

Fred Hutch Data Needs Assessment

Version 1.0 || Jun 7, 2023

Executive Summary

Fred Hutchinson Cancer Center was formed from the merger of Fred Hutchinson Cancer Research Center and the Seattle Cancer Care Alliance in April 2022. Fred Hutch is an independent, nonprofit organization that also serves as the cancer program for UW Medicine. This relationship allows for enhanced care coordination with one of the world’s leading integrated health systems.

In May 2022, Fred Hutch hired [Jeff Leek](#) as Chief Data Officer and J Orin Edson Foundation Professor of Biostatistics. Previously, he was a professor of Biostatistics and Oncology at the Johns Hopkins Bloomberg School of Public Health and co-director of the Johns Hopkins Data Science Lab. Dr. Leek’s responsibility as CDO in the first year was to identify the data needs of the newly merged Fred Hutch and to design a comprehensive data strategy to address those needs. To conduct this study he recruited two individuals - Dr. Sean Kross whose expertise is in the area of human-computer interaction focusing on data science and Dr. Amy Paguirigan who has been a member of the Fred Hutch research community with experience across wet lab, clinical assay development, bioinformatics and computing.

The Fred Hutch data needs assessment ran from April 2022 to April 2023. The needs assessment used a mixed methods approach encompassing informal conversations, formal interviews, and ongoing collaborations addressing acute data needs. Both quantitative and qualitative data were collected from individuals across all divisions and all levels of personnel at Fred Hutch. Recruitment for the needs assessment used multiple channels including announcements on public websites, on Centernet, on social media, outreach to leadership of each division and program, user groups, a Consortium oriented Slack workspace, and through direct collaboration requests. The result of this outreach was a collection of 213 conversations representing 319 individuals and over 1,200 hours of conversation - spanning Fred Hutch research and clinical personnel, as well as a collection of administrative personnel who support research and clinical activities.

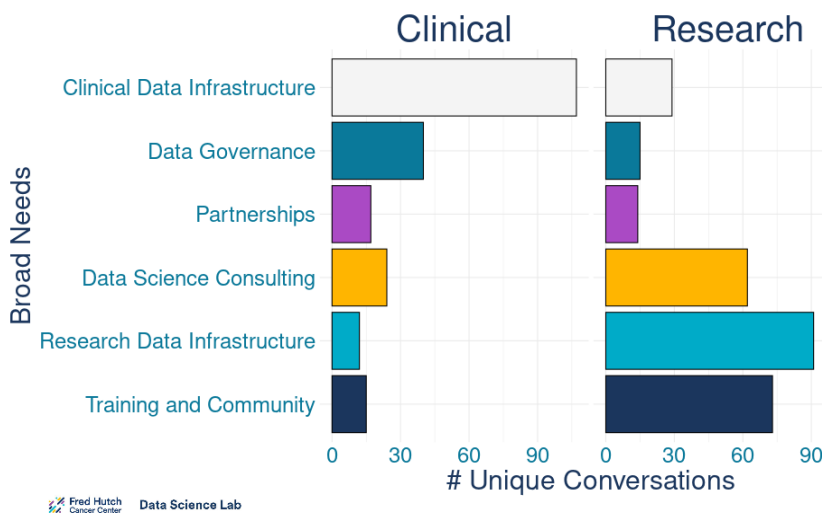


Figure 1 Summary of Data Conversations during Fred Hutch Data Needs Assessment.
 This figure represents the number of unique conversations where specific broad areas of need were raised by members of Fred Hutch’s community. Needs were separated into two groups - individuals who use Fred Hutch patient data for clinical, research, or operations purposes and the broader Fred Hutch research community who leverage a broader range of data resources. Needs were assigned to categories based on a controlled vocabulary with six broad categories (shown here and defined in Appendix A) and 26 specific categories (shown in later figures and defined in Appendix A).

The initial high level take-away of both the quantitative and qualitative data from our assessment show there are two key patterns of need in the Fred Hutch’s data community that broadly reflect our hybrid institution: a Fred Hutch patient clinical data community focused on the research, clinical, and operational use of Fred Hutch patient data and a broader Fred Hutch research data community with diverse researcher needs (**Figure 1**).

Fred Hutch Patient Clinical Data Community

The Fred Hutch patient clinical data community need distribution was driven by challenges centered around the clinical data collected about Fred Hutch patients. The Fred Hutch patient clinical data community is most interested in data related to clinic operations, clinical care, philanthropy, and clinical research such as clinical trials. The top needs are:

1. *Ease of access to comprehensive clinical data in a timely manner*
2. *Appropriate data governance systems for access to these data*
3. *Clinical data analysis support*
4. *Appropriate clinical data storage, management, and analysis infrastructure*
5. *Partnerships to support clinical care, trials and research via clinical data sharing/data use agreements*

The FHCC clinical data are highly regulated and thus administrative needs include expanded information security and compliance support and training, as well as expanded clinical data analytic support. This component of Fred Hutch's data community can be best addressed by creating and managing a centralized data infrastructure, defining clear guidelines for data governance for multiple uses, and leveraging data partnerships to augment data sources.

Broader Fred Hutch Research Data Community

The Fred Hutch research data community need distribution reflected the diversity of Fred Hutch research groups. The clinical research program as we have defined it for this study is focused on clinical research using data about Fred Hutch patients and will be served by Fred Hutch's patient clinical data program. However, there is a broad range of individual research programs and a correspondingly high diversity of data needs that span everything from individual wet labs, to large epidemiological cohort studies, to genetic studies, to massive national and international scale data coordinating centers. The top cross-lab data needs at Fred Hutch include:

1. *Access to training in data science, machine learning and reproducible research*
2. *Research data storage and computing technology and access*
3. *Research data management support to reduce costs, comply with requirements, and archive data*
4. *Data science consulting to adopt modern cloud, software, machine learning and data technologies*
5. *Data science communities for sharing and supporting best practices*
6. *Product development support for R packages, Python libraries, workflow management and Shiny apps*
7. *Partnerships coordination for philanthropy and technology*
8. *Support for data sharing agreements including execution and governance.*

This component of Fred Hutch's data community can be best addressed by creating a collection of smaller, interoperable data products and services that researchers can mix and match to best address their data needs. Due to the long tail of unique data needs for this community beyond these shared groupings, addressing the evolving, fine-grained needs of the community will require active engagement and community connections to continue to adapt to fluidly changing needs over time.

While we have reported here on the initial needs identified during our conversations over the past year with Fred Hutch's community there will be ongoing and new data needs over time. We plan to continue to engage the community through conversation and collaboration to ensure that Fred Hutch's data program is responsive to the diversity of needs of Fred Hutch's community in an ongoing way.

Summary and Next Steps

Fred Hutch's data community is both strong and diverse. Integration and support of all scientists, providers, and administrative personnel to leverage data will have significant positive impacts on patient care, clinical trials, research productivity, and institutional function. Fred Hutch's Strategic Data Plan 2023 outlines how this support can be delivered through an integrative approach leveraging existing data expertise across the enterprise.