



MARCH 2024



Table of Contents

Executive Summary	2
Program Overview and Structure	3
Program Accomplishments	4
Community Engagement.....	4
Impact and the Path Moving Forward	5
Conclusion.....	6



Executive Summary

The RADx Data Hub was established to support researchers in accessing curated and de-identified COVID-19 data, allowing them to find, aggregate, and perform data analyses in a cloud-enabled platform. The centralized cloud resources offered by the Data Hub provide study data and analytic capacities for researchers to better understand diagnostic COVID-19 testing tools, technologies, and future pandemics, especially for underserved populations and those who are disproportionately impacted by COVID-19. Additionally, these resources support rapid review of data requests from authorized users through the NIH RADx Data Access Committee to ensure proposed research purposes are consistent with participant protections and the data use agreement. Access to the Data Hub allows users to query a wide range of health data including clinical, behavioral, social determinants of health, survey, interview, diagnostic and testing results, viral sequences, output from smart sensors, self-reported symptoms, and imaging data.

The RADx Data Hub was built as a modernized, cloud-based data repository by curating collected data elements and creating anonymous COVID-19 research data optimized for cost, performance, and streamlined data access. This platform acts as a centralized data and metadata collection point across the four RADx programs that collect and submit the data (RADx-Digital Health Technologies, RADx-rad, RADx Tech, RADx-UP). Once the data is collected, it is systematically mapped, standardized, and curated across the various RADx studies to ensure consistent data collection and alignment to designated RADx Common Data Elements (CDEs).

The RADx Data Hub was also designed to optimize engagement with the research community. The Data Hub website allows users to explore the various data variables and study information for each registered RADx study. The RADx Data Hub support team also provides community outreach webinars that discuss the RADx data and the Data Hub functionality.

To date, the RADx Data Hub contains 138 studies from across the four RADx programs that are available to researchers. Each study contains data that is harmonized to 12 RADx CDEs and up to 133 mapped variables that are study-dependent. The RADx programs continue to deposit data



validation that the 18 identifiers of PHI are removed to protect participant privacy in the research data. Once that is verified, the data is analyzed to identify, curate, standardize, and map the





programs. The team found that the program-specific data dictionaries sometimes contained weak or missing descriptions for variables when these were merged into a common data dictionary. However, a data dictionary browser designed at the onset of data collection could be a significant improvement to the data collection and harmonization process. Program teams continue to optimize this process, both for uses in the current RADx Data Hub and for the development of future data repositories.

For the process of CDE selection and maintenance, the RADx Data Hub team recommends that the CDEs cover a wide range of health data and should be automated, when possible. The collected data for the RADx Data Hub often required manual mapping and path review, both of which could benefit from automation. If algorithms were to be developed at the onset of data collection, it could simplify data collection and could broaden the evaluation of more variables.

In the case that automation is unattainable, then researchers should consider the scalability and the amount of manual data harmonization/curation that will be required in the data collection process. If these considerations can be applied early in the process, it can improve the efficiency of data processing and can help improve any harmonization algorithms required.

Lastly, early community outreach is critical to promoting a newly data-rich resources that will be adopted and used by researchers. By reaching these communities through webinars, community meetings, social media, etc., the RADx Data Hub Team can inform resources, stimulate scientific research interest and promote early adoption of the platform.

Conclusion

The RADx Data Hub represents one of the largest collections of NIH COVID-19 testing data available to researchers that allows researchers to explore, access, and analyze COVID-related data developed through NIH RADx programs. The Data Hub will provide curated and harmonized data across ~ 200 RADx studies and all data/metadata will be findable, accessible, interoperable, and reusable. These data will be of interest to researchers focused on testing approaches, diagnostic and surveillance testing through wearable sensors and non-traditional approaches, and implementation strategies to enable an enhance testing of underserved, under-

