National Library of Medicine
Network of the National Library of Medicine
National Center for Data Services

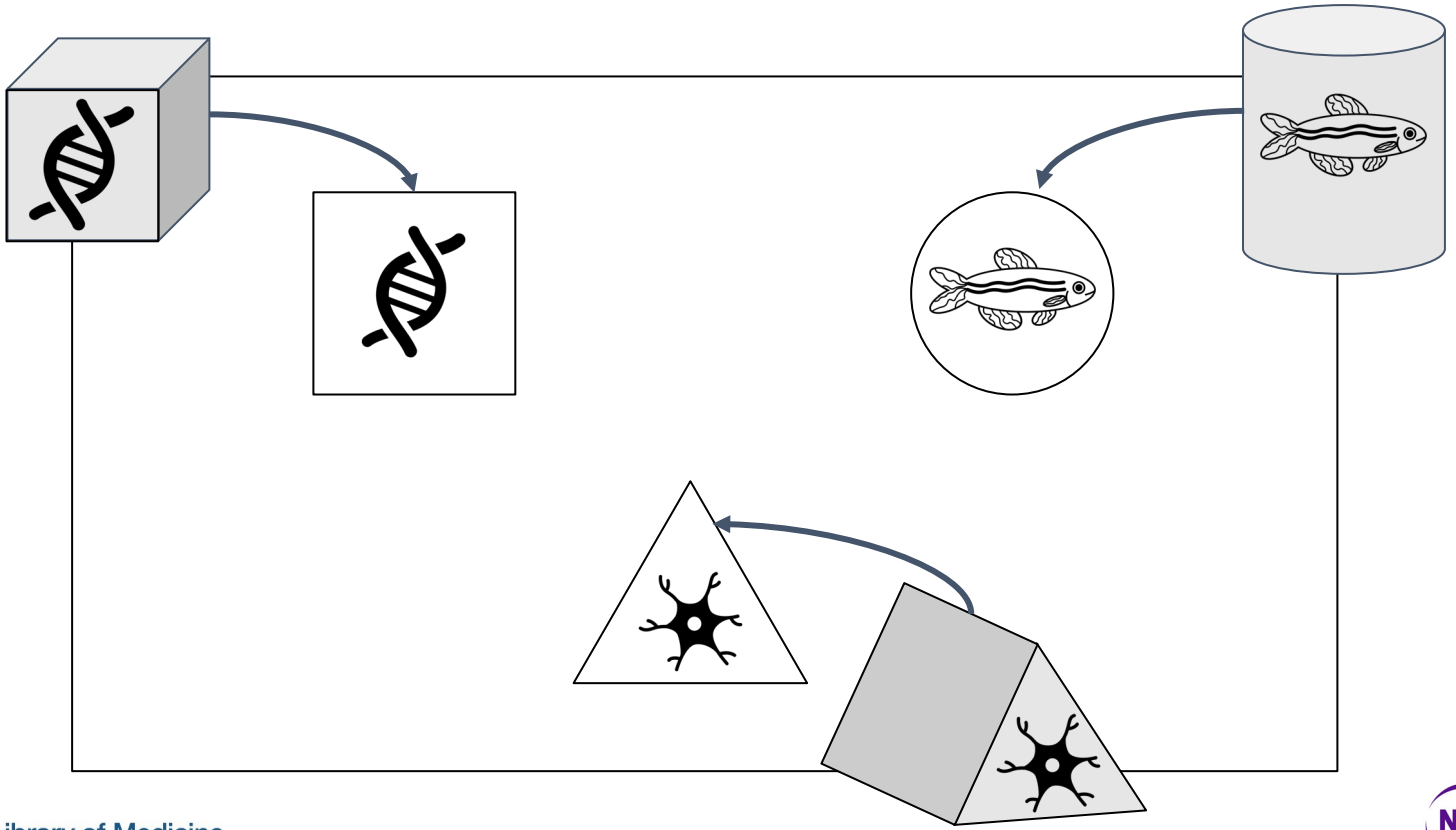
NYU Langone Health

Strategies for an imperfect data sharing ecosystem

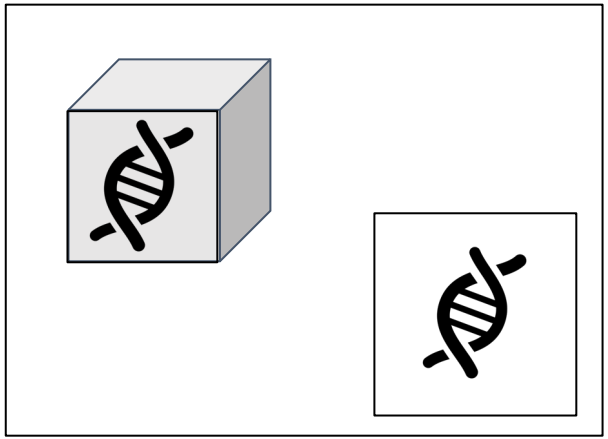
Generalist Repository Ecosystem Initiative Workshop
January 24, 2023

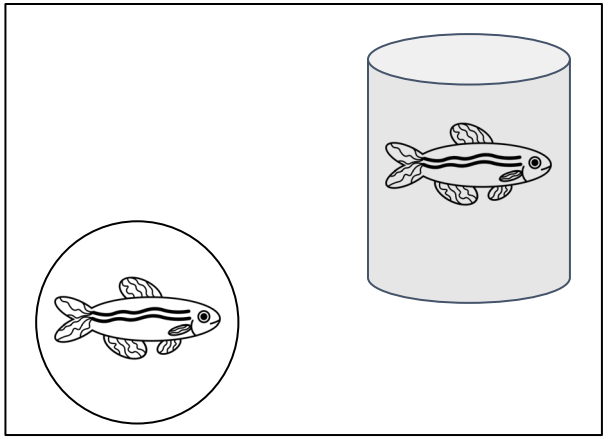
Alisa Surkis, PhD, MLS
Deputy Director/Vice Chair for Research, NYU Health Sciences Library
Director, NNLM National Center for Data Services

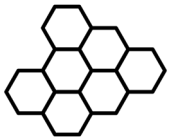
The ideal

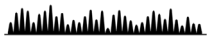


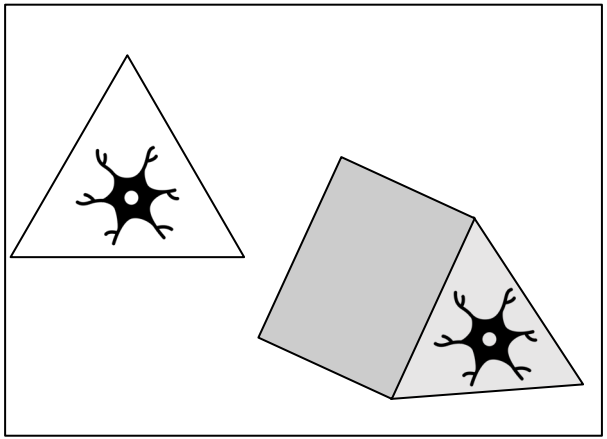
The reality

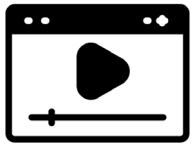




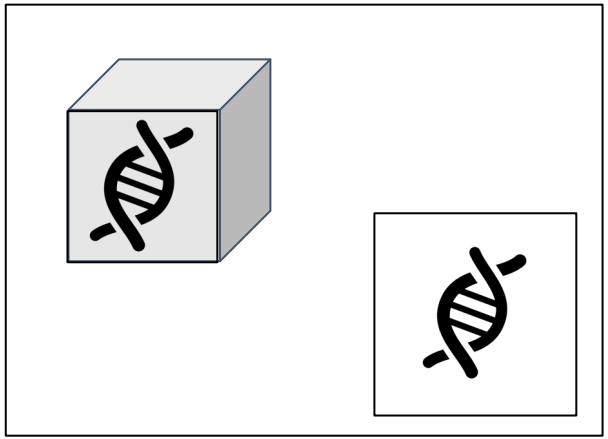


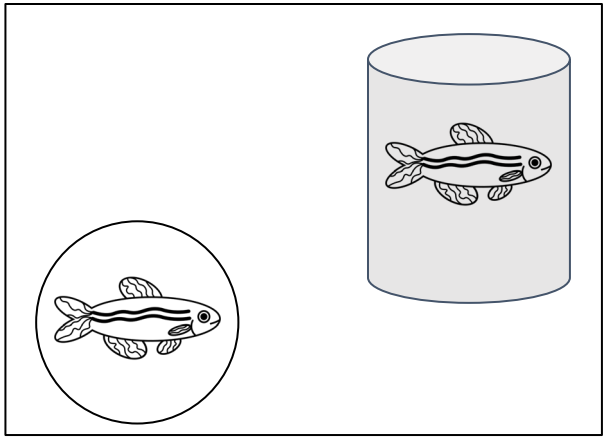


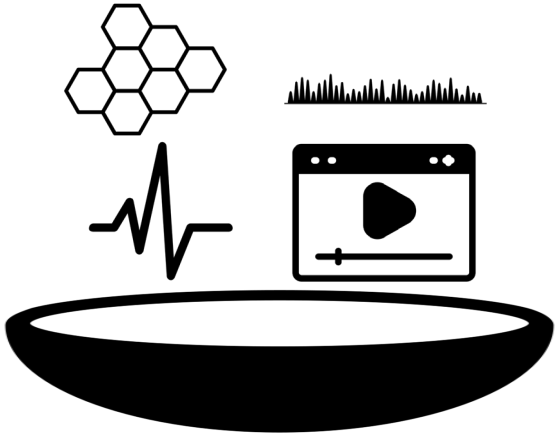


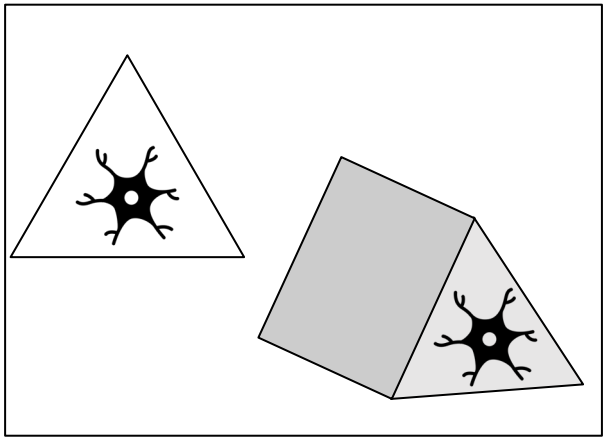


The role of generalist repositories







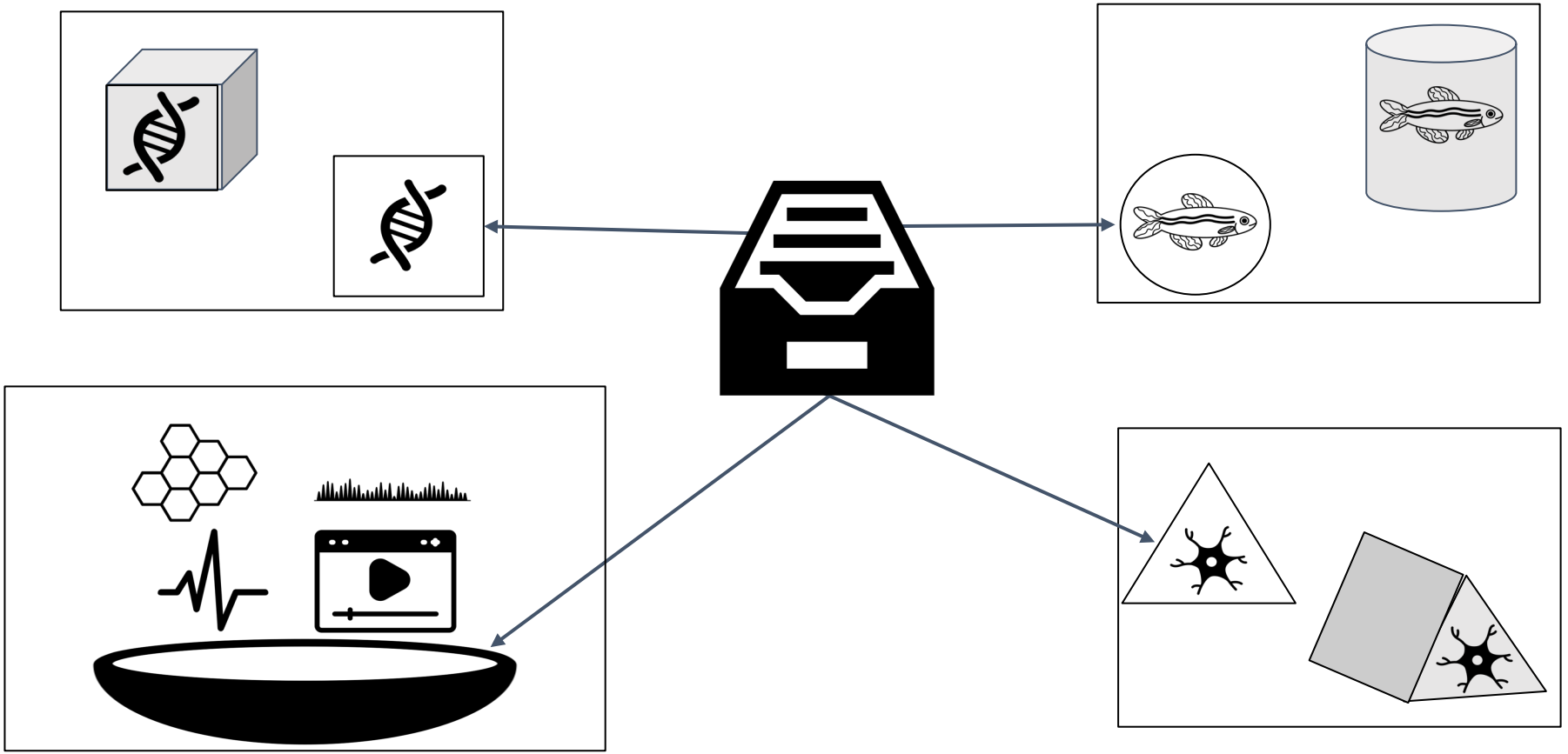


NYU Langone Health

But what about…

- Data discovery across repositories
- Sensitive data that can’t be shared through a repository
- Emergence of new discipline-specific repositories
- Studies with data in multiple repositories

The role of data catalogs



Data discovery

- **Data discovery across repositories**
- Sensitive data that can’t be shared through a repository
- Emergence of new discipline-specific repositories
- Studies with data in multiple repositories

Subject specific repositories have different organizing principles

RCSB PDB
PROTEIN DATA BANK

WormBase

National
COVID
Cohort
Collaborative

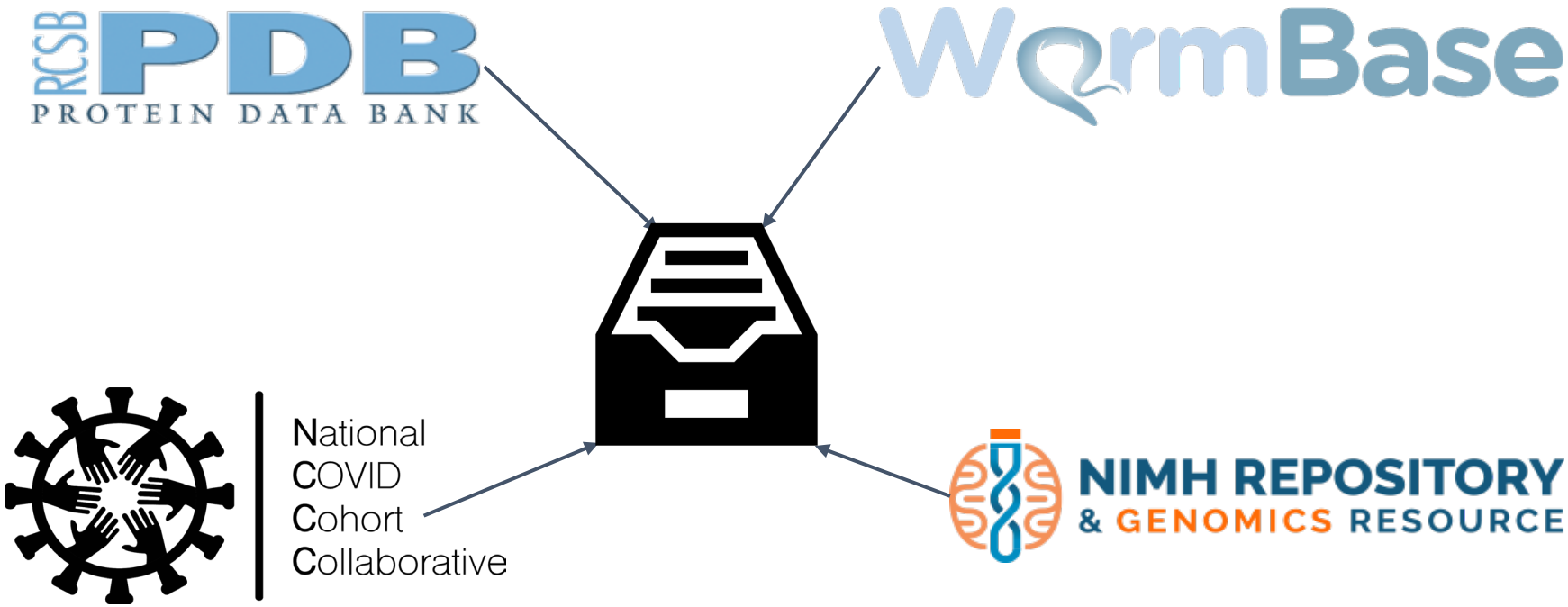
NIMH REPOSITORY
& GENOMICS RESOURCE

As do researchers

A given researcher may want –

- All types of data for a specific organism
- All types of data related to a specific disease
- A particular data type across species
- …

Metadata mapping



Metadata mapping

Sensitive data

- Data discovery across repositories
- **Sensitive data that can’t be shared through a repository**
- Emergence of new discipline-specific repositories
- Studies with data in multiple repositories

NYU Dataset

Syndemic Production among Emergent Adult Men: P18 Cohort Study

UID: 10089

Author(s): Perry Halkitis, Farzana Kapadia*
* Corresponding Author

Description

This ongoing prospective cohort study examines a wide range of individual, social and structural factors as predictors of risk and resilience for behavioral outcomes (e.g., substance use and sexual activity), mental health conditions (e.g., depression, anxiety) and biological outcomes (e.g., HIV seroconversion, viral and bacterial sexually transmitted infections) in a sample of emerging adult men based in the New York City metropolitan area. The study applies a theory of syndemic production to examine these overlapping health states in a cohort of racially/ethnically and socioeconomically diverse young men.

Investigators completed two study waves: 2009 and 2014 (Wave I; n = 600) and 2014–2019 (Wave II; n = 665). Wave I included men between the ages of 18 and19 year old, whereas participants in Wave II were men aged 22 to 23 years old. 41.2% (n = 274) of Wave II participants were retained from Wave I. Over the course of the study, participants completed surveys to provide information on sociodemographic

Access

Restrictions

Author Approval Required

Instructions

Please complete the Data Request Form to request access to this dataset. The information included in the form will be sent to the Corresponding Author who will evaluate your request for their data. The Corresponding Author may ask you to provide additional information if necessary.

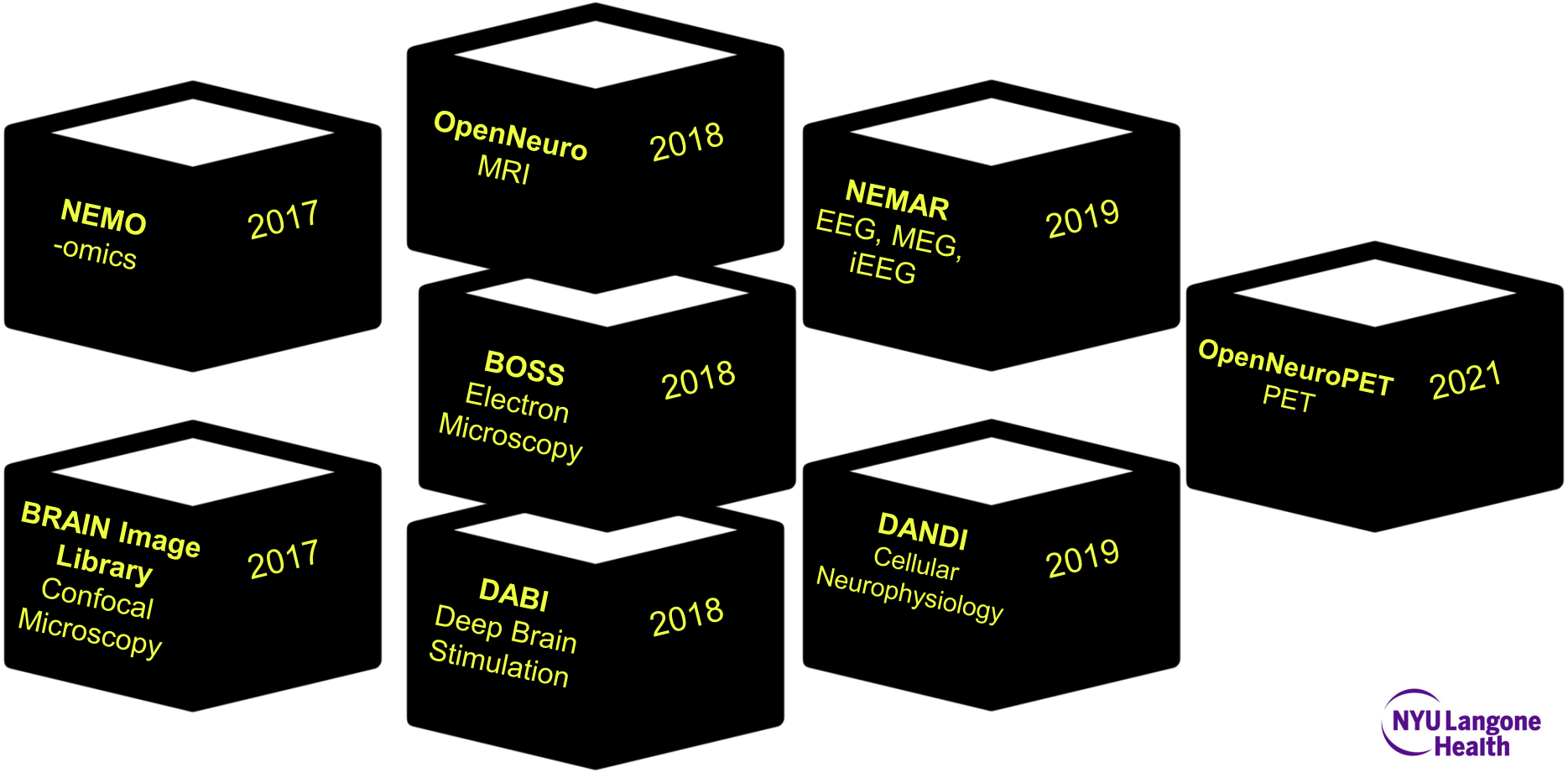
Access via Data Request Form

Form to request access

Changing landscape

- Data discovery across repositories
- Sensitive data that can’t be shared through a repository
- **Emergence of new discipline-specific repositories**
- Studies with data in multiple repositories

NIH BRAIN Initiative funded data archives



NYU Langone Health

Innate and plastic mechanisms in auditory cortex for maternal behavior

 Version 6 Posted on 2020-10-26 - 15:37 authored by Jennifer Schiavo

Data from Schiavo et al. (2020) - Innate and plastic mechanisms in auditory cortex for maternal behavior

AUTHORS (1)

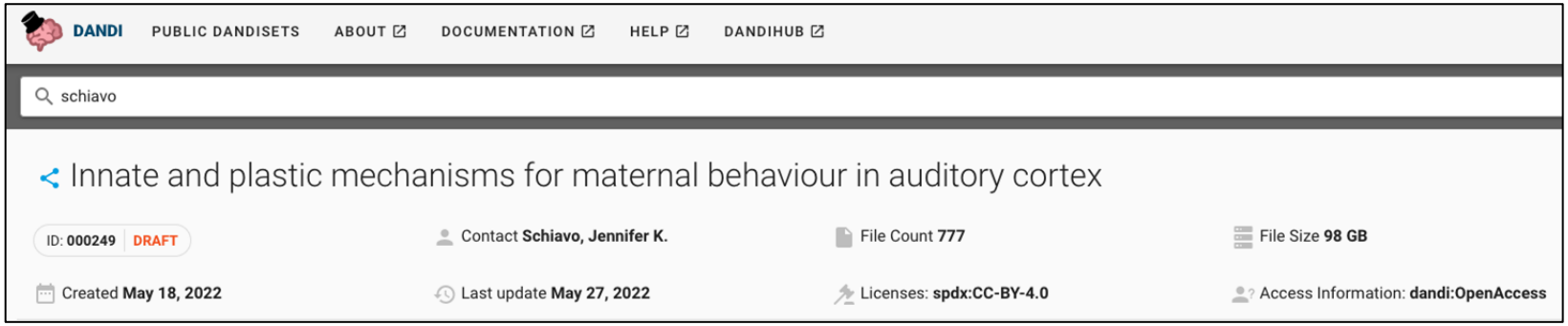
JS Jennifer Schiavo

CATEGORIES

- Behavioural neuroscience

KEYWORDS

Maternal Behavior
auditory cortex (AC)
Plasticity
oxytocin
Behavioral Neuroscience



Assets Summary

Species

Mus musculus - House mouse

Approach

microscopy approach; cell population imaging

Data Standard

Neurodata Without Borders (NWB)
RRID:SCR_015242

Number Of Subjects

54

Variable Measured

OpticalChannel

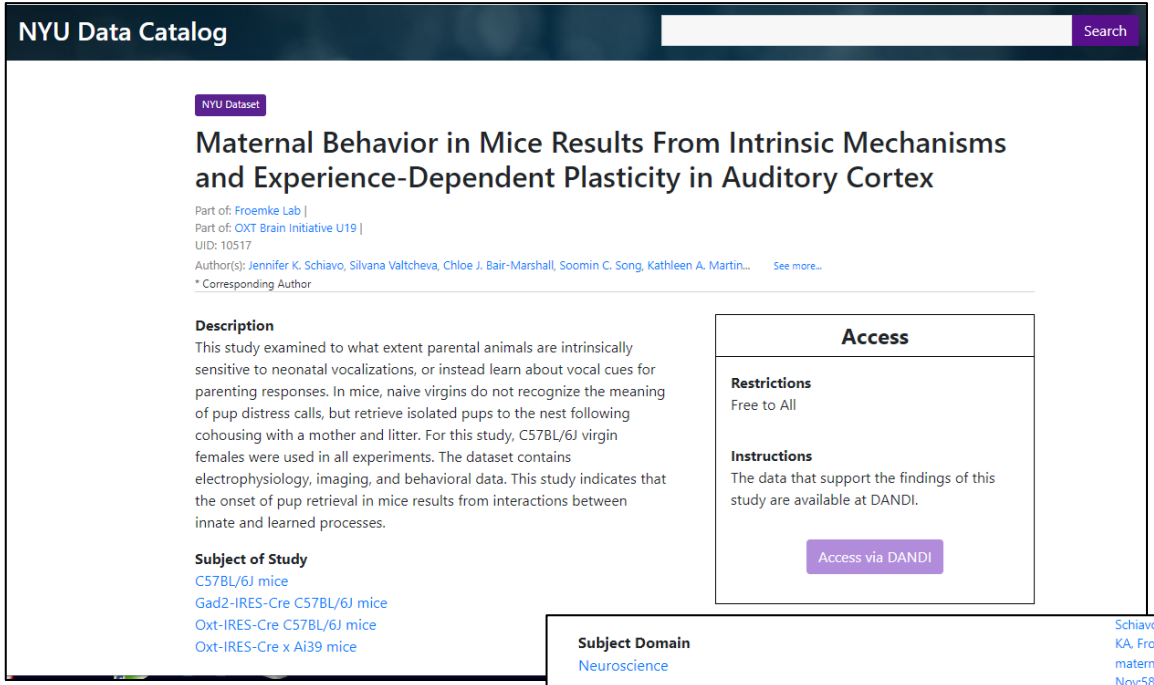
ImagingPlane

TwoPhotonSeries

Measurement Technique

surgical technique

two-photon microscopy technique



Subject Domain
Neuroscience

Keywords
Animal Behavior
Animal Model
Animals
Auditory Cortex
Calcium
Electrophysiology
Mice
Neurons
Neurosciences
Oxytocin

Schiavo JK, Valtcheva S, Bair-Marshall CJ, Song SC, Martin KA, Froemke RC. Innate and plastic mechanisms for maternal behaviour in auditory cortex. Nature. 2020 Nov;587(7834):426-431.

Data Type
Electrophysiological
Behavioral
Imaging

Equipment Used
Avisoft-Bioacoustics CM16/CMPA
Condenser ultrasound microphone

Leica VT1000 S
Vibrating blade microtome

MultiClamp 700B Microelectrode Amplifier
Computer-controlled amplifier

But what about…

- Data discovery across repositories
- Sensitive data that can’t be shared through a repository
- Emergence of new discipline-specific repositories
- **Studies with data in multiple repositories**

NYU Data Catalog

Search

NYU Dataset

Hyperactive CDK2 Activity in Basal-like Breast Cancer Imposes a Genome Integrity Liability

Part of: Possemato Lab |
UID: 10580

Author(s): Vladislav O. Sviderskiy, Lili Blumenberg, Elizabeth Gorodetsky, Triantafyllia R. Karakousi, Nicole Hirsh... See less...
Samantha W. Alvarez, Erdem M. Terzi, Efiyenia Kaparos, Gabrielle C. Whiten, Shakirah Ssebyala, Peter Tonzi, Hannan Mir, Benjamin G. Neel, Tony T. Huang, Sylvia Adams, Kelly V. Ruggles, Richard Possemato*
* Corresponding Author

Description

This study showed that basal-like breast cancer (BLBC) is sensitive to suppression of iron-sulfur cluster (ISC) biosynthesis and identified DNA polymerase epsilon (POLE) as an ISC-containing protein that underlies this phenotype. This dataset includes primary RNA sequencing data and other primary data, as well as supplementary data tied to the publication. The supplementary data contains tables, which includes short hairpin RNA (shRNA) information, shRNA knockdown efficacy, and normalized population doublings upon shRNA expression relative to non-targeting shRNA, correlations between ISCU suppression sensitivity to suppression sensitivity of other genes across CCLE breast cancer cell lines, correlation of POLE sensitivity with gene expression in CCLE cell lines, RNA sequencing data for MDA-MB-231, CAMA-1, and MCF7 cells expressing shRNAs targeting POLE and either a vector control or cDNA POLE rescue, and primary phosphoproteomic data. The data indicate that CDK2 hyperactivity is a genome integrity vulnerability exploitable by targeting POLE.

Access

Restrictions
Free to All

Instructions
Primary RNAseq data available at GEO and other primary data can be accessed at Mendeley Data.

Access via GEO
Primary RNAseq data

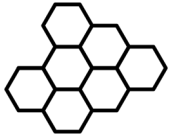
Access via Mendeley Data
Other primary data

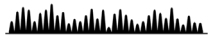
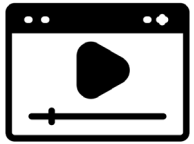
Where do librarians fit in?

Providing support, resources, and/or training to:

- Select an appropriate data repository
- Organize data and assign metadata
- Identify appropriate standards
- Effectively locate appropriate data



Trainings to support DMSP policy

- NIH DMSP Immersion Workshop (2/17/23)
  - 6 hour deep dive on policy and each component of the required DMSP
- NIH DMSP On-demand Webinars
  - 5 webinars covering the policy and practitioner perspectives on different aspects of support
- NIH Data Management and Sharing Plan Overview
- Creating Data Management Plans with DMPTool (2/15/23)

Other data trainings

- Fundamentals of Health Sciences Research Data Management
- Fundamentals of Health Data Science (Sep-Nov 2023)
- Introduction to Research Data Services
- Inclusive Data & Data Accessibility Mini-Course (April 2023)

https://www.nnlm.gov/about/centers/ncds

National Library of Medicine
Network of the National Library of Medicine

National Center for Data Services

Resources

- NIH Data Management and Sharing Policy Toolkit
- Data Repository Finder (in process)
- Data Glossary

Internship Program

- 10 week summer internship in data librarianship for 12 graduate students or recent grads from communities that are underrepresented in the profession

https://www.nnlm.gov/about/centers/ncds

Questions?