

Hiring: Data Scientist



National Institute
of Mental Health



For the full job ad, visit:
[https://nimh-dsst.github.io/
dataSci_job_ad_2024/](https://nimh-dsst.github.io/dataSci_job_ad_2024/)

The Data Science and Sharing Team (DSST) within the National Institute of Mental Health's (NIMH) Intramural Research Program (IRP) in Bethesda, MD, is seeking an **early to mid-career data scientist** for an on-site or hybrid position.

Applications will be accepted through Friday, October 11, 2024

About the NIH, NIMH, & IRP

With a budget of \$2.5B per year, the NIMH is among the largest divisions of the NIH and is the largest funder of research on mental disorders in the world. The NIMH mission is to transform the understanding and treatment of mental illnesses through basic and clinical research, paving the way for prevention, recovery, and cure. The IRP is the largest biomedical research institution on earth, where the unique funding environment allows scientists to conduct both long-term and high-impact science that would be difficult to impossible in a grant-dependent institution.

About the Team

The DSST is responsible for leading and supporting data-intensive scientific projects within the NIMH IRP. We work closely with the Machine Learning Team, as well as other IRP investigators (e.g. Armin Raznahan, Peter Bandettini, Robert Innis). Special emphasis is placed on reproducible analyses of high dimensional genetic, genomic, neuroimaging, and behavioral datasets.

We also work to implement the policy guidance from the White House Office of Science Policy to ensure free, immediate, and equitable public access to federally funded research, foster greater collaboration and innovation, and strengthen public trust.

Everything we create is open source and freely distributed. A full listing of research objects from our team as of December 2023 is available in our Board of Scientific Counsellors Report.

About the Position

Responsibilities

- Building tools and pipelines for reproducible data analysis using Python
- Implementing reproducible pipelines using both High Performance Computing (HPC) and Elastic Cloud Computing (ECC)
- Writing clear documentation and scholarly publications
- Teaching and consulting on data science best practices for reproducibility (e.g. Software and Data Carpentries, Brain-

Hack Global, FAES)

- Creating, expanding, and supporting FAIR data standards in biomedical science (e.g. BIDS, NWB, OME, GO)

This position has minimal supervisory responsibilities, however motivated and productive individuals will have the opportunity to quickly assume a more senior role involving supervision and oversight of additional projects in parallel.

Necessary Qualifications

- A recent graduate degree (PhD preferred) in a STEM field or equivalent knowledge and experience
- Strong coding skills in Python
- Experience with version control and continuous integration with git and GitHub
- Experience with more than one type of biomedical data (e.g. neuroimaging, genomics, next generation sequencing (NGS), transcriptomics, electrophysiology, microscopy)

Desirable but not Required Qualifications

- Experience working with biomedical data repositories and/or biobanks (e.g. The Adolescent Brain and Cognitive Development Project (ABCD), OpenNeuro, UK Biobank, GTEX)
- Experience with distributed, high-performance computing tools such as Docker/Apptainer and batch processing systems such as SLURM
- Recent examples of open access research objects (e.g. publications, code, datasets) on which you were a major contributor

How to apply...

- Email your CV to DATASCI-JOBSEARCH@mail.nih.gov
- Use the body of the email as your cover letter

We especially encourage applications from members of underrepresented groups in the data science and biomedical research community. Other inquiries are also welcome.

The National Institutes of Health is an equal opportunity employer. This position will be based at NIMH in Bethesda, MD via a third party contracting firm who is able to provide visa sponsorship.