

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


Gregory Way, a postdoctoral associate in the Carpenter Lab at The Broad Institute of MIT and Harvard, and his colleagues have published single cell databases on the NIH Figshare instance. In this experiment, his team used CRISPR gene editing technology to knock out important cancer-associated genes in millions of cancer cells in a dish.

They used automated microscopy to measure every cell's morphology in order to determine what cancer cells look like when various genes are compromised.

As part of the National Cancer Institute (NCI), this research data is part of a wider project focused on identifying dependencies in cancer cell lines and helping to develop



Gregory Way



personalized medicine and treatment for individuals diagnosed with cancer. The dataset on the NIH Figshare instance consists of 130GB of SQLITE database files and will continue to be added to by Gregory and his colleagues and released as new versions of the dataset in the future.

“Everything that we do, we try to make public as we’re doing it,” said Gregory. “Throughout the entire research process starting from project initiation we make our code, data, and results available on GitHub. Once we’re ready to formally disseminate our results, we post preprints to BioRxiv.”

These data were originally processed using code stored on GitHub. The processed data has been harvested into the NIH Figshare instance using Figshare’s API. This meant that Gregory was able to move files from Amazon S3 programmatically rather than having to upload the data through the browser interface. Figshare provides documentation and open adaptable code for uploading S3 files through the Figshare API.

“It’s really important to us that the data generators - in this instance, Dr. Francisca Vazquez’s team of the Cancer Dependency Map Project and Dr. Maria Kost-Alimova, who runs the microscopy center at Broad - are sufficiently credited in the record of the data,” said Gregory. “If we’re going to continue sharing data freely, we need to make sure that the data generators are valued and the NIH Figshare instance enables that.”

---

#### References:

1. <https://personal.broadinstitute.org/anne/index.html>
2. <https://doi.org/10.35092/yhjc.9995672>
3. <https://github.com/broadinstitute/cell-health>
4. <https://docs.figshare.com/>
5. [https://docs.figshare.com/#upload\\_files\\_upload\\_s3\\_file\\_to\\_figshare](https://docs.figshare.com/#upload_files_upload_s3_file_to_figshare)