

Exploring a Generalist Repository for NIH-Funded Data

Office of Data Science Strategy (ODSS), National Institutes of Health (NIH)



National Institutes of Health
Office of Data Science Strategy

Table of Contents

Key Takeaways 1

Project Outcomes 2

 NIH Figshare Instance Case Studies 3

Key Takeaways

Overall, this pilot demonstrated that generalist repositories can help fill gaps in the repository landscape for NIH-funded researchers and validated that there is use for a generalist repository that can accept heterogeneous and large files.

ODSS learned three key lessons from the pilot:

- 1. Generalist repositories are growing.** More researchers are depositing data and more publications are linking to generalist repositories.
- 2. Researchers need more education and guidance** on where to publish data and how to effectively describe datasets with detailed metadata.
- 3. Metadata enhancement enables greater discoverability.** Expert metadata review proved to be one of the most impactful and unique features of the pilot Instance.

When compared to uploads indicating NIH funding to the main Figshare repository, the NIH Figshare Instance had uploads with titles that are two times in length and descriptions that are three times in length.

While the [NIH Figshare Instance](#) is now an archive, the data is still discoverable and reusable.

Learn More about Generalist Repositories

NIH encourages researchers to use a generalist repository that meets [OSTP criteria](#) when a domain-specific or institutional repository is not available.

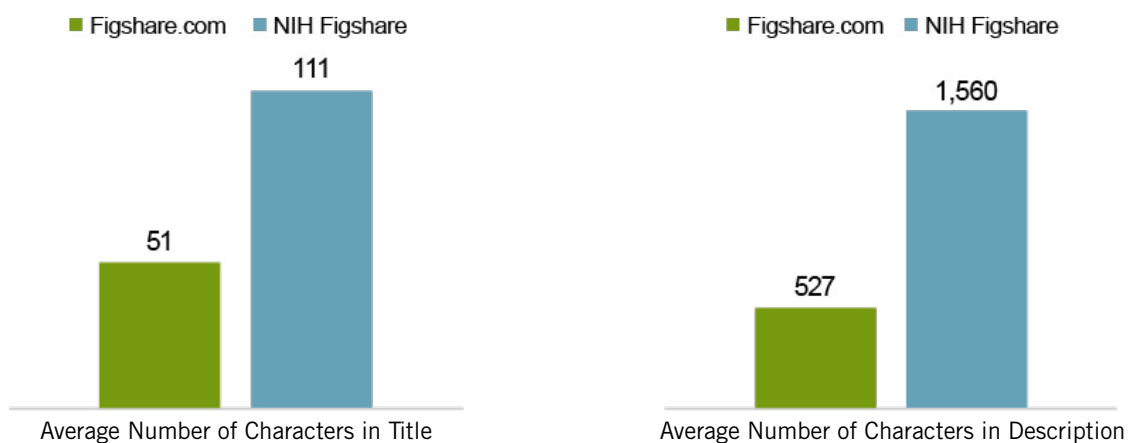
Read Dr. Gregurick's blog post offering "[Some Insights on the Roles and Uses of Generalist Repositories.](#)"

See a comparison of generalist repositories at [FAIRsharing.org](#).

See how generalist repositories are helping researchers [share COVID-19 data](#).

Learn how generalist repositories are [enhancing data discoverability and reuse](#).

When comparing NIH Figshare to Figshare.com, NIH Figshare has titles that are two times longer and descriptions that are three times longer



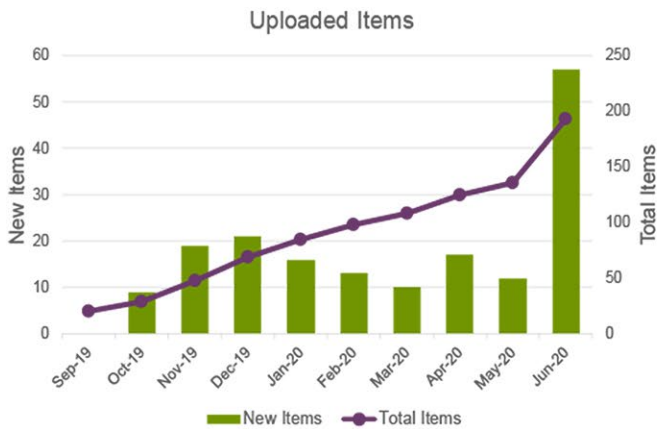
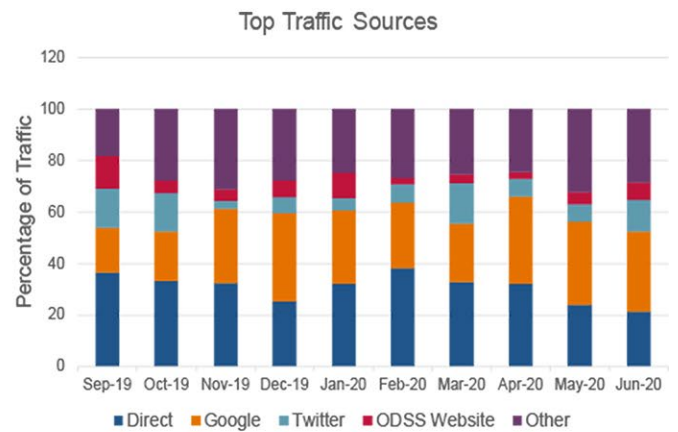
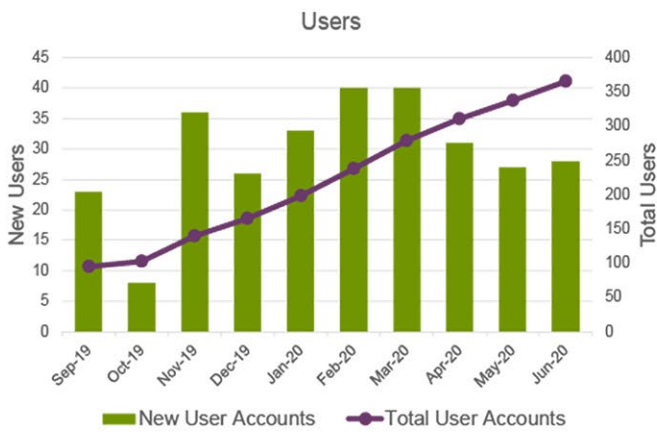
Data and code uploaded to [NIH.figshare.com](#) vs. [Figshare.com](#) users who indicated NIH funding and uploaded datasets or code. July 23, 2019–July 15, 2022.

Want to learn more about this project? Watch a [video](#) of Figshare's founder and CEO, Mark Hahnel, Ph.D., discuss project outcomes and lessons learned, as well as his thoughts on the future of data sharing, or read a summary in this [NIH Record article](#).

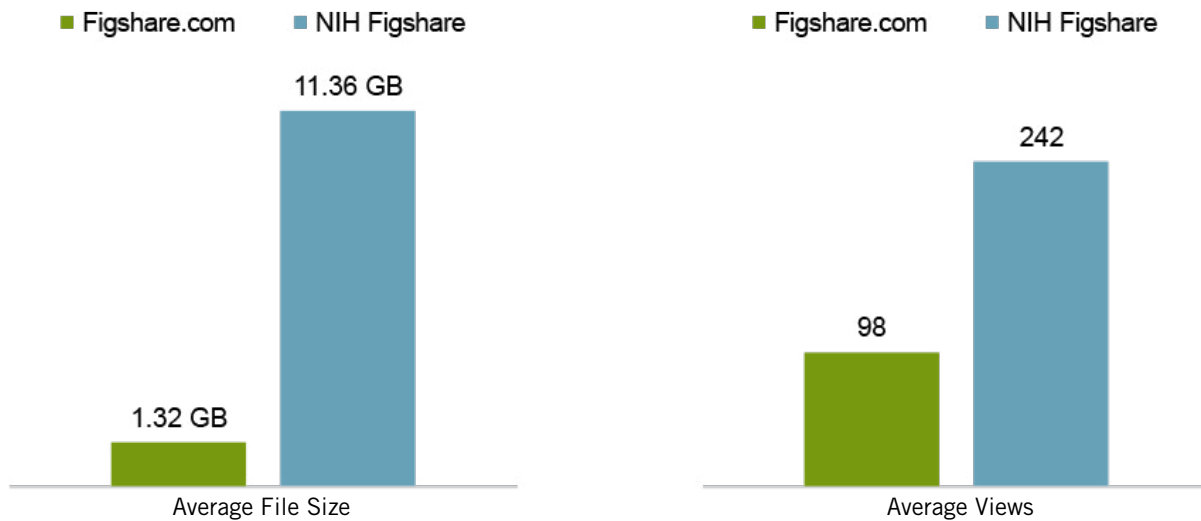
Project Outcomes

Over the course of the one-year pilot, NIH assessed how the NIH Figshare Instance was meeting researchers' needs and how it was making an impact on data sharing and discovery. Key outcomes of the assessment are below—366 users, 209 uploaded items, 1,499 GB of storage used.

- The NIH Figshare Instance received 30,167 total page views.
- Email campaigns and webinars resulted in 31% of new users.
- 22% of data in the NIH Figshare Instance came from intramural researchers across 9 NIH Institutes and Centers.
- 72% of items uploaded were datasets; other item types include code, software, figures, multimedia files, workflows, and online resources.
- The repository Instance includes data funded by 22 different NIH Institutes and Centers.
- 29 different public items in NIH Figshare have each been cited at least once, and five items have been cited twice.
- When compared to uploads indicating NIH funding to the main Figshare repository, the NIH Figshare Instance had files that are eight times in size and with two-and-one-half times as many views.



When comparing NIH Figshare to Figshare.com, NIH Figshare has files that are eight times larger, with two and a half times as many views



Data and code uploaded to [NIH.figshare.com](https://www.figshare.com) vs [Figshare.com](https://www.figshare.com) users who indicated NIH funding and uploaded datasets or code. July 23, 2019–July 15, 2022.

NIH Figshare Instance Case Studies

Five case studies take a deeper dive into the ways the Instance made an impact:

- **Flexible organization for large datasets:** [Storing and sharing x-ray scattering data on the NIH Figshare Instance](#)
 - This case study explores the process of James Fraser and Michael Thompson, researchers in the Fraser Lab at the University of California, San Francisco, uploading x-ray scattering data to the NIH Figshare Instance.
- **Using an API to upload large datasets:** [Making the NIH Figshare Instance part of the research lifecycle: a case study of sharing single cell databases in the Carpenter Lab at The Broad Institute of MIT and Harvard](#)
 - This case study examines how Gregory Way, postdoctoral associate in the Carpenter Lab at The Broad Institute of MIT and Harvard, and his colleagues published single cell databases on the NIH Figshare Instance using Figshare's API.
- **Collections of heterogenous data:** [Using the NIH Figshare Instance to make fMRI and eye movement data associated with a publication openly available](#)
 - This case study explores how Michal Ramot, NIH intramural researcher and visiting fellow at the National Institute of Mental Health, and her colleagues published a collection of neuroimaging research in the NIH Figshare Instance.
- **Reusing and sharing a non-traditional output:** [Using the NIH Figshare Instance to share a cholesterol calculator for reuse and further collaboration and development](#)
 - This case study showcases how Dr. Alan Remaley and Maureen Sampson, intramural researchers at the National Heart, Lung, and Blood Institute, used NIH Figshare to share equation calculators they created for a novel way to calculate low-density lipoprotein cholesterol.

- **Sharing materials supporting a publication from across repositories: [Using NIH Figshare to collect supplementary material associated with a publication](#)**
 - This case study demonstrates how Yosuke Tanigawa and his colleagues at Stanford University used NIH Figshare to share supplementary data supporting a NHGRI-funded publication in *PLoS Genetics* and grouped it in a collection that also includes *PLoS* supplementary materials for the same publication, which were also published on Figshare.

[Read 10 use cases from the pilot](#)

For technical questions about the NIH Figshare archive or using figshare.com to share NIH-funded research, consult this [guide](#) or contact info@figshare.com. For information about sharing NIH data using repositories, contact datascience@nih.gov.