

Show me the data! Data sharing practices demonstrated in published UMass Chan research

Tess Grynoch and Kimberly MacKenzie
Lamar Soutter Library, UMass Chan Medical School

Abstract

In the interest of making data findable, accessible, interoperable, and reusable (FAIR), the National Institutes of Health (NIH) will institute a new Research Data Management and Sharing Policy in January 2023. This policy will require researchers applying for NIH funding to submit a Data Management and Sharing Plan. As 63% of grant dollars received by UMass Chan Medical School researchers comes from the NIH, we explored whether UMass Chan researchers are currently sharing data associated with their published research. PubMed was searched for articles published in 2019 with a UMass Chan researcher as either the first or last author. These articles were then examined for evidence of original or reused data, the type of data, whether the article stated that data was available, and where and how to find that data. Of the 713 articles found with a UMass Chan first and/or last author, 535 were research articles that produced original data or reused existing data. Of those articles with data, 57.4% (307) were NIH funded, 17.2% (92) had a data availability statement, and 10.8% had data deposited in a repository. This poster will highlight how the results of the study will inform our library services for researchers and provide tips on making UMass Chan data FAIR.

New NIH Data Management Plan and Sharing Policy Requirements

What needs to be in a data management plan?

Data Type

What kind of data? How is it processed? What file types?

Tools/Software/Code

What was used? What's needed to replicate the results? Will the code be shared?

Data Standards

Data format, data dictionaries, etc. Does your field have common data standards?

Access, Distribution, Reuse

Who can access the data? What privacy protections are there?

Preservation/Access

Where will data be shared? Will it be findable? What format will it be shared in? When will it be available?

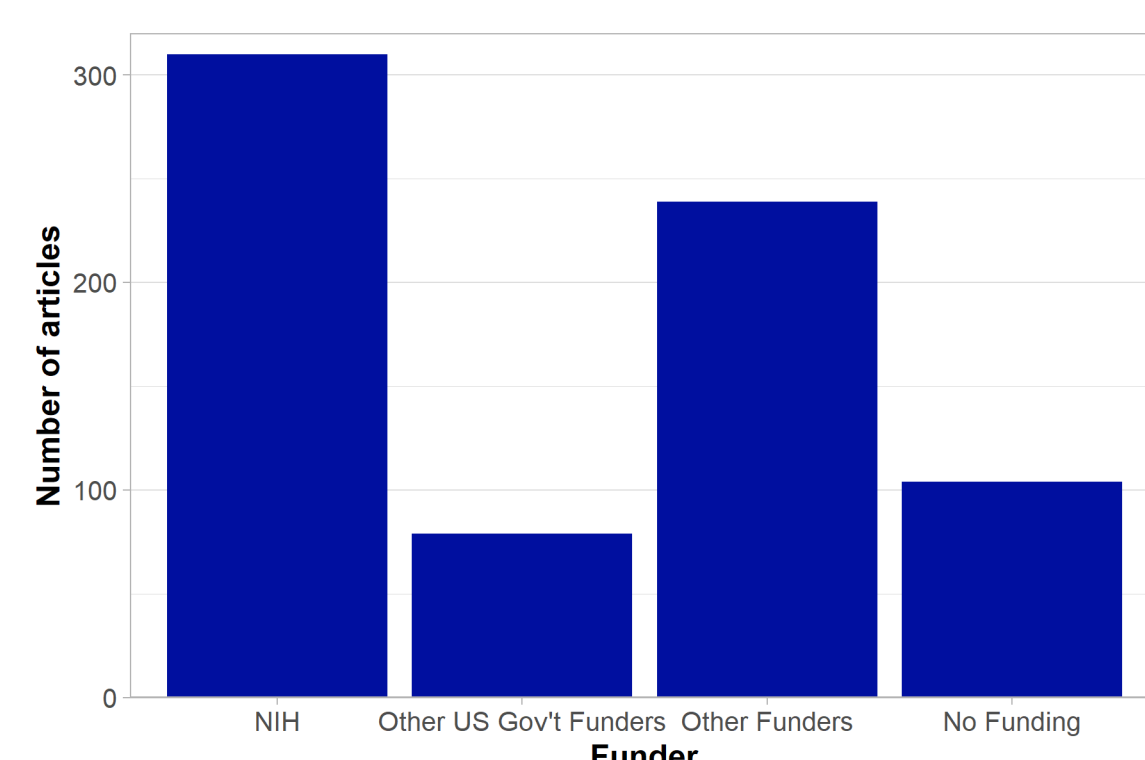
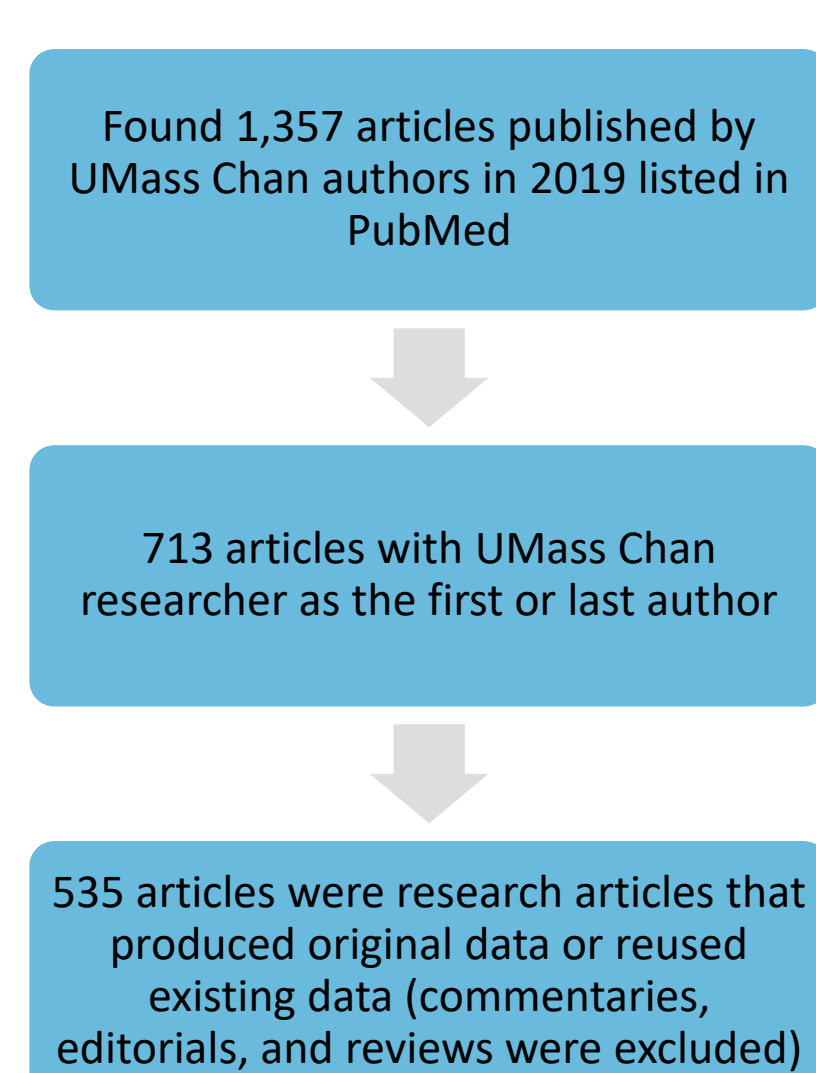


Figure 1. Funding sources for articles published in 2019.

Methods



For each article with original and reused data we noted:

- Which section of the article notes how to access the data
- Article funders
- Whether the article was open access at the time of publication

For articles with original research data, we noted:

- The number of datasets
- Type of data (PCR, microscope images, etc.)
- Where the data is stored
- Does the dataset have a permanent identifier (ex. DOI)
- Data file format
- License attached to the data

For articles with reused research data, we noted:

- Whether the reused data was the author's or someone else's

Results

Table 1. Number of articles with original data or reused data or both.

Articles with...	Number of articles
Original data only	314
Reused data only	174
Both original and reused data	47



Figure 2. Number of articles that had original data sharing mentioned in the data availability statement and whether they were published under an open access license.

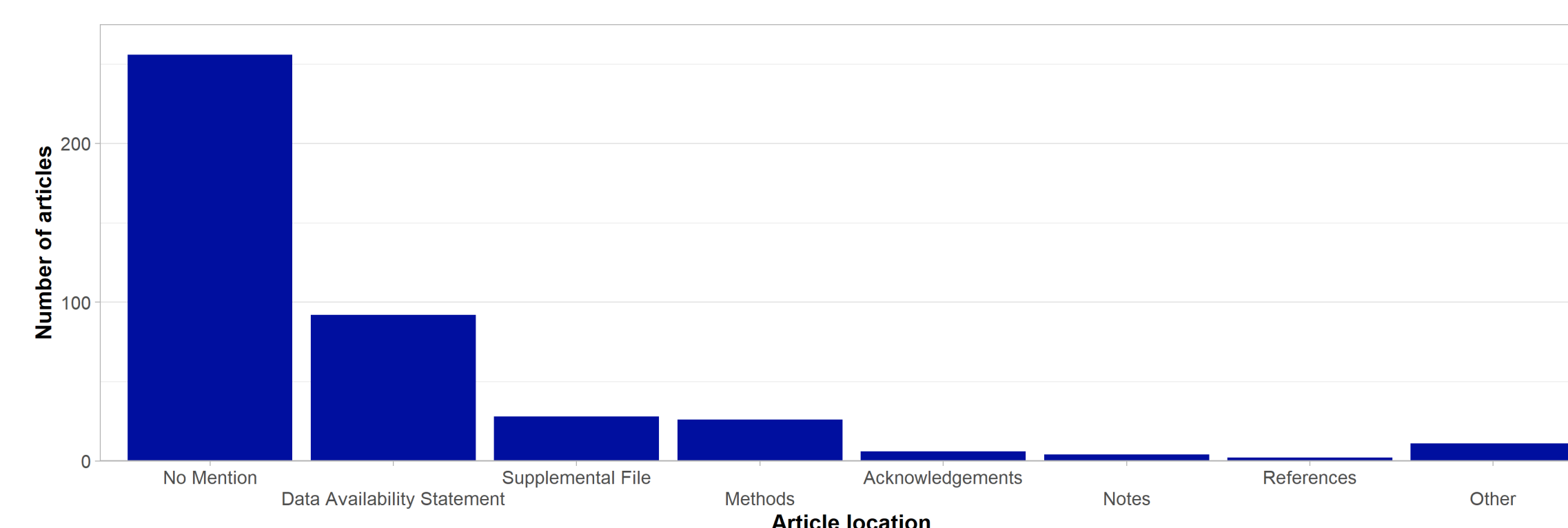


Figure 3. Location in article of original data sharing mention.

Results (continued)

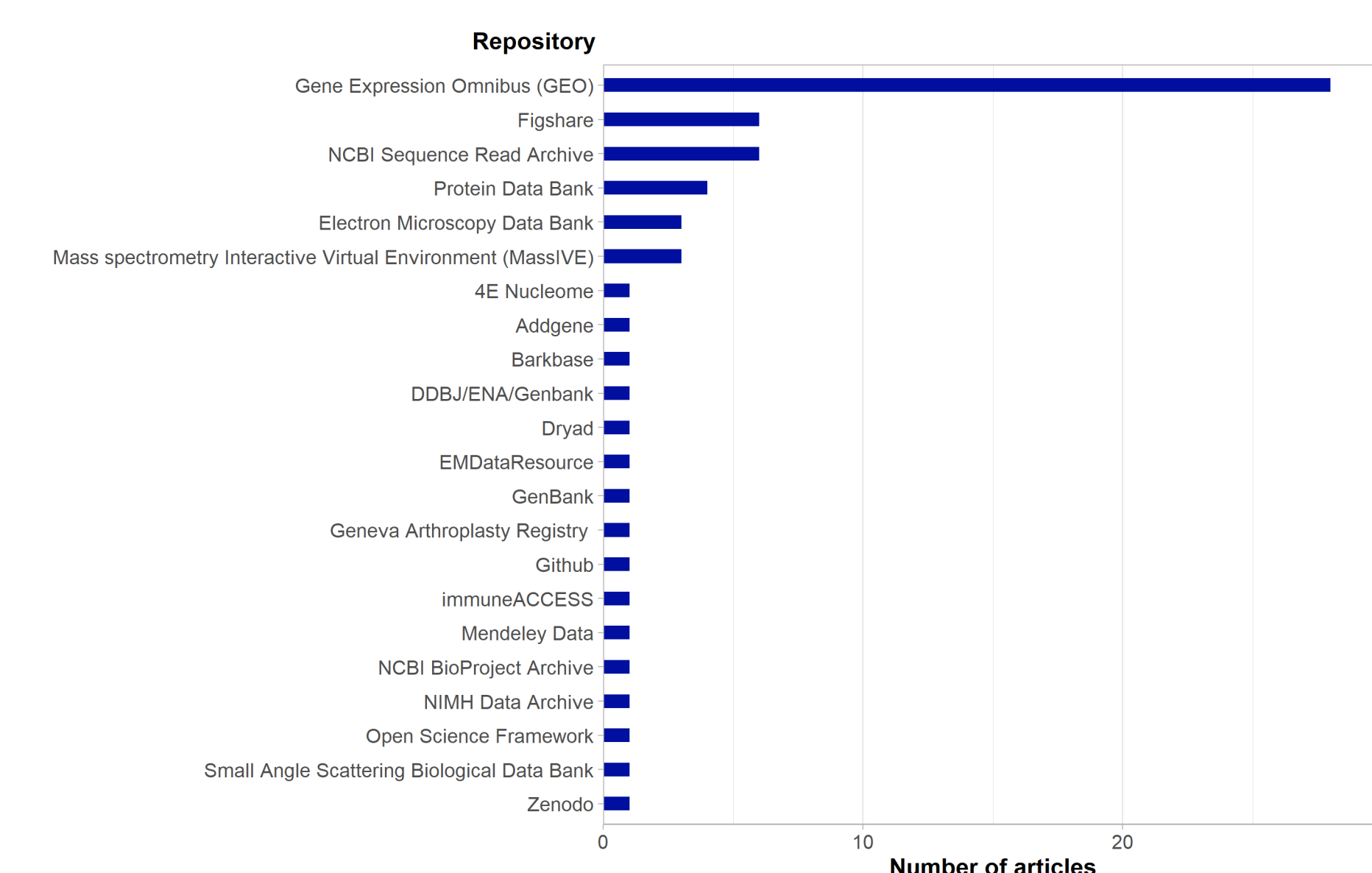


Figure 4. Repositories where original data was shared. DDBJ = DNA Data Bank of Japan; ENA = European Nucleotide Archive; NCBI = National Center for Biotechnology Information; NIMH = National Institute of Mental Health.

Type of data in top 6 repositories

- Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo/
Microarray, next-generation sequencing, and other forms of high-throughput functional genomics data
- Figshare, figshare.com/
Generalist repository, accepts all data types
- NCBI Sequence Read Archive, trace.ncbi.nlm.nih.gov/Traces/sra/
Raw sequence data from next-generation sequencing technologies
- Protein Data Bank, www.rcsb.org/
3D structure data for proteins, DNA, and RNA
- Electron Microscopy Data Bank, www.ebi.ac.uk/emdb/
Electron cryo-microscopy maps and tomograms of macromolecular complexes and subcellular structures
- Mass spectrometry Interactive Virtual Environment (MassIVE), massive.ucsd.edu/
Mass spectrometry data

Acknowledgements

Thank you to Lisa Palmer and Rebecca Reznik-Zellen for designing and completing the pilot study which informed this project.

References

- Go FAIR. (2021). FAIR principles. Retrieved from <https://www.go-fair.org/fair-principles/>
- NIH. (2020). Final NIH Policy for Data Management and Sharing. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
- Utrecht University. (2021). How to make your data FAIR. Retrieved from <https://www.uu.nl/en/research/research-data-management/guides/how-to-make-your-data-fair>

Make Your Data FAIR

Findable

- Deposit the dataset(s) into a trusted and searchable repository
- Make sure the dataset is assigned an identifier such as a Digital Object Identifier (DOI)
- Include clear and complete metadata (data about your data)
- Does your discipline have metadata standards to follow?
- Basic metadata: author, title, abstract, keywords, measurement type, data format, data, size of dataset (the more thorough the metadata, the more findable the dataset)

Accessible

- Metadata should be openly available
- Have clear guidelines on how data can be shared and who to contact about the data (accessible doesn't have to mean completely open!)
- Use open, non-proprietary file formats (.csv instead of .xls, .txt instead of .doc)
- Make clear what software is needed to open the data

Interoperable

- Make programming scripts accessible along with data
- Share "tidy-data" that is analysis ready
- Create a README file and/or a data dictionary that clearly explains what the data is, which figures/tables it is associated with, what variables mean, and describes how the data was analyzed.

Reproducible

- Make sure data is well described
- Give a license that clearly indicates how data can be reused (example: Creative Commons)
- Include license information in the metadata

Questions about any of the FAIR data tips? Contact us!



UMass Chan
MEDICAL SCHOOL

Lamar Soutter Library

Education • Research • Health Care

Empowering the future. Preserving the past.