A Library-based Data Retrieval and Management Training Program for Clinical Researchers

Matthew B. Carson, PhD, Sara Gonzales, MS, Pamela L. Shaw, MS, Dan Schneider, MS, Kristi L. Holmes, PhD

Galter Health Sciences Library & Learning Center, Northwestern University, Chicago, IL

Northwestern Medicine® Feinberg School of Medicine

Northwestern University NUCATS Clinical and Translational Sciences Institute

CTSA Clinical & Translational Science Awards Program

Bridging the skills gap for clinical researchers

In 2018, the Northwestern Medicine Enterprise Data Warehouse (NMEDW) received over 600 research data requests, up nearly 50% from the previous year and more than 200% from 2014. This significant increase in demand for clinical research data highlights the need for a more comprehensive training program for system users. At the same time, resource constraints have made it difficult to provide training for clinician researchers that helps to close the communication gap between these researchers and the data analysts who help them extract their research population data of interest from the database. We believe that by providing proper training we can help these researchers translate their domain knowledge more effectively to their analyst partners, and vice versa. This improved communication pathway would not only increase efficiency but also help to avoid miscommunication that causes delays and can even result in incomplete or incorrect inclusion criteria for research cohorts. In addition, current research reports, which include project title and description, inclusion and exclusion criteria, SQL queries, and data plots are not stored in ways that support reuse and preservation. While the analysts themselves have access to past reports, researchers have no way to search for and explore previous projects that may help them realize efficiencies and advance their own work. These improvements would benefit the NMEDW, with more efficient workflows and less one-on-one analyst time needed; the researcher patrons, who will have an easier research start-up process and a foundation of training and reference studies on which they can build their work; the university, which will save time and research funding that could be applied elsewhere; and the larger community, as we will make project findings and resources available to others who seek to support this critical need on their campuses.
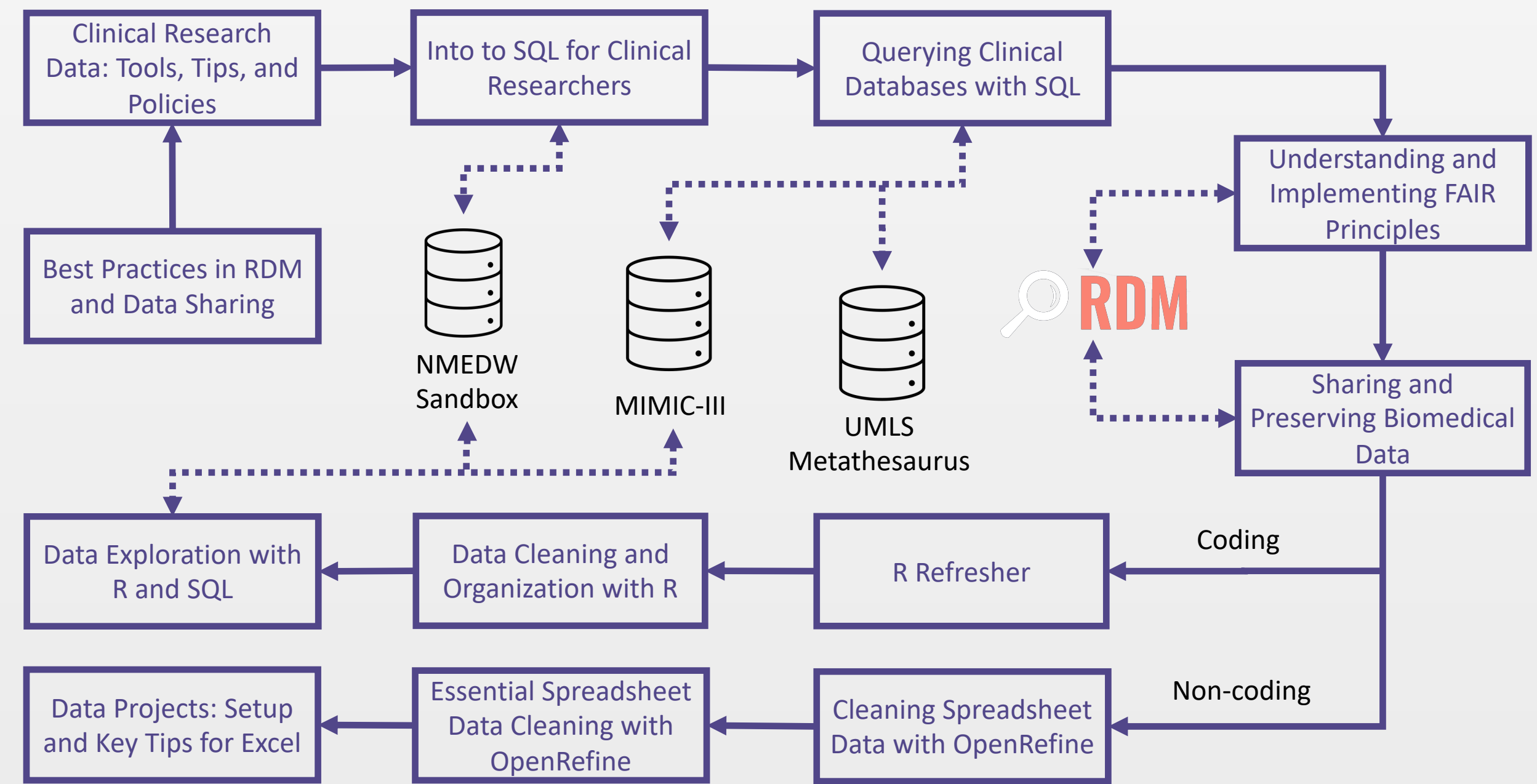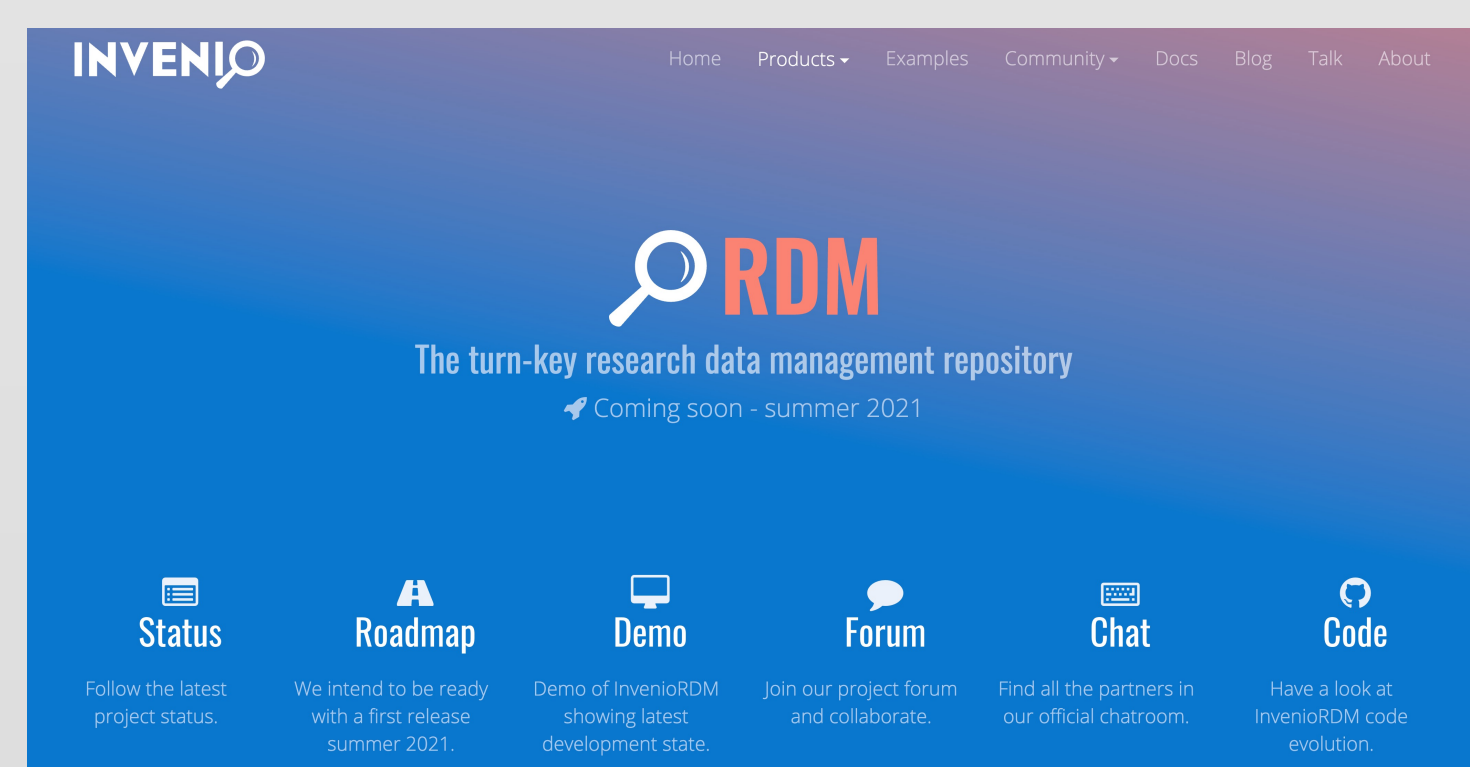

Objective 1: To create an end-to-end Clinical Data Retrieval and Management Program for researchers that teaches them how clinical data is collected, stored, and retrieved, how to identify their research population of interest, how to create practical data retrieval workflows for their clinical research projects, and best practices for ensuring that the research reports for these projects are reproducible and reusable.
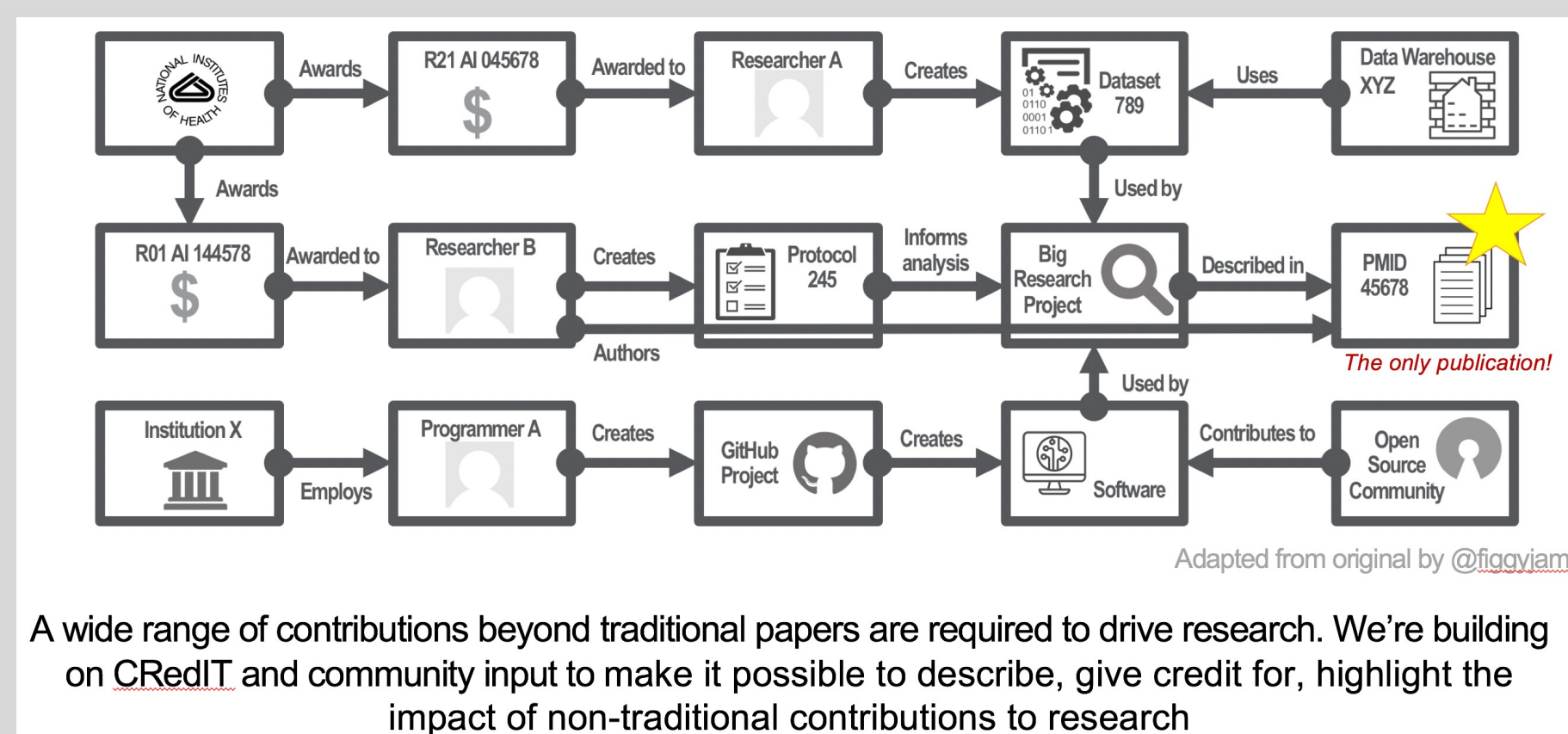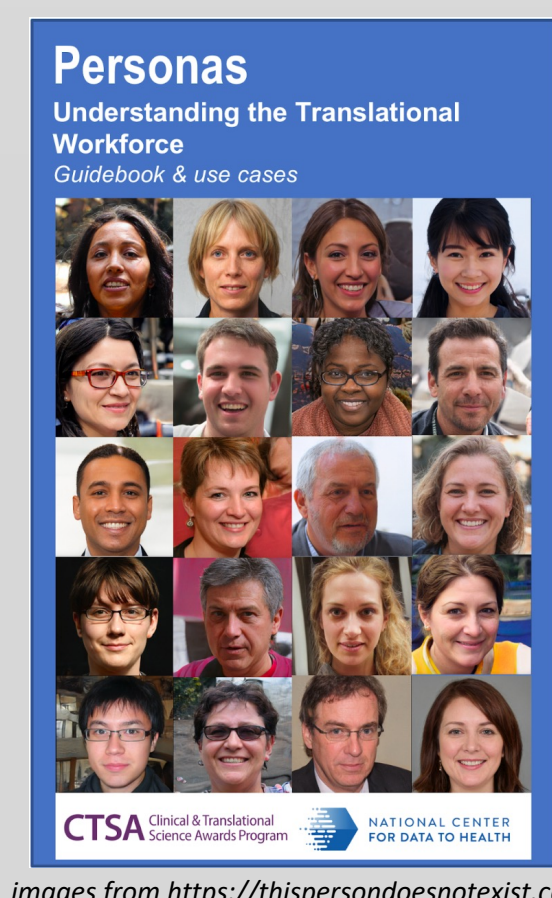
Objective 2: To promote improved communication and collaboration between data analysts and clinical researchers to make them better partners in research projects.

Objective 3: To enhance reusability of clinical reports and database queries by creating workflows and training for preserving them in our next-generation research data management system and making them discoverable to Northwestern's research community.

Objective 4: To use our work and experience to provide a template for clinical research data education and training for other institutions.

Galter DataLab

The **Galter DataLab**[1] 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
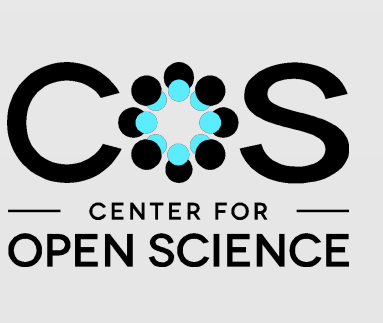
Galter Health Sciences Library & Learning Center

Northwestern INFORMATION TECHNOLOGY RESEARCH COMPUTING SERVICES

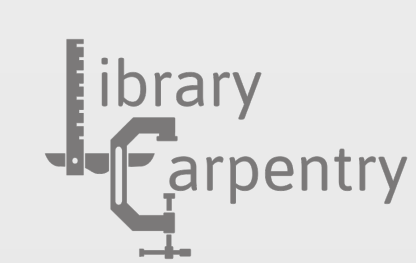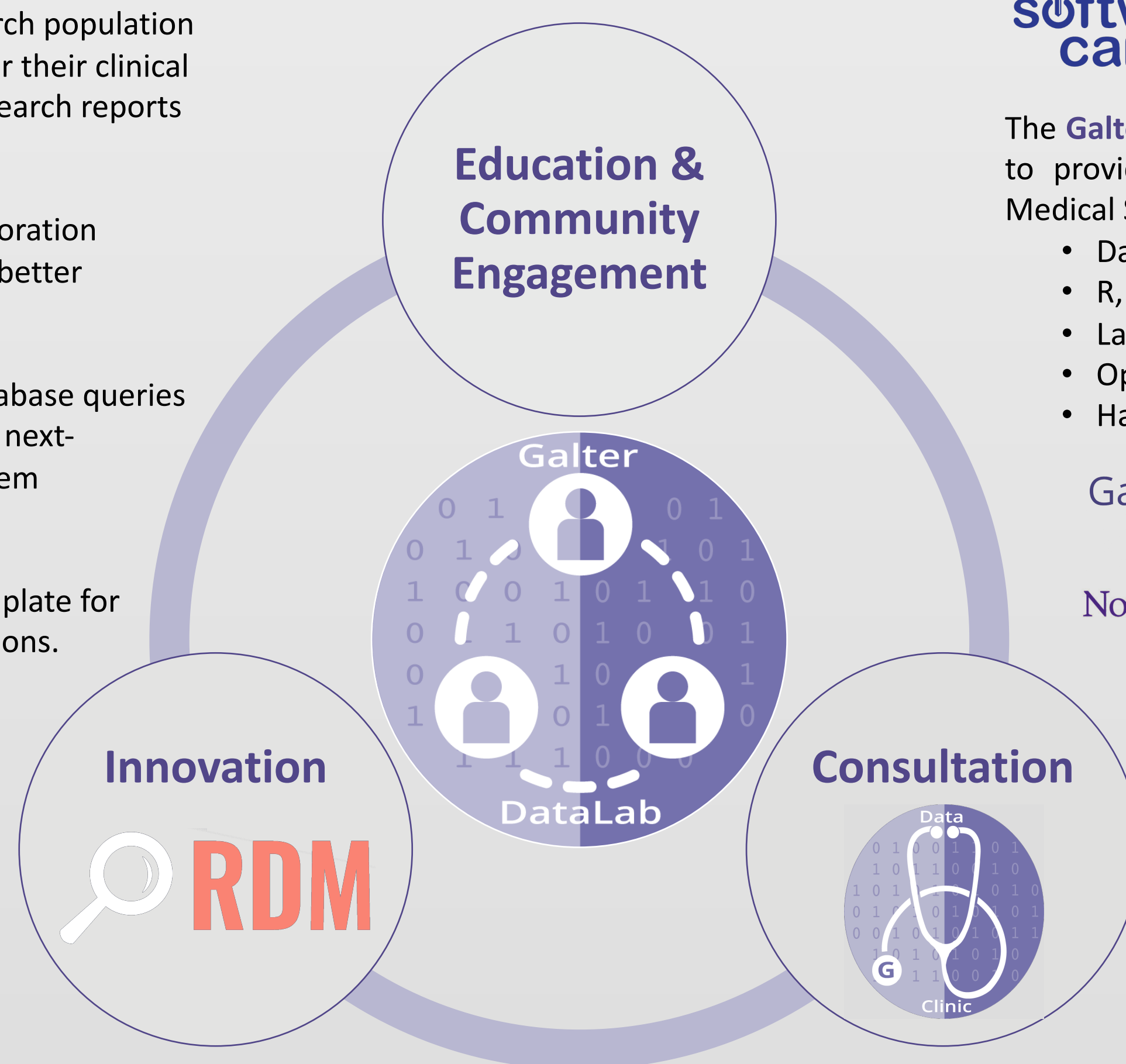NCBI National Center for Biotechnology Information

Proper collection, indexing, and preservation is vital to the discovery and dissemination of research output in scientific research. However, many research communities continue to battle the problem of "silos" at the institutional level that hinder discovery of research output. Galter Library has joined CERN and other international partners to build **InvenioRDM**[2,3], a born-interoperable research data management repository and data catalog that enables best practices in research data management, sharing, and reuse. InvenioRDM makes it easy to collect, preserve, and disseminate a wide range of research products to enhance individual and institutional visibility, promote people and their expertise, support discovery and accessibility by the international scientific community, and promote open and FAIR[4] science. With this project as well as **CTSA Personas**[5] and **Architecting Attribution**, Galter Library promotes good data practice workflows, incorporates standards and persistent identifiers, and accounts for privacy standards required for translational research.

The **DataClinic** follows a primary care model for data-related consultation. We work to solve issues related to our expertise and refer researchers to other resources for specialized or long-term support. *IRB: Institutional Review Board; OSR: Office for Sponsored Research; Research Comp: Northwestern University IT Research Computing Services; NMEDW: Northwestern Medicine Enterprise Data Warehouse; NuSeq: Next-gen Sequencing (NGS) Core*

Consultation → Referral → IRB / OSR / REDCap / Research Comp / NMEDW / Stats Core / NUSeq → Resolution

To reinforce concepts from workshops and consultations, we encourage participants to continue practicing what they've learned using self-paced, online resources. Northwestern members can access **LinkedIn Learning** and **Safari Books Online** through university licenses. Galter Library has partnered with Research Computing and others to offer **Dataquest** license seats on a quarterly basis. We recommend other resources based on subject matter and preferences.

DATAQUEST   in   O'REILLY Safari   **Other training, tools, and resources**

Personas
Understanding the Translational Workforce
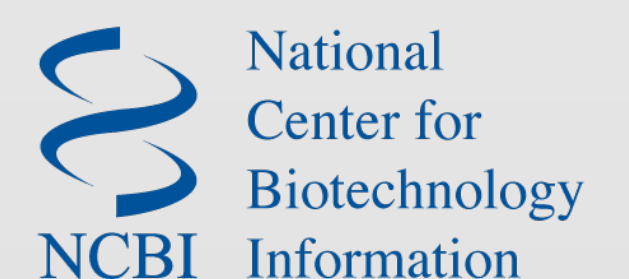Guidebook & use cases

A wide range of contributions beyond traditional papers are required to drive research. We're building on CRediT and community input to make it possible to describe, give credit for, highlight the impact of non-traditional contributions to research

images from https://thispersondoesnotexist.com/

Looking forward

- First Long-term Support (LTS) version of InvenioRDM in Summer 2021

- Continue development of training materials for clinical and translational researchers

- Enhance our workshop catalog to develop a "just-in-time" curriculum with a certificate awarded to attendees upon completion of different tracks

- Continue collaboration with the NMEDW to provide support for advanced users

galterdatalab.github.io/crdm-training

datalab@northwestern.edu
@GalterDataLab

References

1. Galter DataLab [Internet]. Northwestern University Feinberg School of Medicine. 2021. Available at: https://galter.northwestern.edu/datalab
2. InvenioRDM: The turn-key research data management repository [Internet]. European Organization for Nuclear Research (CERN), Northwestern University, and partners. 2021. Available at: https://inveniosoftware.org/products/rdm/
3. Gonzales S, Carson MB, Viger G, O"Keefe L, Allen NB, Ferrie JP, Holmes K. User Testing with Microinteractions. Information Technology and Libraries. 2021 March; 40(1):1. doi: https://doi.org/10.6017/ital.v40i1.12341
4. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18. PMID: 26978244; PMCID: PMC4792175.
5. Gonzales, S., O'Keefe, L., Gutzman, K., Viger, G., Wescott, A., Farrow, B., . . . Holmes, K. (2020). Personas for the translational workforce. Journal of Clinical and Translational Science, 4(4), 286-293. https://doi:10.1017/cts.2020.2