Northwestern University Feinberg School of Medicine

Galter Health Sciences Library & Learning Center

DataLab

The Galter DataLab is focused on collaborative innovation, training & development, and connecting faculty, staff, and students in the Feinberg School of Medicine to data-related resources here at Northwestern. Through our free DataClinic service, we provide consultation and training for all stages of the research data life cycle. We engage the research community by hosting and sponsoring events that promote best practices related to reproducibility and openness in the sciences. In partnership with the National Center for Data to Health (CD2H), we are developing a next-generation repository and data index system that will help scientists across the clinical and translational spectrum to locate tools, data, and expertise that will enhance their research process.

Follow @GalterDataLab on Twitter

DataLab Components

DataClinic

The Galter DataClinic employs a primary care model for data management and analysis. In addition to providing training and best practices support, we offer free consultations for researchers with data issues involving:

- Collection & Management
- Cleaning & Organization
- Analysis
- Visualization
- Preservation

For long-term or specialized support, we refer researchers to our partners here at Northwestern:

To request a Data Clinic consultation, please fill out the contact form.

Data Education & Community Engagement

The Galter Library actively promotes reproducibility and open science best practices by organizing and hosting special events featuring cutting-edge organizations and software tools in the field. See our list of events for more information.

Innovations

Northwestern University, in partnership with Oregon Health & Science University, University of Washington, Johns Hopkins University School of Medicine, and Sage Bionetworks, has been awarded a five-year $25M cooperative agreement from the National Center for Advancing Translational Science (NCATS) to create a new CTSA Program National Center for Data to Health (CD2H). This award will coalesce and coordinate informatics activities across the Clinical and Translational Science Award (CTSA) Program, a network of more than 50 medical research institutions, to provide collaborative clinical and translational research infrastructure. Projects include:

- Development of a next-generation repository and data index system
- Mapping ontologies and metadata
- Data modeling

Read more about CD2H here.

Data-related Resources

For long-term or specialized support, please contact one of our Northwestern University partners. We have highlighted a few important links below:

NU Policies and Procedures

Institutional Review Board (IRB)

- Data Review Protocol

Office for Sponsored Research (OSR)

- Data Use Agreement (DUA)
- Data Retention Policy

FSMIT

- Information Security
- Security Policies
- Guidelines for file storage
- Guidelines for file sharing
- Data Security Plan for information used in clinical research
- FAQs

Clinical Research Support

- REDCap support

- Northwestern Medicine Enterprise Data Warehouse (NMEDW)
- Feasibility Query Tool (I2B2)

Computational and Data Analysis Support

- NUIT Research Computing
- Biostatistics Collaboration Center (BCC)
- NUSeq

Educational Material

- Understanding Academic Research: Free and Low-cost Tools and Workflows
- Data Organization and Documentation
- Data Policies from Funding Agencies

Hosted and Sponsored Events

The Galter Library actively promotes reproducibility and open science best practices by organizing, hosting, and sponsoring special events featuring cutting-edge organizations and software tools in the field.

Workshops

Computational Skills for Informatics [Winter 2018]

Galter Library, NUIT Research Computing, and NUCATS are bringing back Computational Skills for Informatics (CSI). Sessions will be held every other Thursday afternoon from 3:00 - 4:30 pm. We are repeating three popular sessions from past fall.

- Sessions will be approximately 1.5 hours depending on the topic
- Sessions will be held in Galter Library's Learning Resources Center (LRC)
- Space will be limited; register using the links below each class description
- We will provide snacks for the sessions
- Users should bring their own laptops for hands-on sessions

Introduction to R [September 2018]

This two-day workshop is from 12-5pm on both Monday, September 24th and Wednesday, September 26th. The instructor is Christina Maimone, Northwestern IT Research Computing Services. Topics covered include:

- How to use RStudio and projects
- Import and manipulate data
- Basic plotting
- How to write an R script
- Aggregating and summarizing data
- Import and use packages
- Read documentation and get more help

R: Refresher [July 2018]

One of the NUIT Research Computing Summer Workshops, this tutorial is design for R users who have stepped away from the platform for a while and now want to get back to it. In this workshop you will learn how to:

- Use RStudio
- Import and manipulate data
- Create basic plots
- Write an R script
- Aggregate and summarize data
- Import and use packages
- Read documentation and get more help

See instructor Matt Carson's GitHub repository for the workshop materials.

Integrating Reproducible Best Practices into Biomedical and Clinical Research: A Hands-on Workshop for Researchers [April 2018]

Hosted by Code Ocean: With the aim of improving the reproducibility of research, journals and funders are increasingly calling for published research to include associated data and code. This workshop introduces reproducibility best practices for biomedical and clinical research. We will demonstrate and practice tools that help researchers overcome barriers to reproducibility. Goals of this workshop include:

- Define reproducibility and its relevance to biomedical and clinical researchers.
- Learn best practices for planning, documenting, and sharing your reproducible research project.
- Identify guidance to support trial transparency, registration, and reporting.
- Explore the issue of improving the transparency of sensitive data.
- Assess possible tools to improve reproducibility.
- Submit your code for publishing on Code Ocean

Practical Steps for Increasing Openness and Reproducibility: An Introduction to the Open Science Framework (OSF) [November 2017]

Hosted by the Center for Open Science: Targeted towards graduate students, postdocs, and faculty across disciplines engaged in quantitative research, this workshop provides a foundation for incorporating reproducible, transparent practices into current workflows. Participants learn to:

- Create meaningful project documentation
- Implement Version Control
- Write pre-analysis plans
- Use open source tools like the Center for Open Science's Open Science Framework to easily implement these concepts in a scientific workflow.

See video tutorials on the OSF here.

Hackathons

Post-ISMB 2018 Chicago Bioinformatics Hackathon at Northwestern University School of Medicine [July 2018]

Hosted by NCBI: The hackathon focuses on genomics as well as general Data Science analyses including text, image and sequence processing. This event is for researchers, including students and postdocs, who have already engaged in the use of large datasets or in the development of pipelines for analyses from high-throughput experiments. Some projects are available to other non-scientific developers, mathematicians, or librarians. The event is open to anyone selected for the hackathon and willing to travel to Chicago. Working groups of five to six individuals will be formed into five to eight teams. These teams will build pipelines and tools to analyze large datasets within a cloud infrastructure. Example subjects for such hackathons include:

- Prediction of infection-prone metagenomic states

- Endogenous Retroviral Expression in Cancer
- Expression of Structural Variants in Sepsis
- An Online Bioinformatics Pipeline Design Engine
- Disease clustering from literature based on limited training data (phenotypic information)
- Graphical User Interface for Gene Expression calculated on the fly from raw data

---

Contact DataLab
