

Galter DataLab

Overview for Cunningham Fellows: May 10, 2023

Matt Carson, Head of Digital Systems
Galter Health Sciences Library & Learning Center

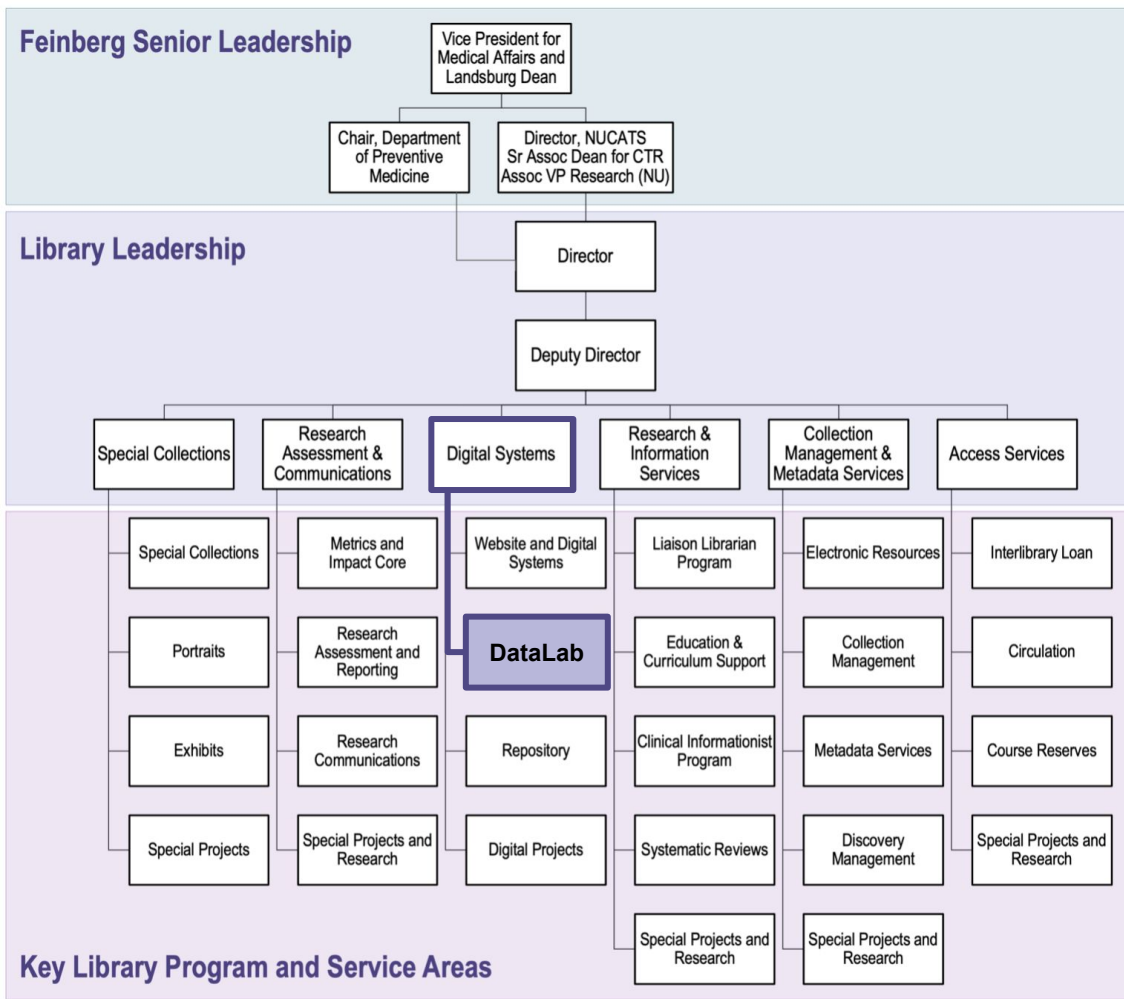


What We Are

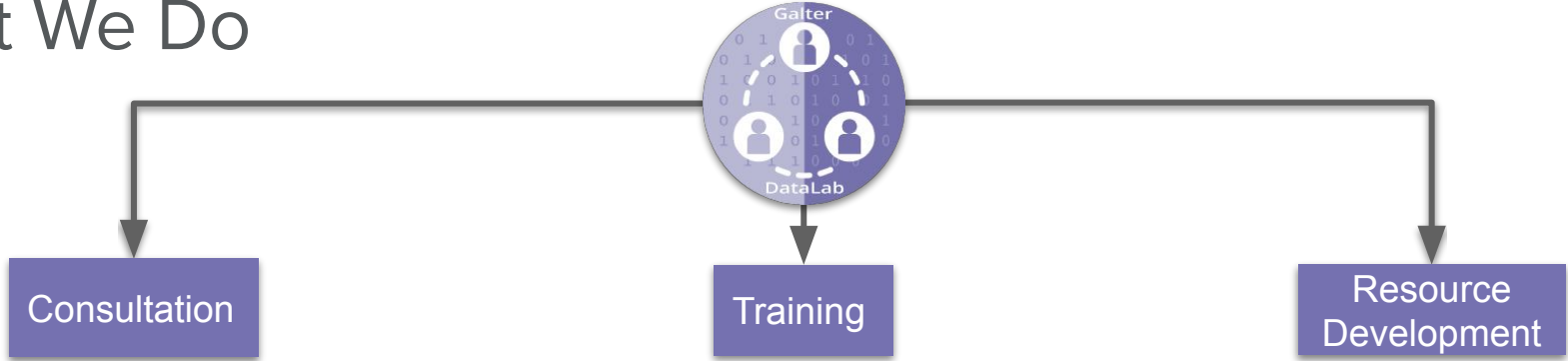
***A research support core that provides
consultation, training, and resource
development for Feinberg and beyond***



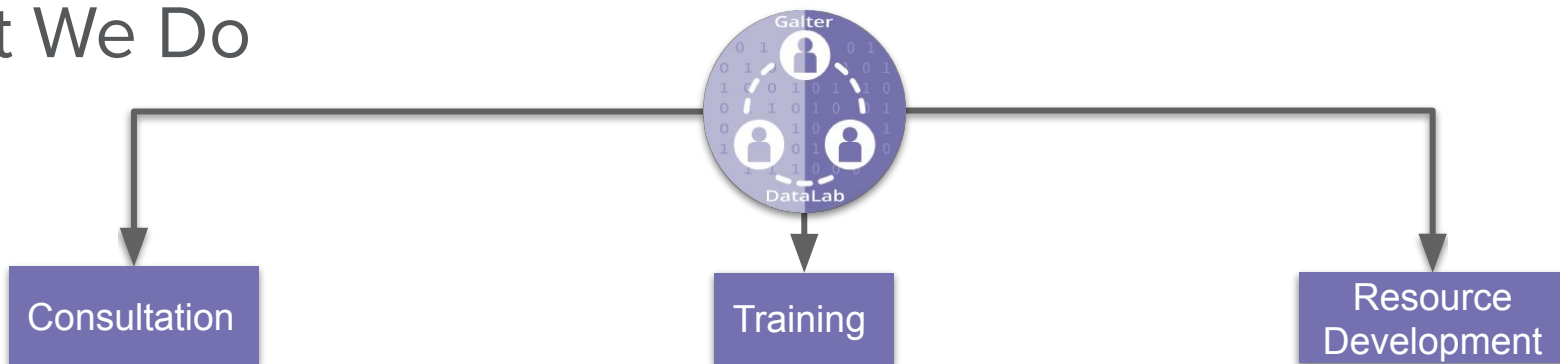
Where We Are



What We Do



What We Do

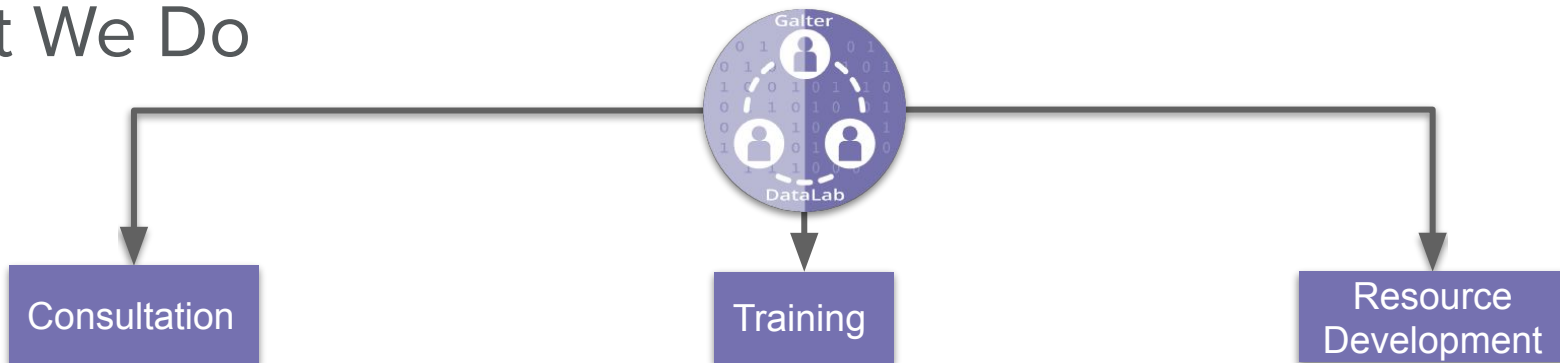


Support for a variety of data-related topics:

- * Policy Compliance
 - NIH Data Sharing
 - NIH Public Access
- * Collection & Management
- * Cleaning & Organization
- * Analysis
- * Visualization
- * Preservation
- * Metadata issues (e.g., MeSH, UMLS Metathesaurus)
- * Databases and SQL
- * Online bioinformatics databases and tools
- * PubMed-related issues
- * R and Python



What We Do



Support for a variety of data-related topics:

- * Policy Compliance
 - NIH Data Sharing
 - NIH Public Access
- * Collection & Management
- * Cleaning & Organization
- * Analysis
- * Visualization
- * Preservation
- * Metadata issues (e.g., MeSH, UMLS Metathesaurus)
- * Databases and SQL
- * Online bioinformatics databases and tools
- * PubMed-related issues
- * R and Python



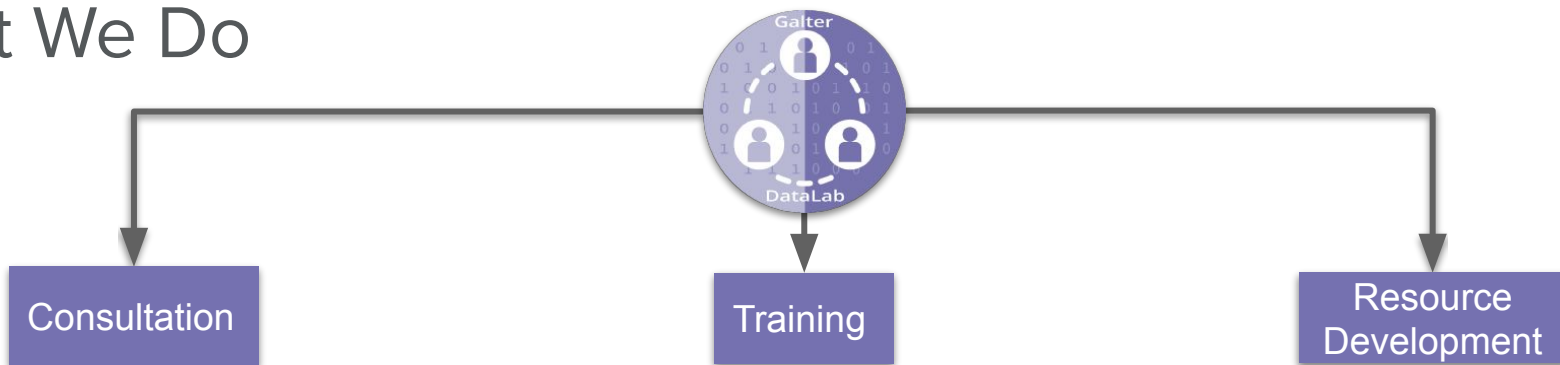
The Galter DataLab works with a variety of contributors and partners to provide workshop and other educational content for Feinberg Medical School faculty, staff, and students. Topics include:

- Data Management
- R, Python, SQL
- Large-scale computing
- Open Science and Reproducibility
- Hackathons

Galter Health Sciences Library & Learning Center
Northwestern INFORMATION TECHNOLOGY RESEARCH COMPUTING SERVICES
National Center for Biotechnology Information



What We Do



Support for a variety of data-related topics:

- * Policy Compliance
 - NIH Data Sharing
 - NIH Public Access
- * Collection & Management
- * Cleaning & Organization
- * Analysis
- * Visualization
- * Preservation
- * Metadata issues (e.g., MeSH, UMLS Metathesaurus)
- * Databases and SQL
- * Online bioinformatics databases and tools
- * PubMed-related issues
- * R and Python



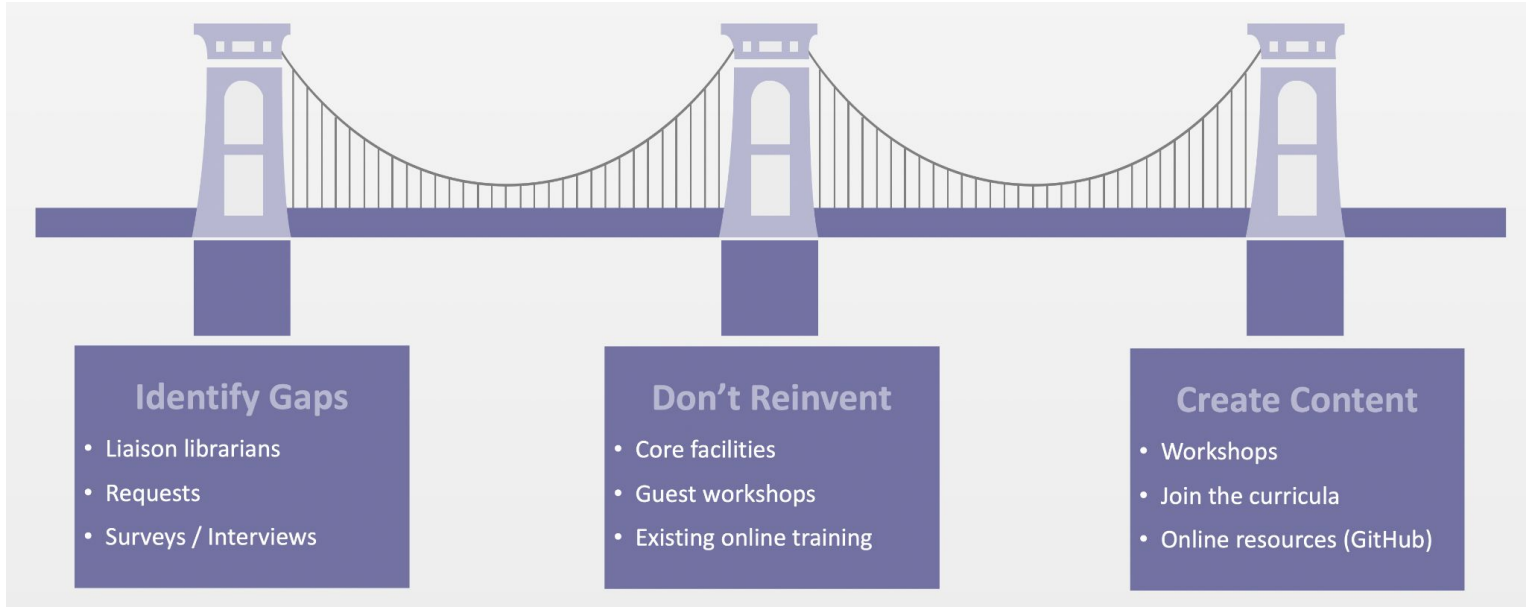
The Galter DataLab works with a variety of contributors and partners to provide workshop and other educational content for Feinberg Medical School faculty, staff, and students. Topics include:

- Data Management
- R, Python, SQL
- Large-scale computing
- Open Science and Reproducibility
- Hackathons

Logos: software carpentry, library carpentry, DATA CARPENTRY, CODE OCEAN, OPEN SCIENCE, Galter Health Sciences Library & Learning Center, Northwestern UNIVERSITY INFORMATION TECHNOLOGY RESEARCH COMPUTING SERVICES, National Center for Biotechnology Information, NCBI.



Consultation & Training Strategy

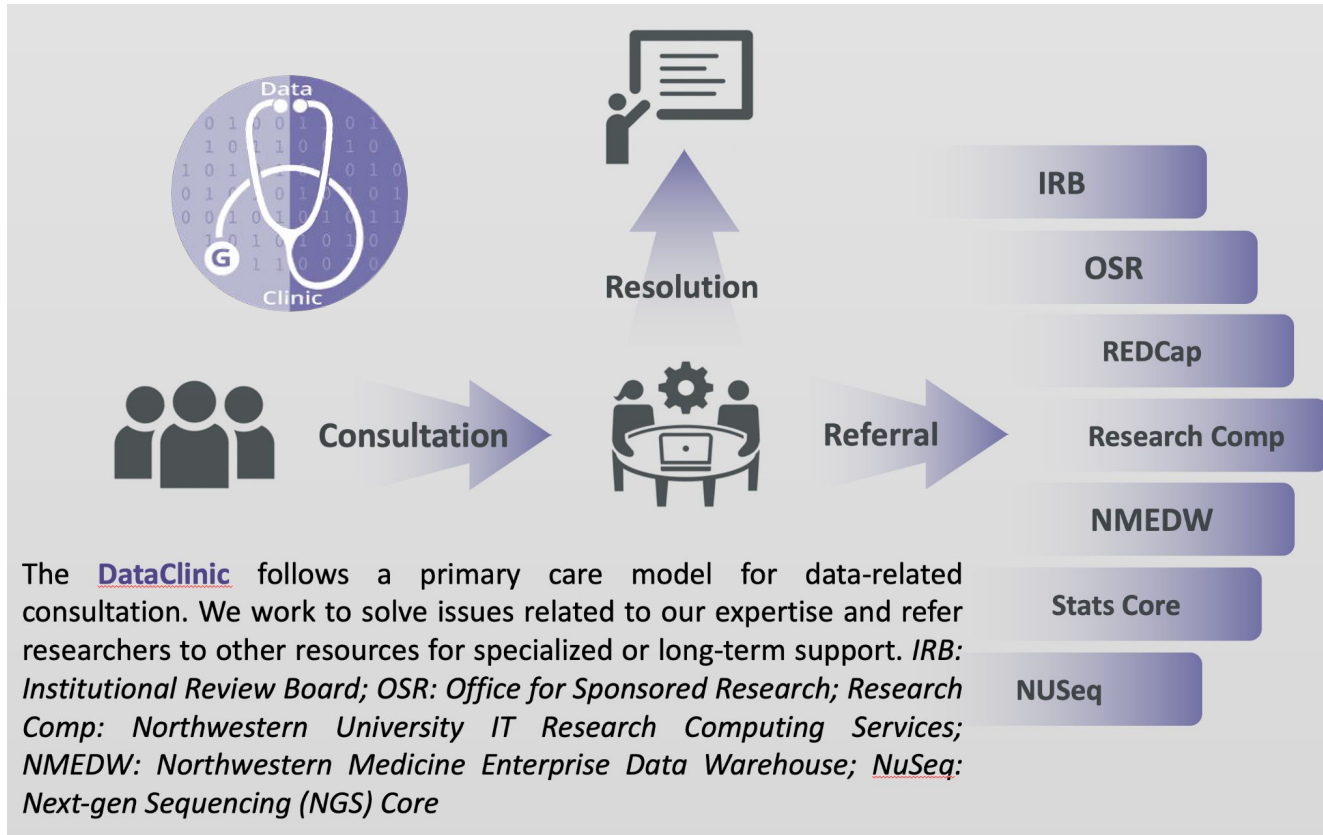


Consultation



Consultation

dataclinic@northwestern.edu



Training



Resources

- Data training @ Galter
- LibGuides
- Boilerplate language for grants/DMPs
- Example DMPs & best practices
- DMP Tool
- Prism (Institutional Repository)

Services & Other Activities

- Open office hours
- Liaison Librarians
- Meetups & “social coding” events
- Communications

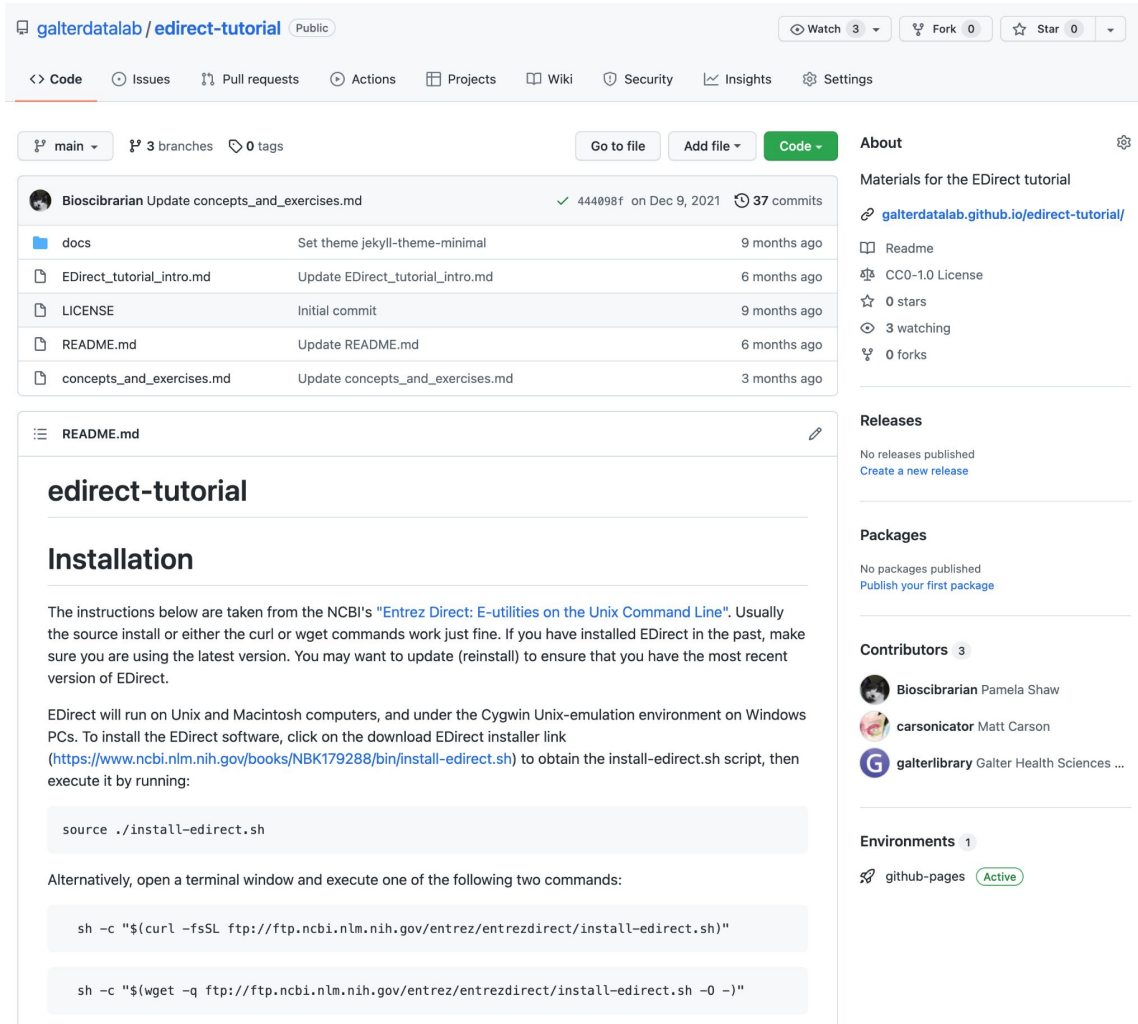
Data Training @ Galter

- **Getting to know the NIH Policy on Data Management and Sharing (monthly, begins August)**
- **10 Simple Rules for Conducting a Data Inventory**
- **10 Simple Rules for Data Cleaning with Excel**
- **10 Simple Rules for File Organization**
- **Best Practices in Research Data Management and Data Sharing**
- **Cleaning Spreadsheet Data with OpenRefine**
- **Data Projects: Setup and Key Tips for Excel**
- **Essential Spreadsheet Data Cleaning with OpenRefine**
- **Introduction to SQL for Clinical Researchers**
- **Making your Data FAIR**
- **Principles of Effective Data Visualization**
- **Others in development**



Tutorial Example

This **Entrez Direct** tutorial by Pamela Shaw walks the user through NCBI's E-utilities, a set of tools that let you search database statistics, text, documents, etc. from the many Entrez Databases



The screenshot shows the GitHub interface for the repository 'galterdatalab / edirect-tutorial'. At the top, there are navigation links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. Below this, the repository name and 'Public' status are shown, along with 'Watch 3', 'Fork 0', and 'Star 0' buttons. The main content area displays a commit history table with columns for author, commit message, and date. The current commit is by 'Bioscibrarian' updating 'concepts_and_exercises.md'. Below the commit history, the 'README.md' file is open, showing the title 'edirect-tutorial' and an 'Installation' section. The installation instructions describe how to run EDirect on various operating systems and provide a terminal command to execute the installer script. The right sidebar contains metadata for the repository, including a link to the tutorial page, license information (CC0-1.0), and statistics (0 stars, 3 watchers, 0 forks). It also lists releases, packages, contributors (Bioscibrarian, carsonicator, galterlibrary), and environments (github-pages).

Author	Commit Message	Date
Bioscibrarian	Update concepts_and_exercises.md	444098f on Dec 9, 2021
	docs	Set theme jekyll-theme-minimal 9 months ago
	EDirect_tutorial_intro.md	Update EDirect_tutorial_intro.md 6 months ago
	LICENSE	Initial commit 9 months ago
	README.md	Update README.md 6 months ago
	concepts_and_exercises.md	Update concepts_and_exercises.md 3 months ago

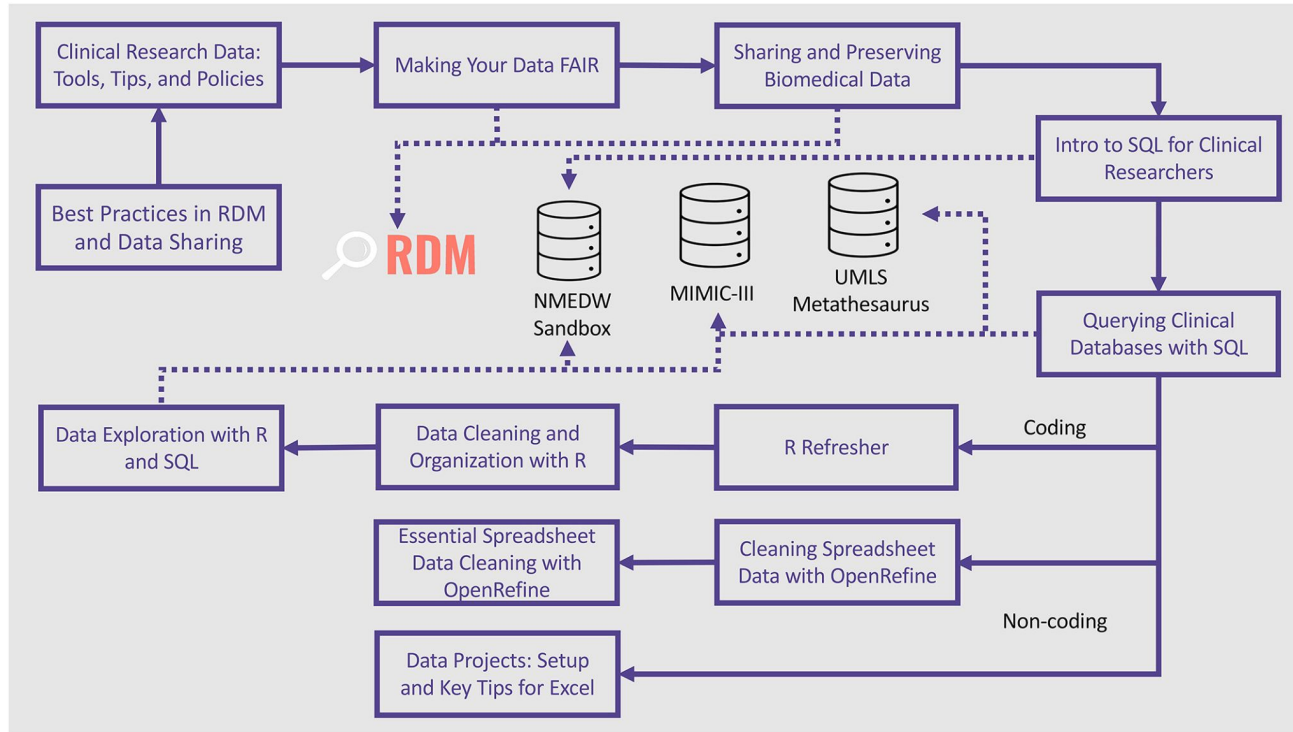
```
source ./install-edirect.sh
```

```
sh -c "$(curl -fsSL ftp://ftp.ncbi.nlm.nih.gov/entrez/entrezdirect/install-edirect.sh)"
```

```
sh -c "$(wget -q ftp://ftp.ncbi.nlm.nih.gov/entrez/entrezdirect/install-edirect.sh -O -)"
```



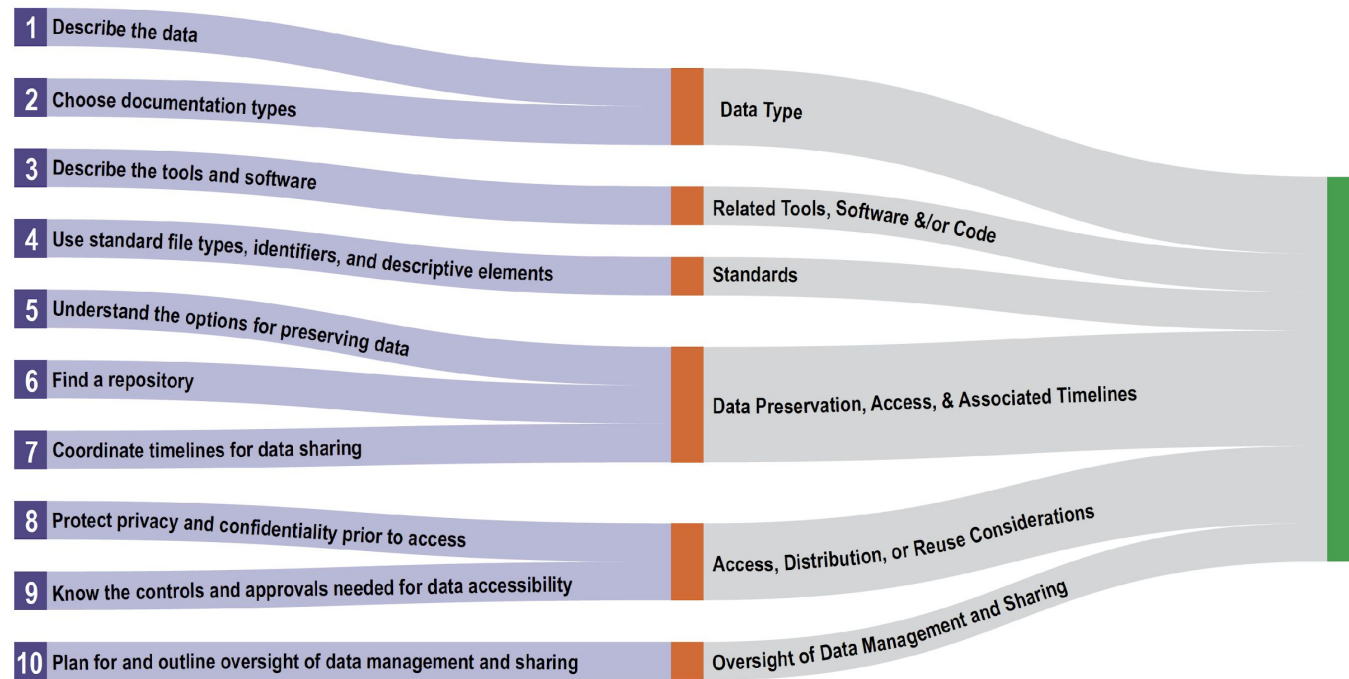
Clinical Research Data Management (cRDM) Program



Carson, MB, Gonzales, S, Shaw, P, Schneider, D, Holmes, K. Bridging the gap: A library-based collaboration to enhance data skills for clinical researchers. *Learn Health Sys.* 2023; 7(2):e10339. doi:[10.1002/lrh2.10339](https://doi.org/10.1002/lrh2.10339)



10 Simple Rules for Maximizing the Recommendations of the NIH Data Management and Sharing Plan



Sankey diagram to show the relationship between the “Ten Simple Rules” (purple bars) described in the paper and the elements of an NIH Data Management and Sharing Plan, as described by NIH (gray bars).



Resource example: LibGuide

Elements of an NIH Data Management and Sharing Plan

Alignment with & reinforcement
of Feinberg policies

Creating an NIH Data Management and Sharing Plan

This guide presents information on the NIH Data Management and Sharing Policy taking effect on January 25, 2023, requiring Data Management and Sharing Plans (DMSPs) for all NIH-supported research.

Getting Started

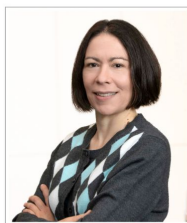
[Creating an NIH Data Management Plan](#)

[Data Management and Sharing: Galter and FSM Resources](#)

[Additional NIH Resources](#)

[Additional Resources from Across the Web: Includes Sample DMS Plans](#)

Data Librarian



Sara Gonzales

[Email Me](#)

Contact:

Sara Gonzales
Senior Data Librarian
Galter Health Sciences
Library & Learning Center

Introduction

As of January 25, 2023, the NIH has implemented an updated [Data Management and Sharing Policy](#) that will apply to all NIH-funded research that generates scientific data, regardless of funding mechanism. Compliance with this policy requires submission of a Data Management and Sharing Plan (DMSP) with the funding application, outlining the plan for managing, preserving, and sharing project data. This guide provides recommendations for creating a compliant DMSP.

For questions or a consultation, please contact ResearchData@northwestern.edu

[NIH Data Management and Sharing Policy and supplemental guidance links:](#)

- [NOT-OD-21-013: The Final NIH Policy for Data Management and Sharing](#)
 - [Plan writing supplement: Elements of an NIH Data Management and Sharing Plan](#)
 - [Cost-defining supplement: Allowable Costs for Data Management and Sharing](#)
 - [Repository selection supplement: Selecting a Repository for Data Resulting from NIH-Supported Research](#)

[See also:](#)

- NIH's website on data sharing, <https://sharing.nih.gov/>, for all the latest updates and links
- NIH's [FAQs for the NIH Policy for Data Management and Sharing \(DMS Policy\)](#)
- NIH's [Implementation Details for the NIH Data Management and Sharing Policy](#)
- NIH's [DMP Format Page](#) (a sample template for creating a Data Management and Sharing Plan)
- The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)'s [Data Management and Sharing Website](#)

Consult the sections of this guide for section-by-section guidance on creating an NIH Data Management Plan.

Data to Share vs. Data Not to Share

What data should be shared?

- Sufficient data to validate and replicate study findings
- Data produced through the study but not necessarily resulting in a publication
- Null findings on which no publication will be made

Which data are not eligible to be shared?

- Laboratory notebooks
- Preliminary analyses
- Completed case report forms

Online, Self-paced Training



DATAQUEST

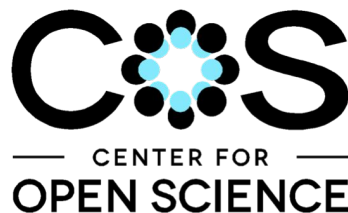


O'REILLY®
Safari

Other
training,
tools, and
resources



Education & Community Engagement



Galter Health Sciences Library
& Learning Center

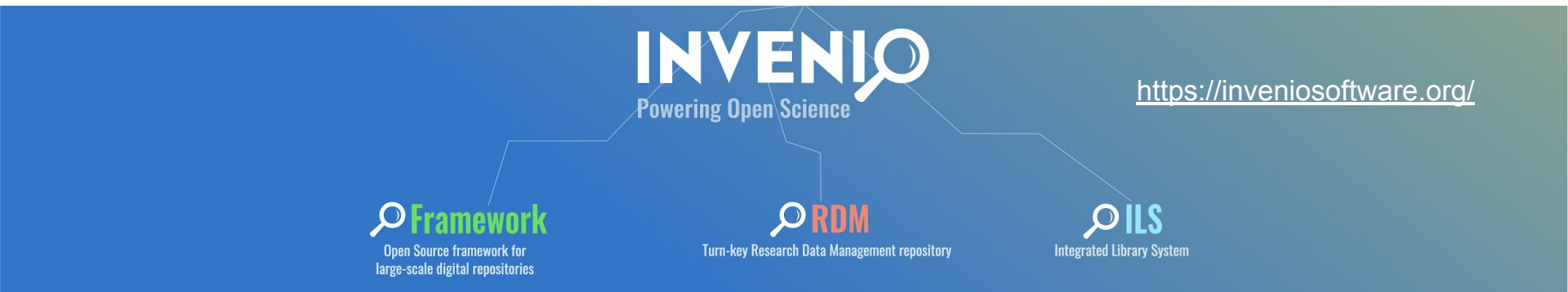
Northwestern | INFORMATION TECHNOLOGY
RESEARCH COMPUTING SERVICES



Resource Development



Invenio: CERN's Repository Framework



Free Open Source Software

Invenio is Free Open Source Software supported by a committed community of multidisciplinary institutions.

[Code](#) | [Docs](#) | [Examples](#)



Friendly and Responsive Community

Although Invenio was born at CERN, its community is growing bigger every day. Talk to the team now in our chatroom or forum.

[Chatroom](#) | [Forum](#) | [Get Involved](#) | [Events](#)



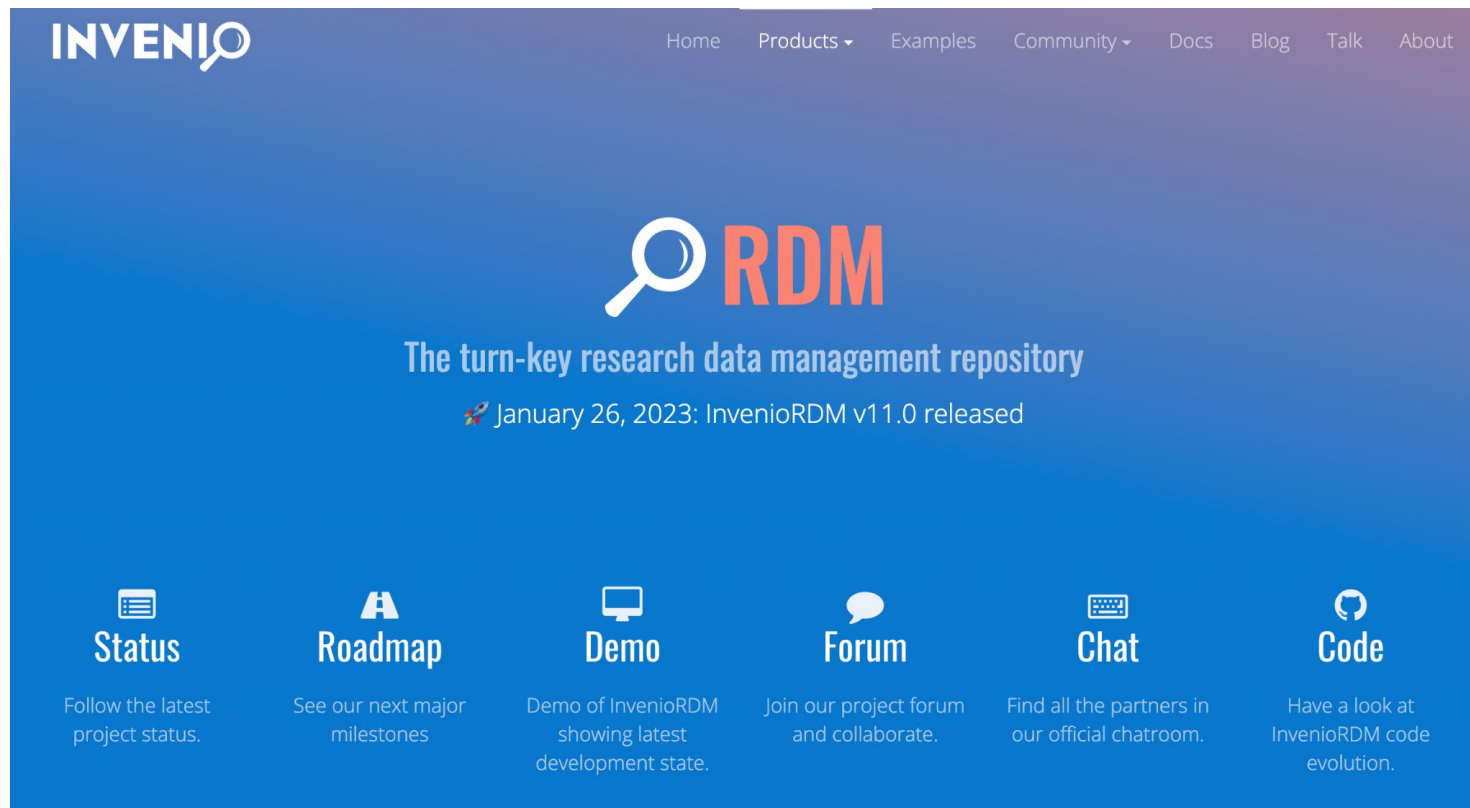
The "Safe bet"

Invenio community has been around for 20 years. **Solid services** have been built on top of it to ensure long-term confidence.

[Live services](#) | [Products](#)



InvenioRDM



The screenshot shows the InvenioRDM website homepage. At the top left is the 'INVENIO' logo with a magnifying glass icon. The navigation menu includes 'Home', 'Products', 'Examples', 'Community', 'Docs', 'Blog', 'Talk', and 'About'. The main heading is 'RDM' with a magnifying glass icon, followed by the tagline 'The turn-key research data management repository'. A news item states 'January 26, 2023: InvenioRDM v11.0 released'. Below this are six feature cards: 'Status' (project status), 'Roadmap' (milestones), 'Demo' (development state), 'Forum' (collaboration), 'Chat' (official chatroom), and 'Code' (code evolution).

INVENIO

Home Products Examples Community Docs Blog Talk About

RDM

The turn-key research data management repository

🚩 January 26, 2023: InvenioRDM v11.0 released

- Status**
Follow the latest project status.
- Roadmap**
See our next major milestones
- Demo**
Demo of InvenioRDM showing latest development state.
- Forum**
Join our project forum and collaborate.
- Chat**
Find all the partners in our official chatroom.
- Code**
Have a look at InvenioRDM code evolution.



NIH Generalist Repository Ecosystem Initiative (GREI)

The mission of GREI is to establish a common set of capabilities, services, metrics, and social infrastructure; raise general awareness and facilitate researchers to adopt FAIR principles to better share and reuse data.

This initiative will further enhance the biomedical data ecosystem and help researchers find and share data from NIH-funded studies in generalist repositories.

Goals of the Generalist Repository Ecosystem Initiative



1

Make it easier for researchers to **share data**



2

Enable the improved **discoverability** of NIH-funded data across generalist repositories



3

Support greater **reproducibility** of NIH-funded research by ensuring data associated with publications is readily available



4

Avoid **duplication** of data across repositories



5

Encourage NIH-funded researchers to be both contributors and consumers to **increase the reuse of data**



NIH Generalist Repository Ecosystem Initiative (GREI)

Objectives



- Catalog Use Cases Supported
- Implement Open Metrics
- Prepare Training Materials
- Conduct Outreach
- Commit to 'Cooperation'
- Implement Best Practices for Data Repositories
- Support Discovery of NIH Funded Data
- Adopt Consistent Metadata Models
- Facilitate QA/QC
- Connect Digital Objects

Participating Repositories



Contact

GREI@nih.gov • <https://bit.ly/3en3rRy>

<https://doi.org/10.6084/m9.figshare.21318270>



NNLM National Evaluation Center (NEC)

- Provides evaluation materials and resources NNLM-funded grants and activities
- Based in Galter Library for the cooperative agreement period 2021-2026
- DataLab team projects:
 - RDM 101 & 102 courses
 - NNLM Personas
 - Data Science Interns
 - NNLM Data Warehouse

NIH National Library of Medicine
Network of the National Library of Medicine

What do you want to search for today?

Hello, Matthew Carson
[Update my profile](#)
[Log out](#)

Home About Us Membership Initiatives Funding Training Guides & Resources

Home / About Us / Centers / NEC

View Edit Layout Revisions

CENTER: NNLM National Evaluation Center

The NNLM National Evaluation Center (NEC) provides innovative evaluation frameworks and evidence-based evaluation tools and practices to NNLM. NEC collaborates with:

- other NNLM Regions, Offices and Centers to develop strategies and standardized approaches for evaluating outreach and education services
- the NNLM Training Office (NTO) to provide continuing education opportunities for NLM, RMLs and network members
- the NLM Office of Engagement and Training (OET) to ensure NNLM evaluation strategies are aligned with the NLM Strategic plan.

The NEC is located at the Galter Health Sciences Library and Learning Center, Northwestern University, Feinberg School of Medicine in Chicago.

Staff Directory
Looking for assistance? Browse our staff directory to find contact information, and get to know the NEC staff by reading their personal bios.

Funding Opportunities
Browse open funding opportunities that are available to NNLM members.

Classes & Training
Browse NNLM's upcoming training opportunities, course catalog, and recordings of past webinars.

Galter Health Sciences Library & Learning Center | 320 E. Superior Street | Chicago, Illinois | 60611
tel: 312-503-8126 |





Data Infrastructure

The NEC is working with the National Web Services Office (NWSO) to build data infrastructure that will enhance reporting and analytics capabilities for the NNLM community

NWSO

- NNLM website
- CiviCRM
- Data Reporting System (DRS)

NEC

- Survey Data
- NEC Data Warehouse
- Tableau Dashboards
- Zenodo Community



Galter Health Sciences Library & Learning Center



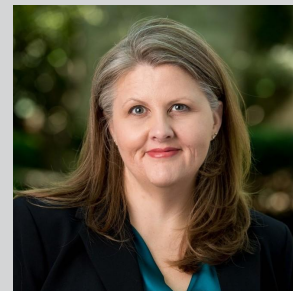
Matt Carson



Sara Gonzales



Pamela Shaw



Kristi Holmes

*Developed resources reported in this workshop are supported by the **National Library of Medicine (NLM)**, **National Institutes of Health (NIH)** under cooperative agreement number **1UG4LM012346**, by the **Network of the National Library of Medicine Evaluation Center**, grant number **1U24LM013751-01**, and by the **National Institutes of Health's National Center for Advancing Translational Sciences**, Grant Number **UL1TR001422**. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.*



[galterdatalab.github.io](https://github.com/galterdatalab)



datalab@northwestern.edu



[@GalterDataLab](https://twitter.com/GalterDataLab)

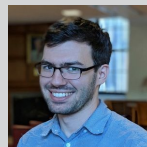
InvenioRDM and Prism

Teams

- The Invenio team @ CERN & RDM collaborators ([here](#))
- Galter Health Sciences Library & Learning Center

Guillaume Viger
Sara Gonzales
Karen Gutzman
Josh Elder
Austin Sharp

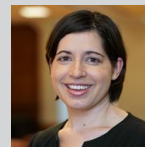
Eric Newman
Mahonry Estrada
Gretchen Neidhardt
Matt Carson



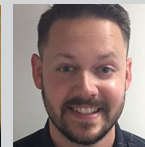
Guillaume
Viger



Sara
Gonzales



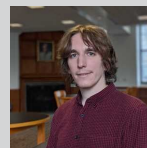
Karen
Gutzman



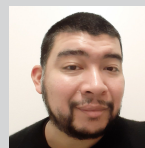
Josh Elder



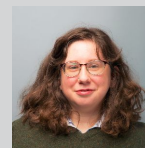
Austin
Sharp



Eric
Newman



Mahonry
Estrada



Gretchen
Neidhardt



Matt Carson



Kristi Holmes

Support

Work presented here is supported in part by:

CERN Knowledge Transfer Fund

NUCATS: UL1TR001422 (NCATS)

★ all of the InvenioRDM project partners



Northwestern University
NUCATS
Clinical and Translational Sciences Institute

Acknowledgements



Northwestern Medicine Enterprise Data Warehouse

Northwestern | INFORMATION TECHNOLOGY
RESEARCH COMPUTING SERVICES