

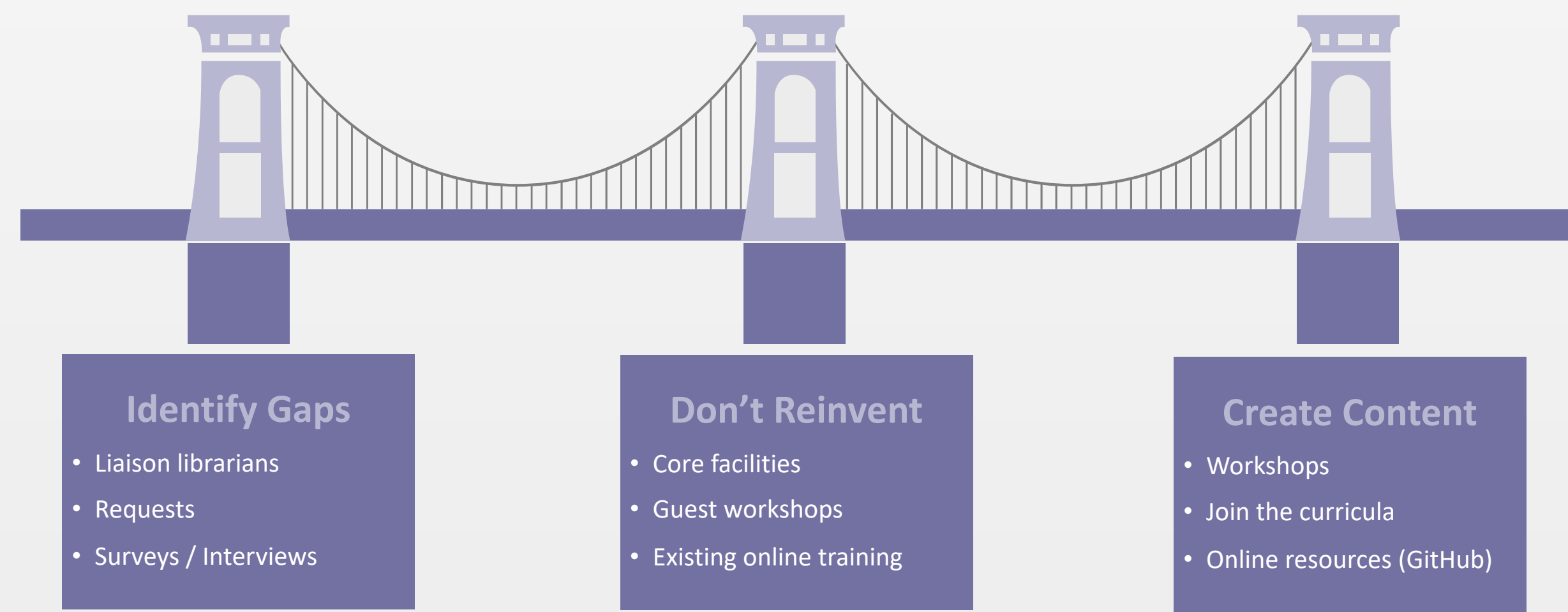
Galter DataLab: Expanded Informatics Training Through a Health Sciences Library

Matthew B. Carson, PhD, Pamela L. Shaw, MS, Sara Gonzales, MS, Kristi L. Holmes, PhD
Galter Health Sciences Library & Learning Center, Northwestern University, Chicago, IL



Bridging the skills gap for informatics researchers

The interdisciplinary nature of today's scientific research programs demands an equivalent approach to training. This is particularly relevant in areas such as informatics education, where biomedical and technical topics tend to compete for space in crowded curricula. For many students with non-technical backgrounds, graduate programs do not include sufficient technical training (basic computer literacy, programming languages, data management and analysis, data workflow management, etc.) to give them the necessary tools for success with cutting-edge research. Additionally, curricula are often slow to change and do not always support training for skills demanded by current job markets. There is a need for supplementary training methods that help those with domain knowledge acquire technical expertise. Using a variety of practices such as topical workshops and online, self-paced training platforms, demand for technical training in R, Python, SQL, and other popular tools can be met. With a mission to serve the Feinberg Medical School, the Galter Health Sciences Library & Learning Center, in partnership with university core facilities and centers, is the ideal nexus for training and consultation in best practices related to data management and analysis as well as the development of infrastructure that will serve to connect researchers, their output, and their expertise with others in the research community. Building on the library's extensive education catalog^{1,2}, we have recently established the Galter DataLab³ to support these objectives for the Feinberg School of Medicine.



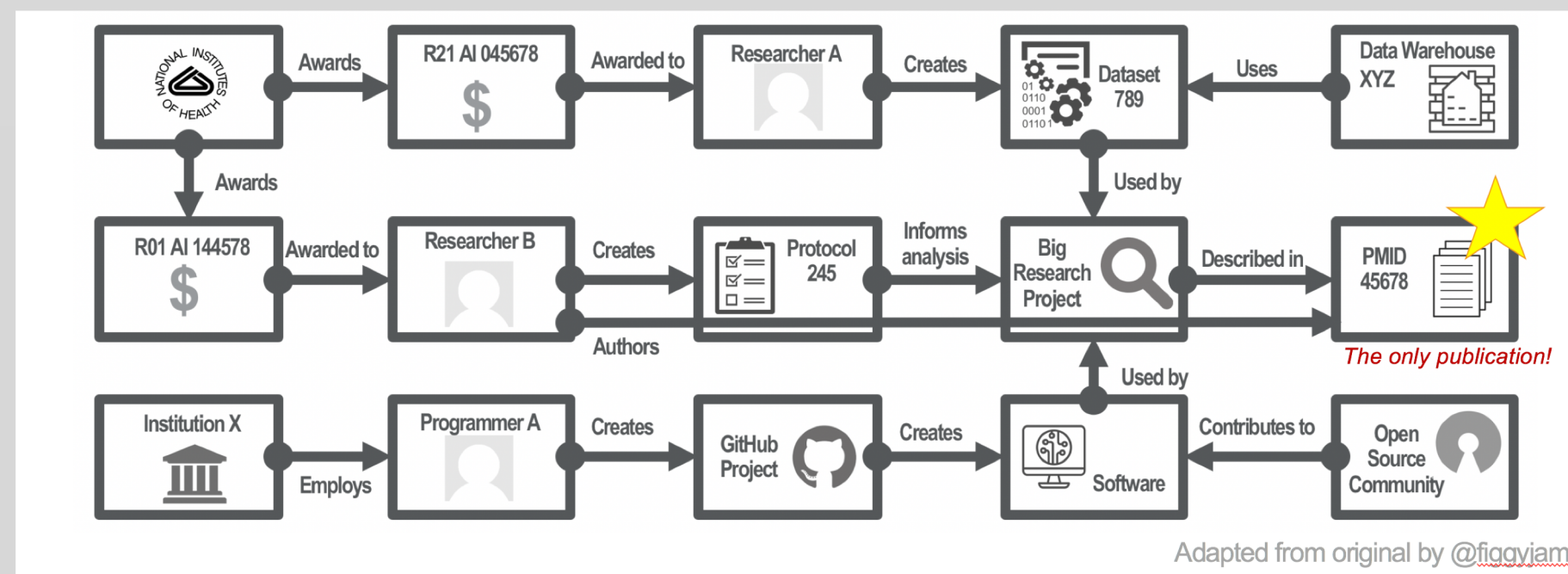
Our strategy is based on three principles. First, we identify researcher needs either through Galter's liaison librarians, by request, or by surveys and interviews. Second, we make use of existing resources, both local and remote, whenever possible. Third, if necessary, we create content that addresses researcher's needs and subsequently make that content available to the community.



Hosted by Galter Library, NUIT Research Computing, and NUCATS, the **Computational Skills in Informatics (CSI) workshop series**² on Northwestern's Chicago campus covers both general and domain-specific computational topics to help researchers improve their workflows.

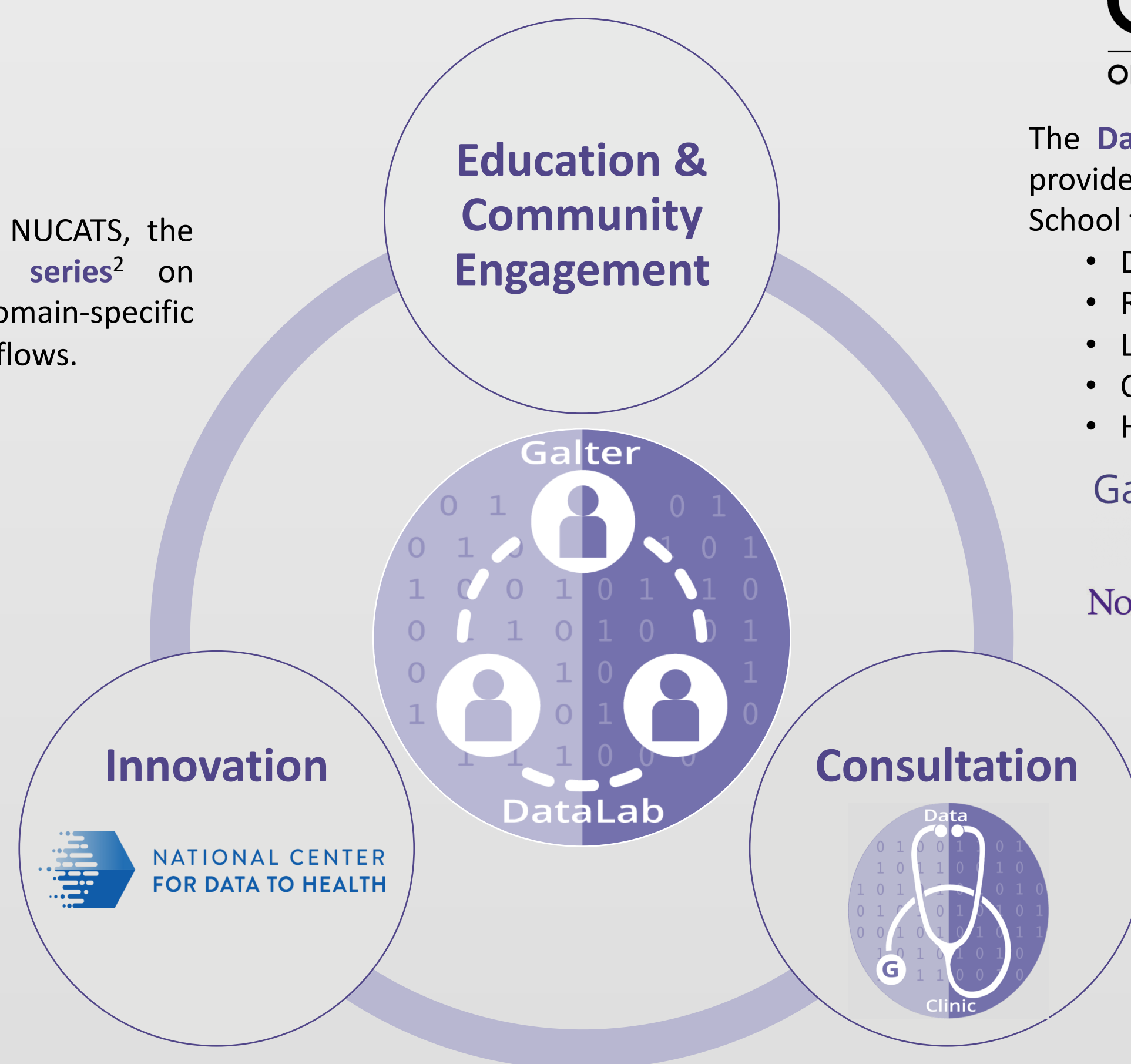


Proper collection, indexing, and preservation is vital to the discovery and dissemination of research output in scientific research. However, many research communities continue to battle the problem of "silos" at the institutional level that hinder discovery of research output. The National Center for Data to Health (CD2H)⁴ is a CTSA-wide initiative that aims to create a biomedical informatics ecosystem of people, education, expertise, standards, and software tools. As part of this effort, Galter Library is building infrastructure that can be easily deployed and managed either locally or on a cloud-based platform to collect, record, preserve, and disseminate a wide range of digital works across the translational community to enhance their visibility, promote people and their expertise, support attribution of their work, aid discovery and accessibility by the international scientific community, and support open and FAIR-TLC⁵ science. With projects such as **InvenioRDM**⁶, **Personas**, and **Architecting Attribution**, we promote good data practice workflows, incorporate standards and persistent identifiers, and account for privacy standards required for translational research.



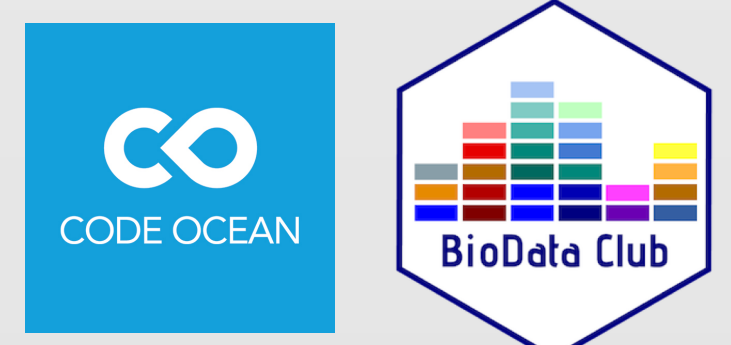
A wide range of contributions beyond traditional papers are required to drive research. We're building on CRediT and community input to make it possible to describe, give credit for, highlight the impact of non-traditional contributions to research

What is the DataLab?

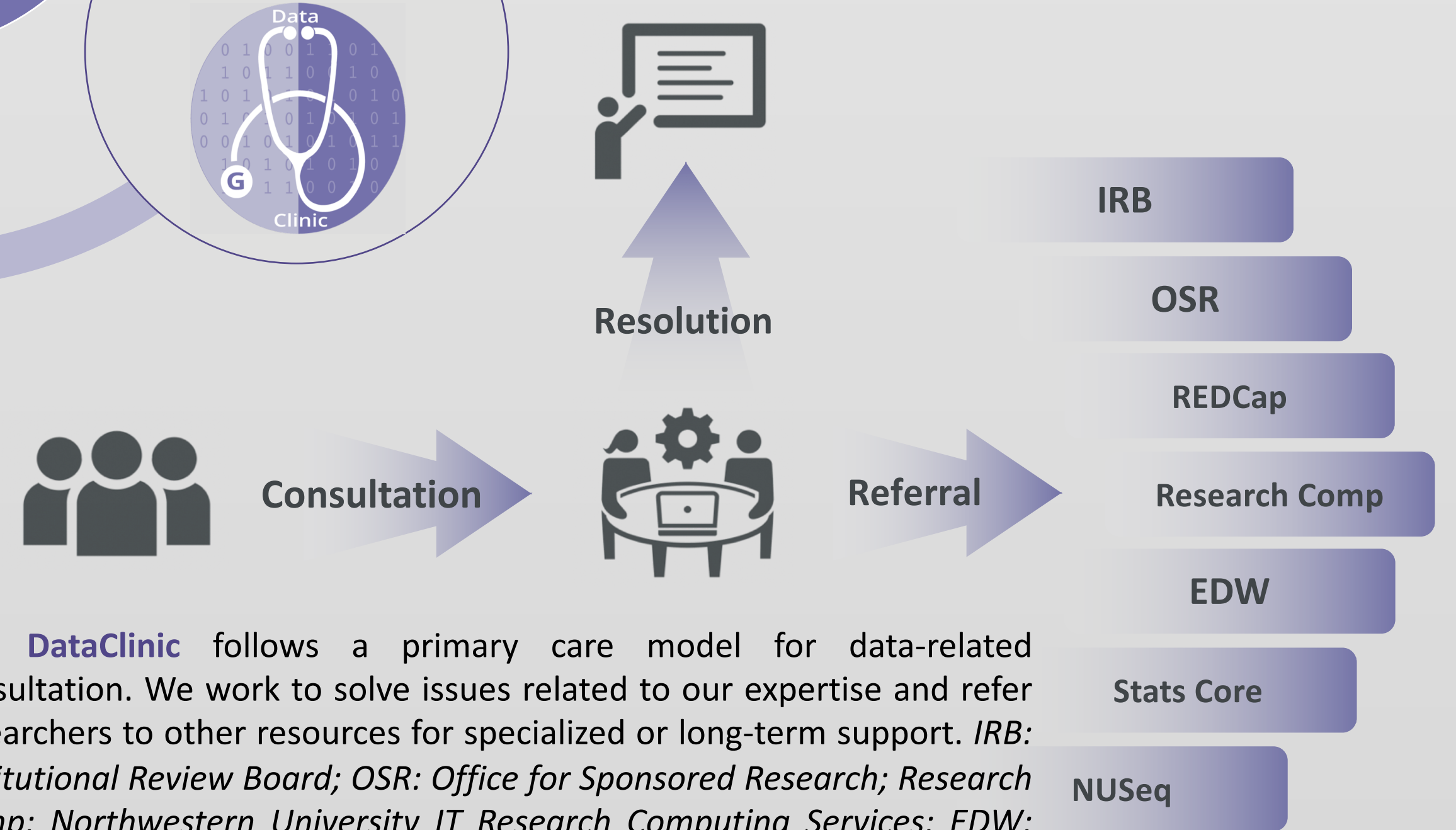
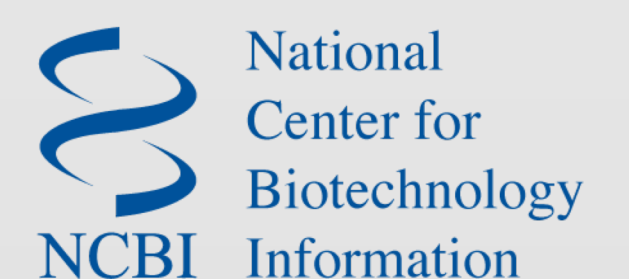


The **DataLab** works with a variety of contributors and partners to provide workshop and other educational content for Feinberg Medical School faculty, staff, and students. Topics include:

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



Galter Health Sciences Library & Learning Center
Northwestern INFORMATION TECHNOLOGY RESEARCH COMPUTING SERVICES



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; **EDW**: Enterprise Data Warehouse; **NUSeq**: Next-gen Sequencing (NGS) Core

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



Other training, tools, and resources

Looking forward

- Develop a data science workshop series for clinical researchers
- Enhance our workshop catalog to develop a "just-in-time" curriculum with a certificate awarded to attendees upon completion of different tracks
- Coming in Fall 2019:
 - R and Python workshop series on NU's Chicago campus in collaboration with NUIT Research Computing Services
 - "Integrating reproducible best practices into biomedical and clinical research: A hands-on workshop for researchers" hosted by Code Ocean and Galter DataLab



References

- Starren J, Carson M, Valenta AL. Mapping the new Core Informatics Competencies for Clinical and Translational Scientists to a Teachable Syllabus. AMIA Inspire Conference, June 28-30 2016; Columbus, OH.
- Computational Skills for Informatics Workshops [Internet]. Galter Health Sciences Library, Northwestern University Feinberg School of Medicine (AU); 2018 [cited 2018 Dec 11]. Available from: <https://galter.northwestern.edu/guides-and-tutorials/computational-skills-for-informatics-workshops>
- Galter DataLab [Internet]. Northwestern University Feinberg School of Medicine (AU); 2018 [cited 2018 Dec 11]. Available from: <https://galter.northwestern.edu/datalab>
- CTSA National Center for Data to Health (CD2H) [Internet]. U.S. Department of Health and Human Services (AU), National Institutes of Health (AU), National Center for Advancing Translational Sciences (AU); 2018 [cited 2018 Dec 11]. Available from: <https://ctsa.ncats.nih.gov/cd2h/>
- Haendel M, Su A, McMurry J, & et al. FAIR-TLC: Metrics to Assess Value of Biomedical Digital Repositories: Response to RFI NOT-OD-16-133. December 15, 2016. Zenodo. <http://doi.org/10.5281/zenodo.203295>
- Gonzales S, Viger G, Carson MB, Holmes K. "Beyond Accessibility: Taking the Institutional Repository to the Next Generation". Poster presented at the Research Data Alliance's 13th Plenary in Philadelphia, PA, April 1 - 4, 2019. doi: [10.18131/g3-b7wj-vh65](https://doi.org/10.18131/g3-b7wj-vh65) [third prize in the Best Poster Award competition]

This work was supported by the National Institutes of Health's National Center for Advancing Translational Sciences, grants U24TR002306 and UL1TR001422.