH Resources



NIH SCIENTIFIC DATA SHARING

Search

NIH Staff | FAQ | Contacts & Help

DATA MANAGEMENT AND SHARING POLICY GENOMIC DATA SHARING POLICY OTHER SHARING POLICIES ACCESSING DATA ABOUT

Home > Data Management and Sharing Policy > Data Management & Sharing Policy Overview

Data Management & Sharing Policy Overview

Learn what is expected of investigators and institutions under the 2003 NIH Data Sharing Policy and the 2023 NIH Data Management & Sharing Policy.

<https://sharing.nih.gov/data-management-and-sharing-policy>

Includes:

- The policy itself and all associated supplemental information
- FAQs on the policy
- Assistance on budgeting and new allowable costs
- Guidance on NIH resources for managing and sharing data, like NIH repositories

Step #2: Collaborate

Example at NYU Langone Health

Development of a Cross-Institutional Working Group to Address the Policy

- Coordinate and collaborate on preparations for the policy change
- Ensure diverse expertise and viewpoints are represented in preparations
- Develop **resources**, **outreach strategy**, and **education strategy**
- **Members:**
 - NYU Health Sciences Library, Office of Science and Research, Cores, Legal, IRB, IT, Sponsored Projects Admin, 'Data Hubs'

Step #3: Be Realistic

Outreach

Education

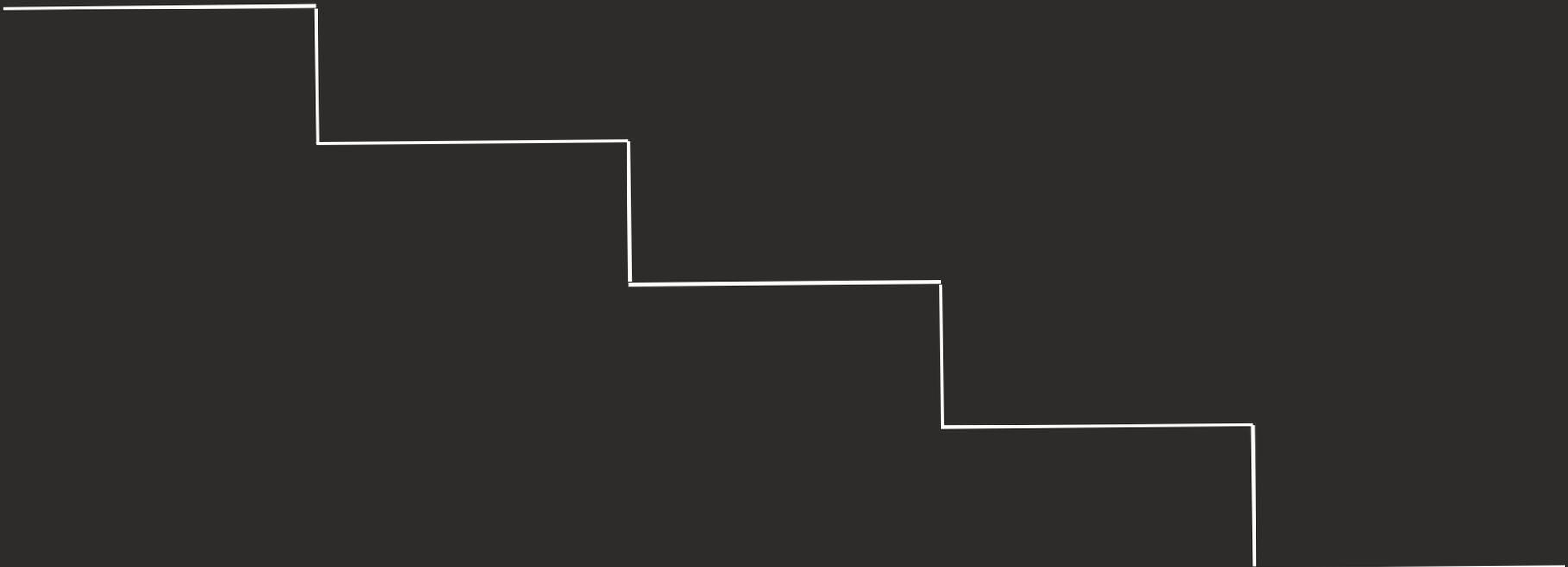
**Resource
Creation**

**DMS Plan
Review**

maDMPs

FTE and level of data expertise at the institution

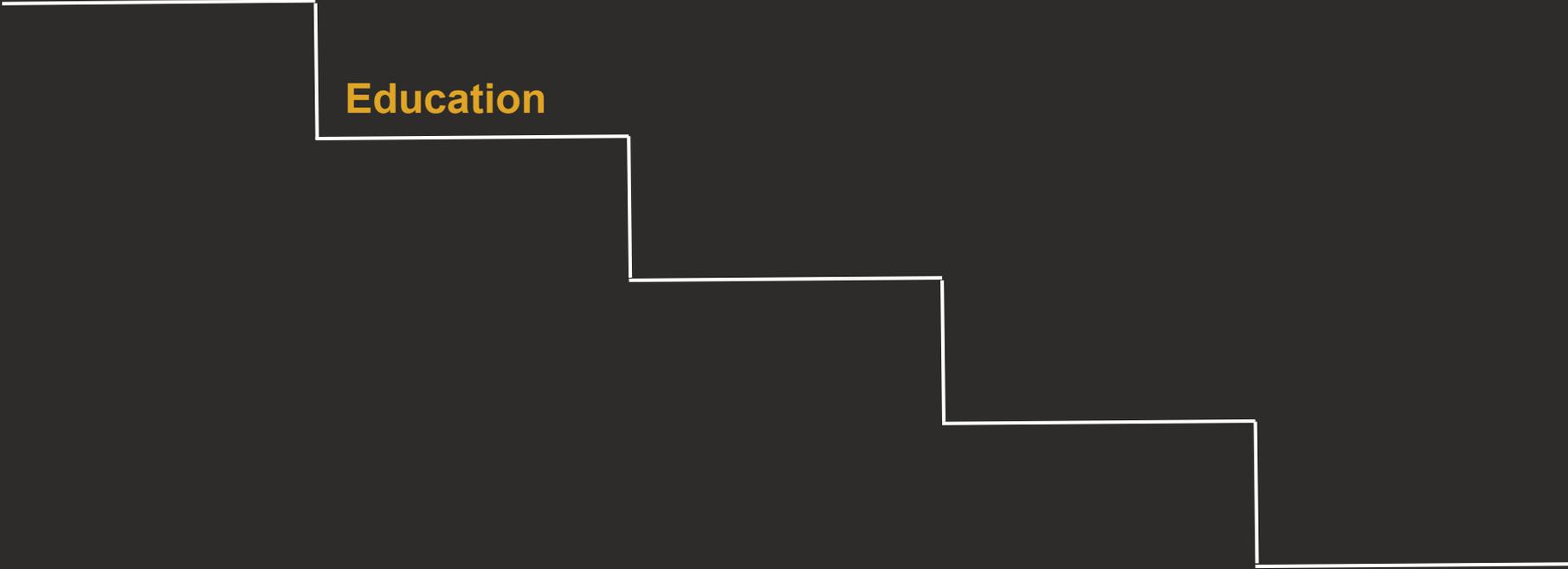
Outreach

- 
- Ensuring researchers and librarians know about the policy change
 - Can include in-person outreach (e.g., short presentations at faculty meetings) and online outreach (e.g., notices throughout your institution's web presence)

Example: Outreach at NYU Langone

Outreach

- Plan developed based on pre-existing communications strategies with our Office of Science and Research
- Plan includes:
 - Creation of a '[NIH DMS Policy Outreach Homepage](#)' that guides researchers to resources and answers basic questions on the policy and linking to that page across the NYU Langone web presence
 - Creation of a group email to triage questions on the policy



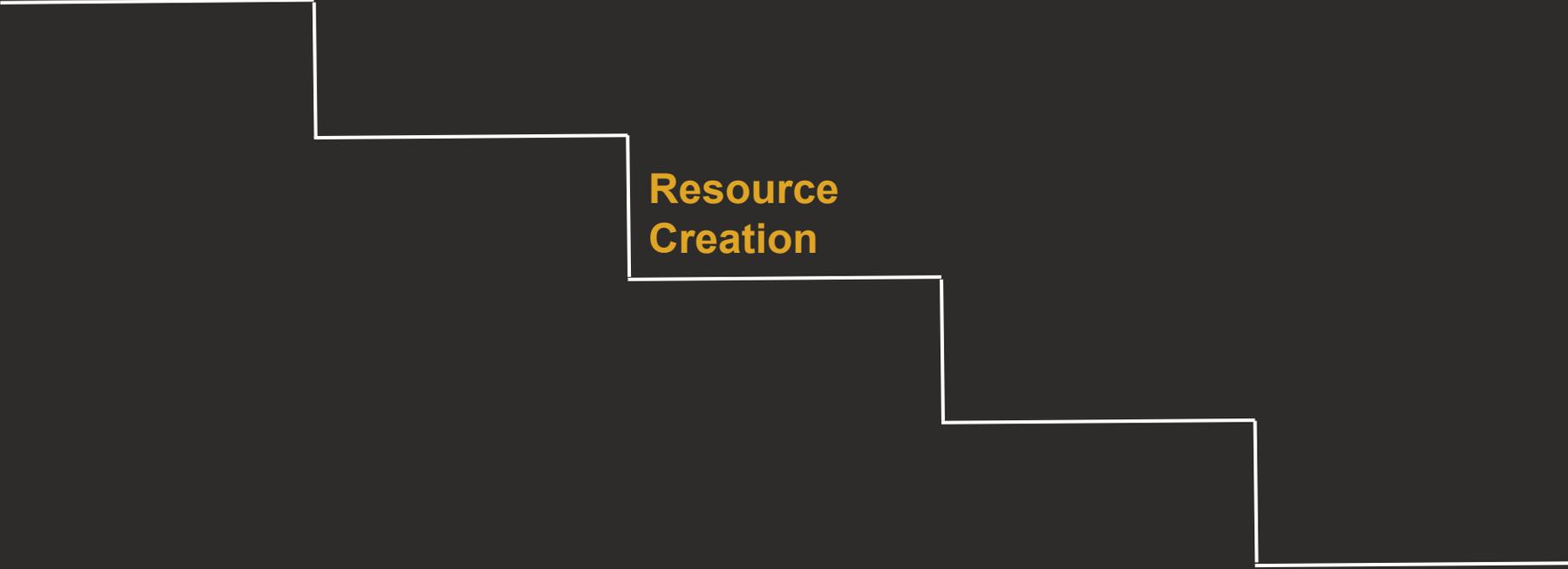
Education

- Ensuring researchers and librarians know how to comply with the policy
- Will need to address creating a DMSP and some data management best practices
- May include asynchronous or synchronous classes/materials

Example: Education at NYU Langone

Education and Classes

- Library spearheaded creating educational materials and work group provided feedback and created material when certain expertise was required
- **Classes created:** (1) 15 minute asynchronous online class; (2) 60 minute synchronous workshops at the library; (3) On demand classes/presentations for departments



Resource Creation

- Create or collect and make findable resources, materials or guidance that can help researchers comply
- Resources may include institution-specific guidance on writing a DMSP or materials created by the NIH (e.g., NIH.Sharing.Gov)

Resource Creation

Data management and sharing plan guidance

- Includes guidance specifically for NYULH researchers and point to additional internal and external resources
- Make it available in multiple formats, including on DMPTool
- **Collaborators:** Librarians at other institutions, Legal, IRB, Cores, IT, Office of Science and Research

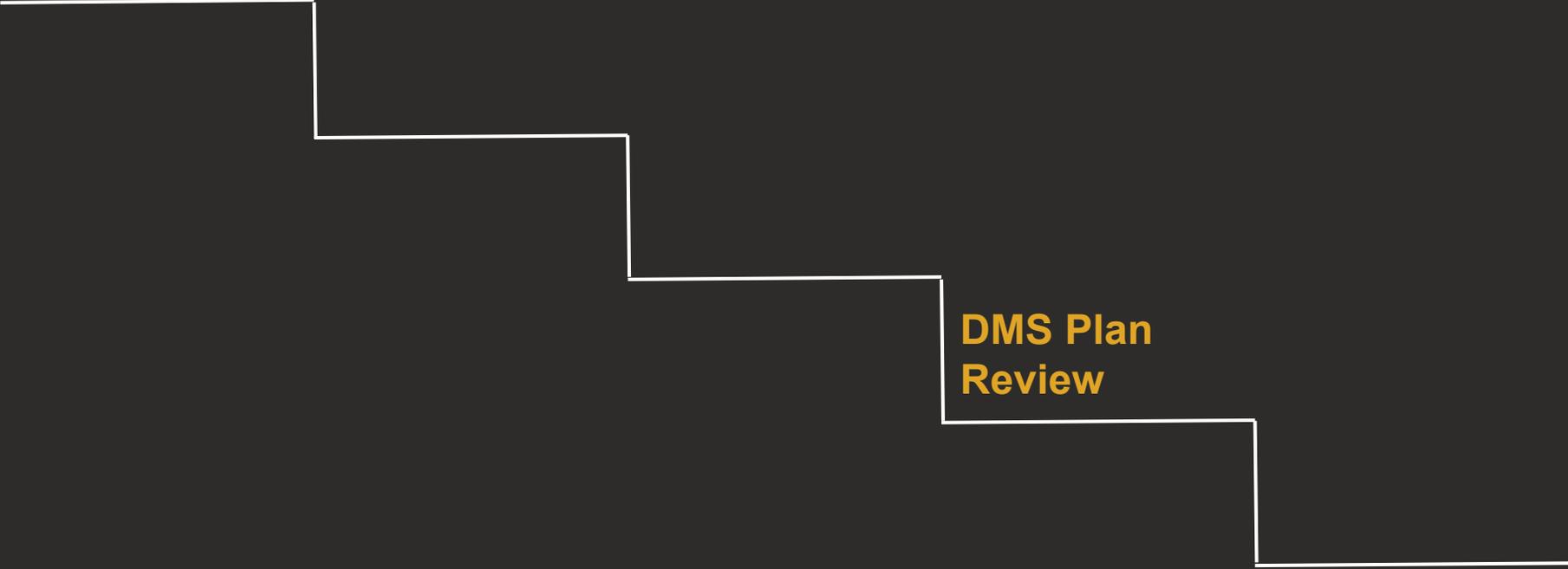
Resource Creation

Data management and sharing plan guidance

- Includes guidance specifically for NYULH researchers and point to additional internal and external resources
- Make it available in multiple formats, including on DMPTool
- **Collaborators:** Librarians at other institutions, Legal, IRB, Cores, IT, Office of Science and Research

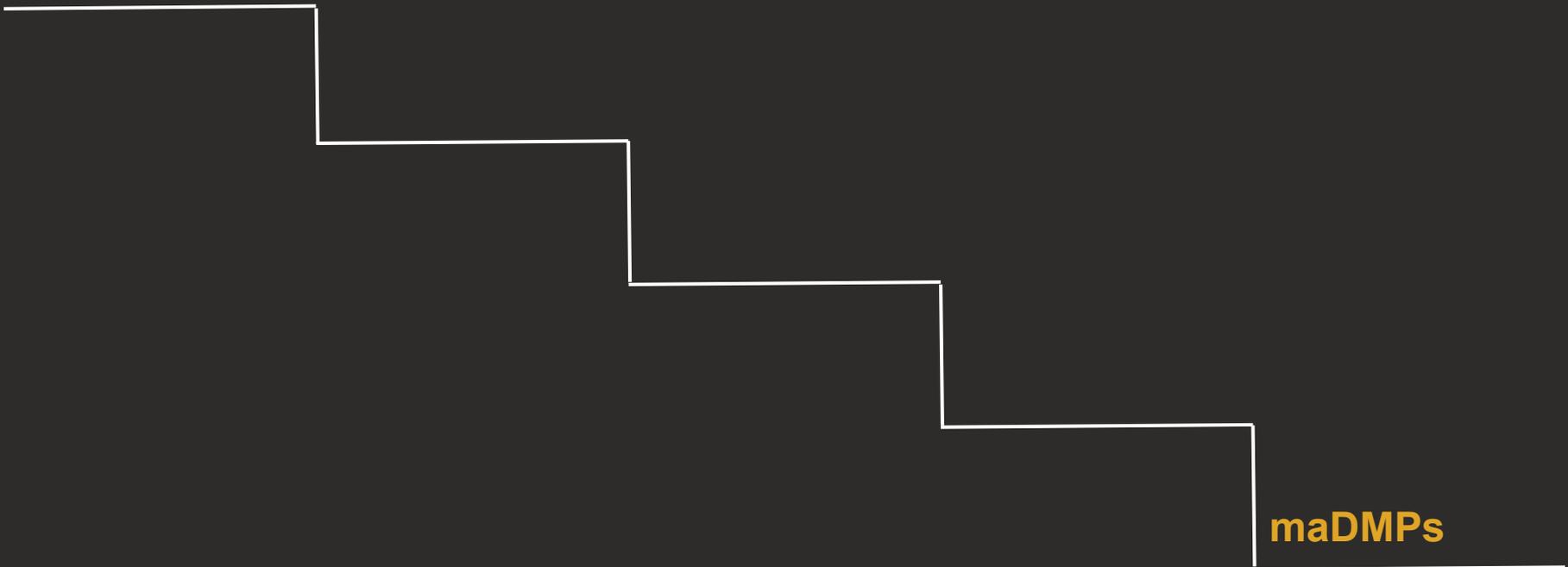
The Data Repository Finder

- Designed to help researchers locate repositories that meet NIH guidelines by asking researchers about their data/research and pointing them to suggested repositories
- Created using open source code shared by Cornell libraries
- **Collaborators:** Librarians at other institutions, IT



DMS Plan Review

- Reviewing, editing, or providing feedback on DMSPs at point of grant submission to provide guidance on data management and sharing best practices
- May require significant resources depending on the number of researchers who request this service



maDMPs

- “Machine-actionable Data Management Plan”
- A framework undergoing extensive research in feasibility and implementation
- Meant to automate data management and sharing processes by making information in a DMSP ‘machine-readable’

Step #4: Reuse Tools and Resources

Example Tools and Resources

Toolkit for the NIH Data Management and Sharing Policy from the National Center of Data Services at the Network of the National Library of Medicine

- Link:
<https://www.nlm.nih.gov/guides/nlm-toolkit-nih-data-management-and-sharing-policy>
- Some overlapping resources from the librarian-led collaboration on OSF
- Additionally, contains:
 - Class slides on the policy if you want to teach to your community
 - Links to relevant NIH resources
 - Recordings of previous classes on the policy

Example Tools and Resources

DMPTool

- Link: <https://dmptool.org/>
- Helps you write DMS Plans that meet funder requirements, including the new NIH policy
- Includes guidance on what to include in each section and example language for each section

Librarian created resources working through RDAP, NCDS, and MLA-Data Sig

- Link: <https://osf.io/uadxr/>
- Contains:
 - Policy Readiness Checklist
 - DMS Plan Checklist for Researchers
 - Example DMS Plans
 - Glossaries

WHO WHAT HOW

From a shared resource laboratory perspective

FAIR

Findable, Accessible, Interoperable, Reusable

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).

WHO

Campus

Most of the labs are located in the same campus, many exceptions.

Labs

200
+

People

Basic scientists, Clinician
scientists

IT

Highly centralized, with a dedicated team to support
research

Cores

Under one division (DART) and interactive among
each other

Health Sc.
Library

Rodent Genetic Eng.

Flow Cytometry

C.Biospecimen
Res.

Pre-clinical imaging

Genome
Technology

Bioinformatics

Microscopy

Proteomics

Immune
monitoring

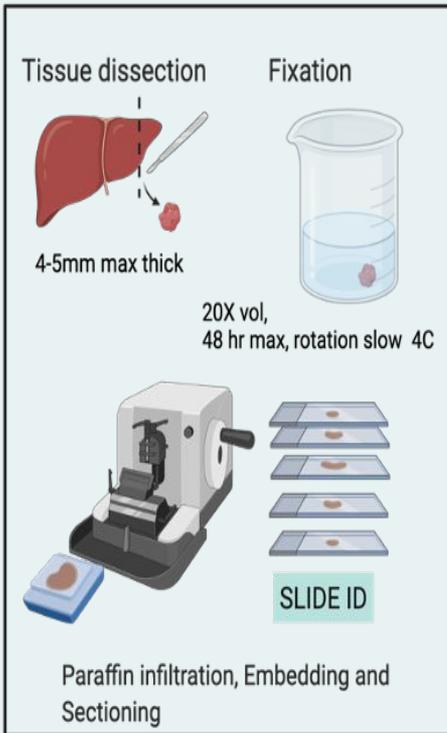
ExpPath

Metabolomics

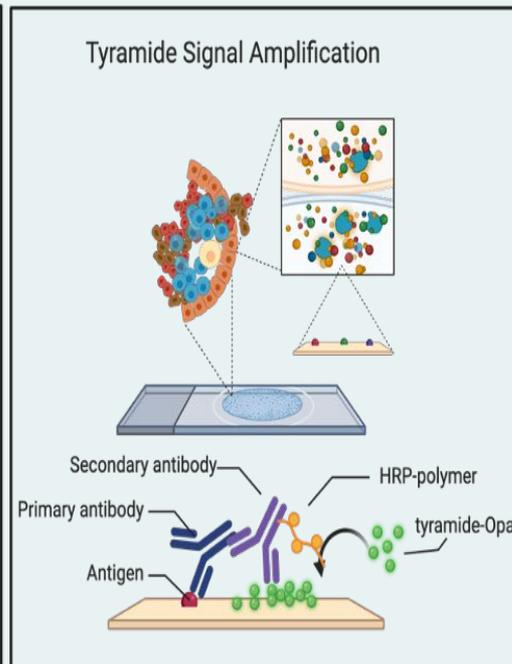
Cryo EM

WHO

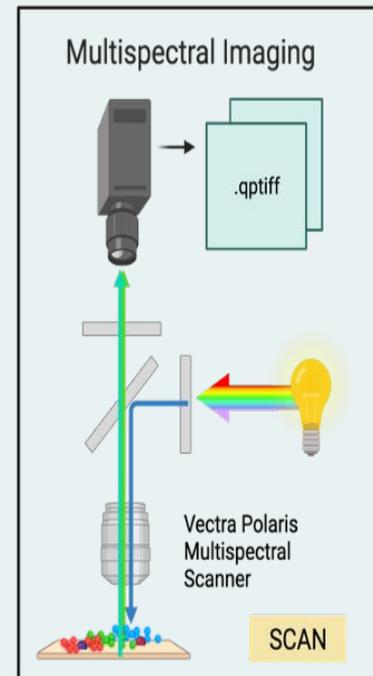
ExpPath Example



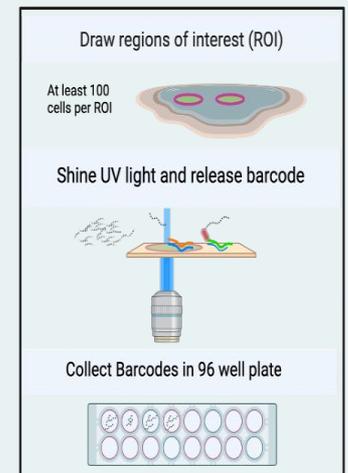
Processing



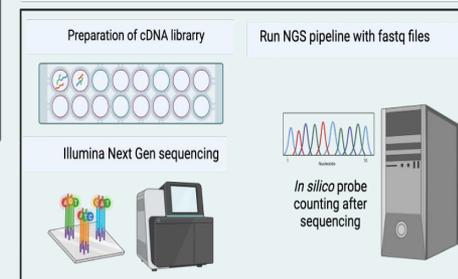
Staining



Slide Scanning



Genome Technology Core



Spatial profiling

WHAT

Standards

Wilkinson, M. *Sci Data* **3**, 160018 (2016).

“Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components.”

A FEW EXAMPLES

Proteomics

<http://www.proteomexchange.org/>

Flow Cytometry

MIFlowCyt: The minimum information about a flow cytometry experiment *Cytometry A*. 2008 Oct;73(10):926-30.

<https://doi.org/10.1002/cyto.a.20623>

Data File Standard for Flow Cytometry, version FCS 3.1 *Cytometry A* 2010 Jan;77(1):97-100 DOI: 10.1002/cyto.a.20825

Metabolomics

COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access *Metabolomics*

2015;11(6):1587-1597. doi: 10.1007/s11306-015-0810-y

Multiplexed imaging

MITI minimum information guidelines for highly multiplexed tissue images. *Nat Methods*. 2022 Mar;19(3):262-267. doi:

10.1038/s41592-022-01415-4. <https://github.com/miti-consortium/MITI>

cryo-EM

Evolving data standards for cryo-EM structures *Struct Dyn*. 2020 Jan 24;7(1):014701. doi: 10.1063/1.5138589.

Histotechnology

Guide for collecting and reporting metadata on protocol variables and parameters from slide-based histotechnology assays to enhance reproducibility *J Histotechnol* 2022 Nov 1;1-16. doi: 10.1080/01478885.2022.2134022.



Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging

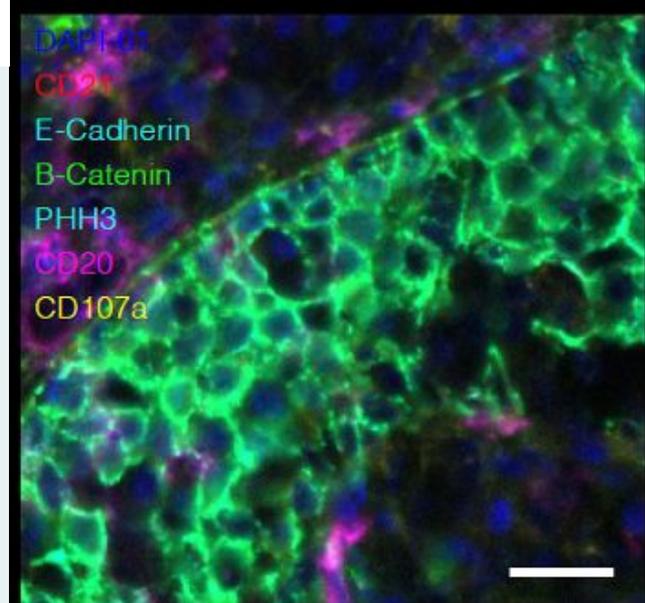
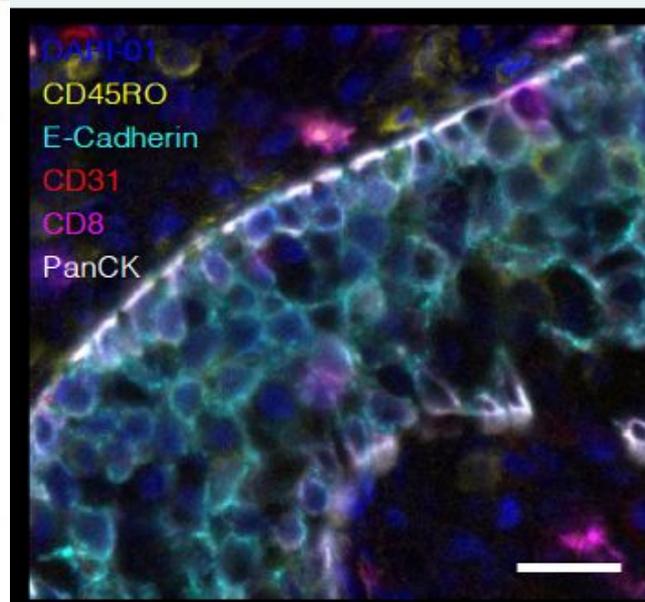
John W. Hickey^{1,26}, Elizabeth K. Neumann^{2,3,26}, Andrea J. Radtke^{4,26} ✉, Jeannie M. Camarillo⁵, Rebecca T. Beuschel⁴, Alexandre Albanese^{6,7,25}, Elizabeth McDonough⁸, Julia Hatler⁹, Anne E. Wiblin¹⁰, Jeremy Fisher¹¹, Josh Croteau¹², Eliza C. Small¹³, Anup Sood⁸, Richard M. Caprioli^{2,3,14}, R. Michael Angelo¹, Garry P. Nolan¹, Kwanghun Chung^{6,7,15,16,17,18}, Stephen M. Hewitt¹⁹, Ronald N. Germain⁴, Jeffrey M. Spraggins^{3,14,20}, Emma Lundberg²¹, Michael P. Snyder²², Neil L. Kelleher⁵ and Sinem K. Saka^{23,24} ✉

Yellow : Academic Institution USA

Magenta: Commercial Company

Cyan: Academic Institution Korea

Green: Academic Institution Europe (Sweden/EMBL)



Supplementary Table S4. Suggested minimum antibody metadata to be reported for multiplexed antibody-based assays. Proposed metadata to be collected on individual antibodies implemented in multiplexed antibody-based imaging assays. Validation should be performed according to previously published recommendations¹⁴⁴⁻¹⁴⁸. Also further see *Schapiro, Yapp, Sokolov et al., 2021*, for guidelines on a detailed metadata database structure encompassing all steps of multiplex experiments.

Suggested minimum antibody metadata	Rationale
UniProt Accession Number	Identifies the target protein.
Target Name	Provides a common name for the target protein.
RRID	Allows for universal identification of an antibody.
Antibody Name	Provides common name for the antibody used.
Host Organism and Isotype	Describes the species in which the antibody was raised (e.g. Mouse IgG1).
Clonality	Identifies the antibody as monoclonal or polyclonal.
Vendor	Provides information on the source of the antibody.
Catalog Number	Provides information on the source of the antibody.
Lot Number	Allows for monitoring of lot-to-lot variation.
Recombinant	Classifies the antibody as recombinant or not.
Dilution/Concentration	Provides a recommended usage (e.g. 1:100, 5 µg/ml).
Conjugate/Format	Offers details on the format and mode of detection (e.g. Fluorophore (AF647), Metal isotope (¹⁶⁴ Er), Oligo barcode and detection sequences. If custom conjugated, identify conjugation kit (Vendor, Catalog Number) or conjugation chemistry/protocol.
Downstream Platform Used	Identifies the imaging platform used (e.g. CODEX, MIBI, etc.).
Organ/tissue/cell line and sample preparation used for validation	Provides details on sample format and preparation protocol (e.g. fixation and antigen retrieval method if applicable).
Cycle Number	Identifies order of antibody immunolabeling or visualization for cyclic imaging methods.
Special Considerations (If known)	Describes whether a particular antibody is sensitive to imaging order (must go first or last in cyclic imaging methods) or is incompatible with other antibodies (steric hindrance, cross-reactivity with unconjugated antibodies of the same host species).
DOI for Validation Protocol	Details the protocol used to validate the antibody including positive and negative controls and example images. To enhance reproducibility, we recommend that all protocols be made publicly available via protocols.io.
ORCID ID of author	Identifies the individual who validated the antibody used in the assay.

HOW

Institutional Infrastructure eLab notebook

The screenshot displays the eLab notebook interface. On the left, a sidebar shows a list of projects under the heading 'Projects / GeoMx'. The main workspace is titled 'CODEX Runs' and contains a table with 23 rows of marker data. The table has columns for marker ID, antibody, fluorophore, and description. The interface includes a search bar, filters, and a 'Share' button.

ID	Marker	Antibody	Fluorophore	Description
2	CD31-001	AF-750	EP3095	endothelial cell marker/vascular differentiation marker
3	CD20-007	AF-750	L26	B cell marker/elevated in T cell lymphoma
4	PanK-019	AF-750	AE1/AE3	Tumor marker
5	CD21-032	Atto-550	EP3093	strongly expressed in mature B cells, follicular dendritic cells; weakly expressed in immature thymocytes / lymphocytes; lost in plasma cells during development
6	CD45RO-017	Atto-550	UCHL1	Memory T cells
7	E-cadherin-014	Atto-550	4A2C7	tumor suppressor protein, Epithelial cell marker, normally expressed at junctions between epithelial cells
8	K14-002	Atto-550	EPR3692	expressed in mitotically active basal layer cells, downregulated with carcinogenesis
9	Mac2/Galectin-3-035	Atto-550	M3/38	Microglia/actin filaments/myocardium/not specific for macrophages
10	B-catenin1-020	Atto-550	12F7	epithelial-mesenchymal transition (EMT)-associated marker
11	CD8-026	Atto-550	C8/144B	surface cell marker for cytotoxic T cells, also on NK cells, cortical thymic dendritic cells.
12	Ki67-047	Atto-550	B56	Mitotic activity marker/nuclear
13	CD11c-024	Cy5	118/A5	Dendritic cell marker/monocytes/granulocytes/subset of B cells/ and M2 macrophages
14	CD4-003	Cy5	EPR6855	surface cell marker for helper T cells, monocytes, macrophages, and dendritic cells
15	CD45-CX021	Cy5	D9M8I	Transmembrane protein-Lymphocytes
16	HLA-DR-033	Cy5	EPR3692	MHC classII cell surface receptor, expressed in activated T cells. HLA-DR is a marker of T cell activation and has been shown to be increased in autoimmune diseases and in patients with HIV infection.
17	CD68-015	Cy5	KP1	Macrophages and other mononuclear phagocytes
18	pHisH3-030	Cy5	HTA28	Proliferation marker/identifies cells undergoing mitosis
19	CD3e-045	Cy5	EP449E	T cell marker, detectable in early thymocytes
20	CD107a-006	Cy5	HGA3	General marker of NK cell activity and activated CD8+ T cells degradation/stimulation.
21	PDPN-BX019		NC-08	Mucin-like transmembrane protein widely expressed in various specializations throughout the body, colocalizes with nestin-primarily expressed in neural crest Specific lymphatic vessel marker
22	TFAM-BX029		18G102B2E11	Mitochondrial transcription factor A, ubiquitous cytoplasmic expression
23	LIF-BX006		M1506B09	granulocytes, alveolar cells type 2, smooth muscle cells, immune cell markers

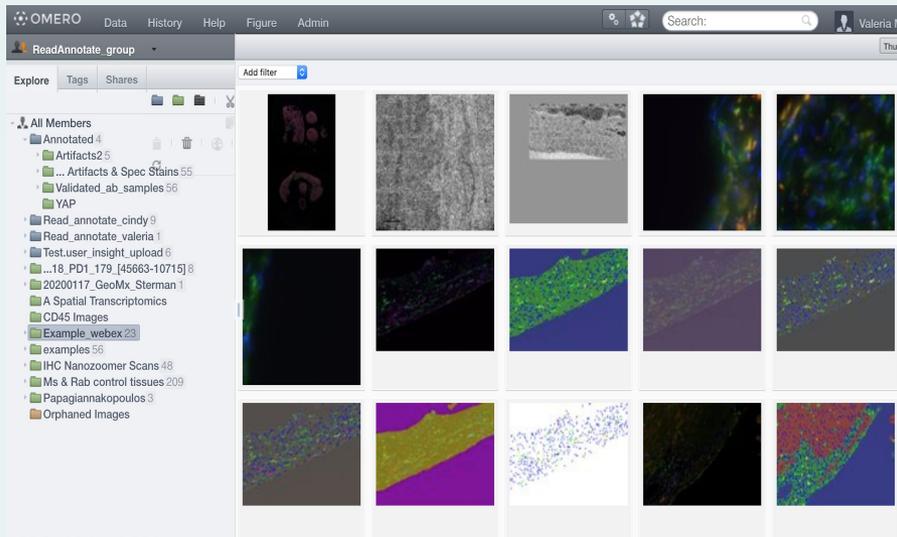
HOW

Institutional/Global Infrastructure Image Data Management Support



Group management - collaboration
Rendering different image formats:

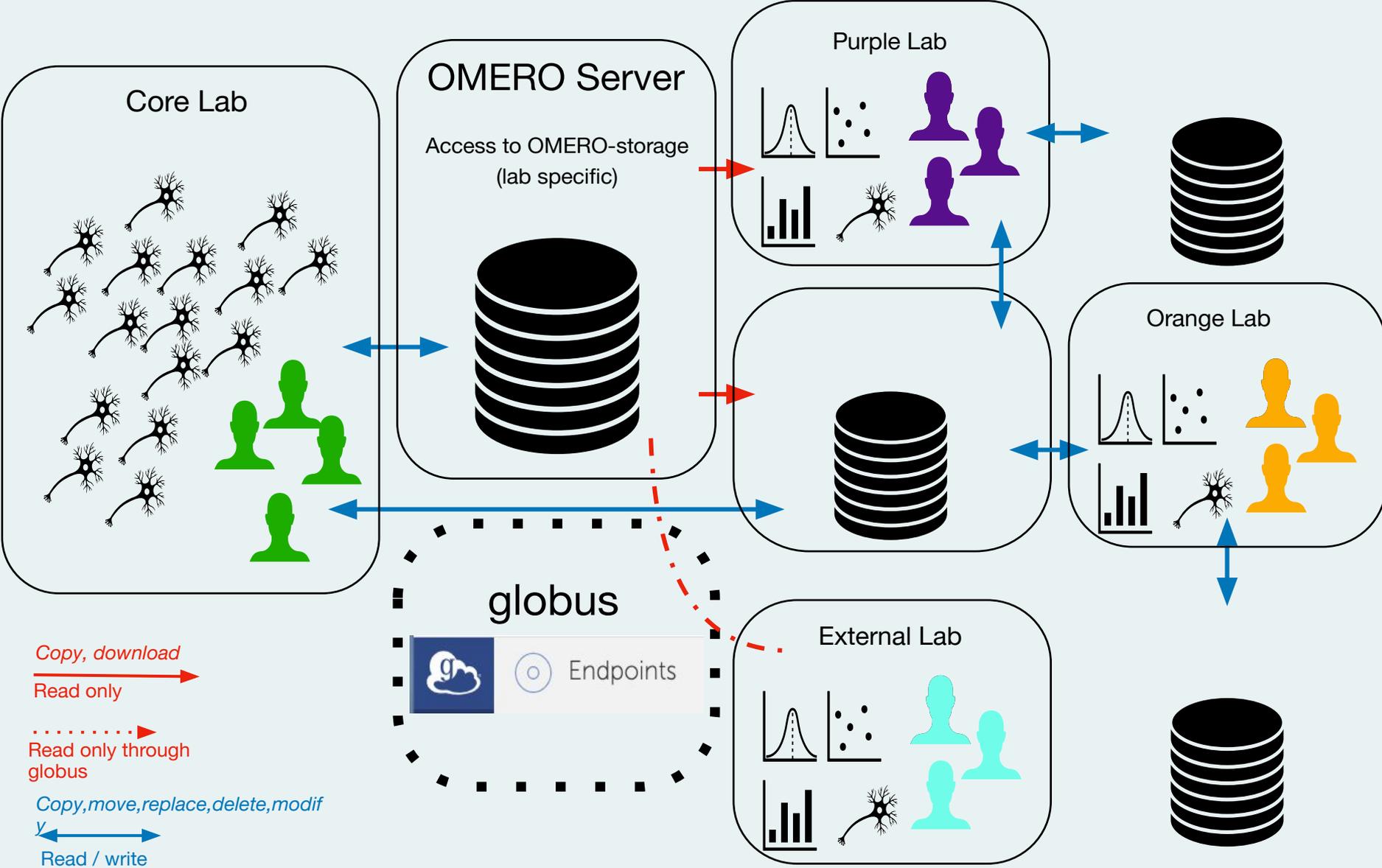
Track Experiments, preserve data provenance,
metadata is maintained and available, excellent for
data management plans.



Date	Total users	NYU groups	External Groups
20210914	473		13
20211025	501	179	18
20220706	627	190	24
20220920	666	193	26
20220923	647	194	26
20221006	651	195	26

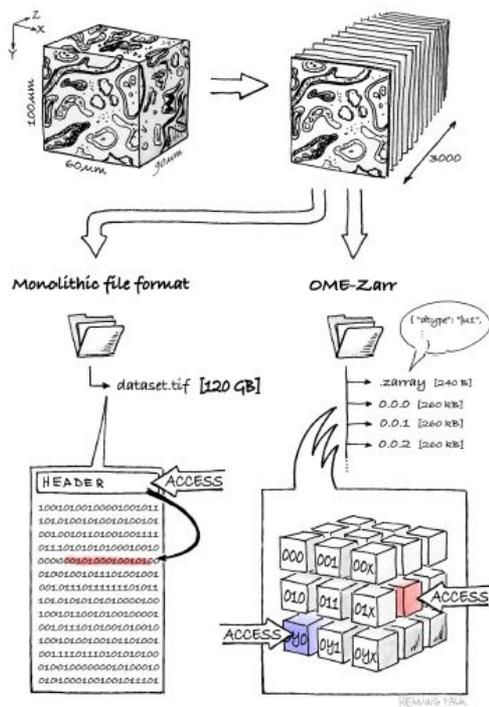
Institutional Infrastructure

image transfer workflows

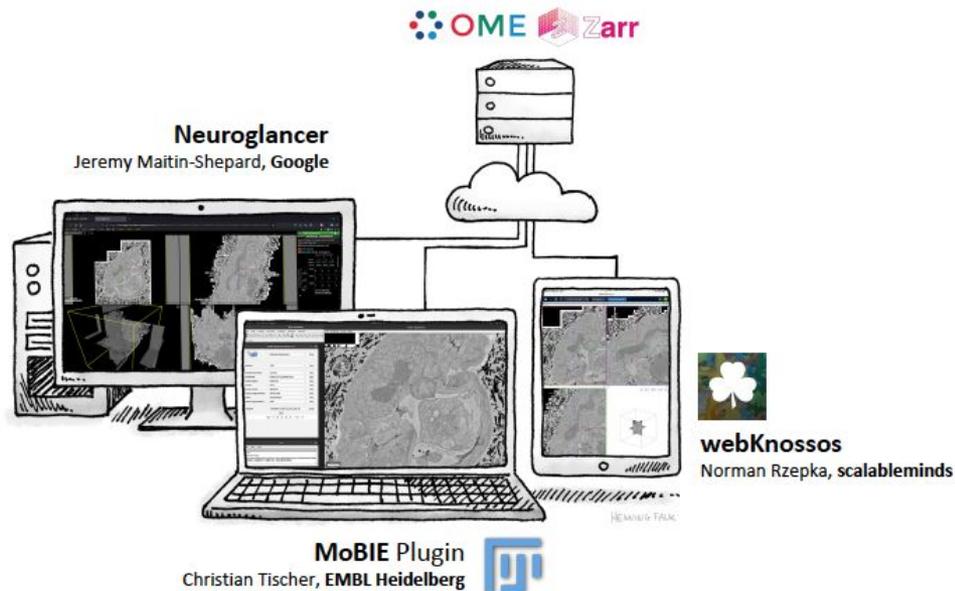


Global Infrastructure

Image Data-Metadata! Management work in progress




Zarr
Readable metadata
Well-defined layout
Configurable chunk size & compression



"Monolithic vs. chunked" by @DrHenningFalk
©2022 @NumFOCUS is used under a CC-BY 4.0 license.

"Multiple clients" by @DrHenningFalk
©2022 @NumFOCUS is used under a CC-BY 4.0 license.

SHARE RESEARCH ARTICLE



SARS-CoV-2 productively infects human gut enterocytes



Mart M. Lamers^{1,*}, Joep Beumer^{2,*}, Jelte van der Vaart^{2,*}, Kèvin Knoops³, Jens Puschhof², Tim I. Breugem¹, Raimond B. G. R...

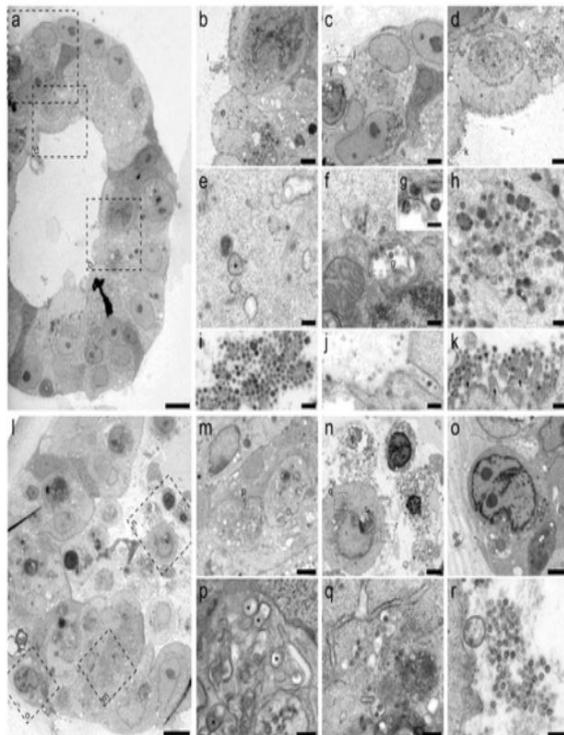
+ See all authors and affiliations



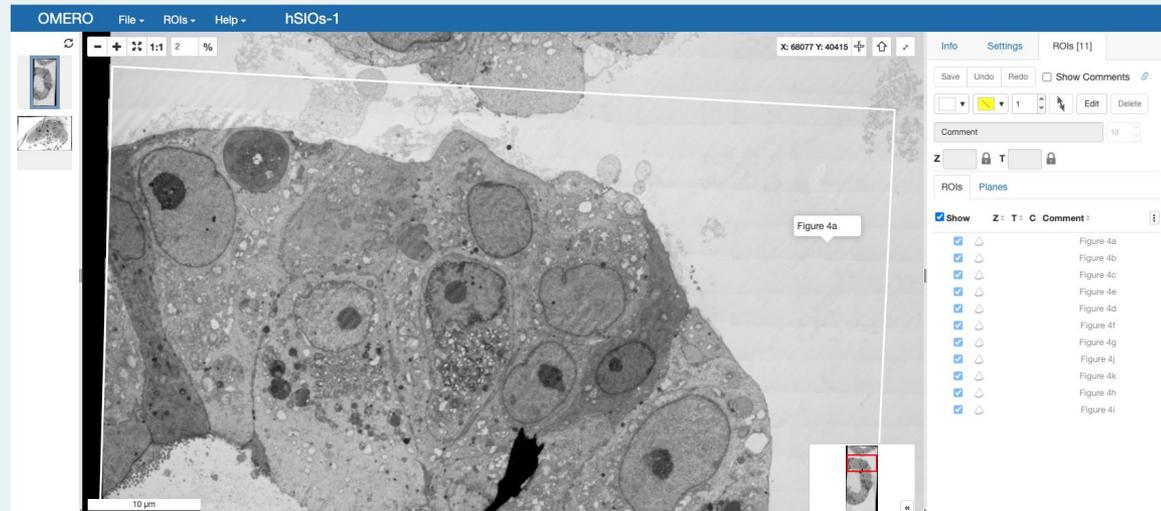
Science 01 May 2020:

eabc1669

DOI: 10.1126/science.abc1669



[IDR SarsCov2 organoids://idr.openmicroscopy.org/webclient/?show=image-9822151](https://idr.openmicroscopy.org/webclient/?show=image-9822151)



Moving Forward

Creation of Discipline Specific Resources

- Different sets template language for researchers with specific concerns, including: (1) Sharing human subjects data; (2) Managing and sharing data created at the Cores; (3) Describing data security protocols at our institution
- Cores are collecting data management plans in 'spreadsheet' format, outlining the data/formats and using RRID's for each Core lab
- Cores, Library, and IT collaborating on setting up OMEROPlus as the NYU GSoM non-clinical image repository

Moving Forward

Investigation of machine-actionable Data Management Plans (maDMPs)

- To make a DMS Plan machine-actionable means to format the information in such a way that it can be used to automate systems or processes related to the management and sharing of data
- Current strategy is to wait at least a year, assess submitted DMS Plans and possible processes for automation
- **Example:** If a DMS Plan states that a researcher will use a specific piece of software, IT automatically will receive an alert when the grant is funded so that they can begin working with the researcher sooner, avoiding unnecessary delays and frustrations.

Conclusions

- We are not in this alone
- Consider existing resource and community, both inside and outside your institution
- Set boundaries