Journal of Diabetes Science and Technology
1–4

DOI: 10.1177/19322968241306129
journals.sagepub.com/home/dst
Sage

How the Diabetes Research Hub Will Modernize and Enhance Diabetes Data Utilization

Alessandra T. Ayers, BA[1], Cindy N. Ho, BA[1], Emma Friedman, MA[2], Juan Espinoza, MD[2], Shahid N. Shah, MSc[3], and David C. Klonoff, MD, FACP, FRCP (Edin), Fellow AIMBE[4]

Abstract

The Diabetes Research Hub (DRH) is a centralized data management system and repository that will revolutionize how diabetes data are gathered, stored, analyzed, and utilized for research. By harnessing advanced analytics for large datasets, the DRH will support a nuanced understanding of physiological patterns and treatment effectiveness, ultimately advancing diabetes management and patient outcomes. This is an opportune time for researchers who are collecting continuous glucose data and related physiological data sources, to leverage the capabilities of the DRH to enhance the value of their data.

Keywords

CGM, diabetes, repository, research hub

Introduction to the Diabetes Research Hub

The Diabetes Research Hub (DRH), developed through a collaboration of Diabetes Technology Society, Ann & Robert H. Lurie Children's Hospital of Chicago, and the Netspective Foundation, is a comprehensive platform aimed at accelerating breakthroughs in diabetes research. It will integrate diverse data sources including continuous glucose monitor (CGM) data, patient demographics, and related fitness and nutritional data generated from clinical trials and other studies to empower researchers. The DRH will support collaboration among research institutions, health systems, device manufacturers, and funding agencies. It provides both "offline" (edge-based) tools and a cloud-based SaaS platform for data processing, advanced analytics, machine learning integration, and tools for data visualization and sharing. The current and future components of the DRH are shown in Figure 1.

Importance of CGM

Continuous glucose monitors have revolutionized diabetes management by providing real-time, minute-by-minute glucose level tracking, offering a dynamic view of patients' glycemic control in everyday life. The CGMs capture crucial data on how patients truly respond to treatments over time, particularly the variability of glucose levels throughout the day and night. The DRH has integrated CGMs as a foundational element for understanding diabetes beyond traditional clinical trials. The data from CGMs are essential for exploring personalized care and enhancing the precision of diabetes treatments. Furthermore, CGMs can also be used to assess people with prediabetes and people without diabetes, expanding their use in other fields of research.[1]

While CGMs are pivotal for monitoring glucose, their value grows when integrated across multiple institutions and across studies over time. The DRH will facilitate these integrations, allowing researchers to harness a complete view of diabetes management by connecting CGM data across different settings and research efforts.

Correction (January 2025): Editorial updated to correct the title. Please see <https://doi.org/10.1177/19322968251317190> for more details.

[1]Diabetes Technology Society, Burlingame, CA, USA
[2]Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
[3]Netspective Communications LLC, Silver Spring, MD, USA
[4]Diabetes Research Institute, Mills-Peninsula Medical Center, San Mateo, CA, USA

Corresponding Author:
David C. Klonoff, MD, FACP, FRCP (Edin), Fellow AIMBE, Diabetes Research Institute, Mills-Peninsula Medical Center, 100 South San Mateo Drive, Room 1165, San Mateo, CA 94401, USA.
Email: dklonoff@diabetestechnology.org

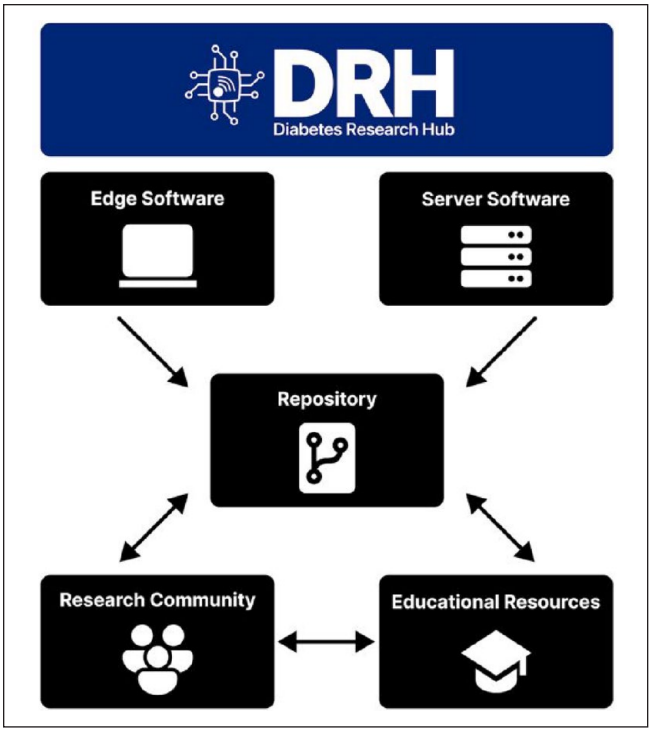

Figure 1. Five different components of the DRH, including edge software, server software, repository capabilities, the ability to build a research community, and educational resources. Figure courtesy of Ashley DuNova.
Abbreviation: DRH, diabetes research hub.

The DRH involves a diverse group of key stakeholders, including research institutions, the National Institutes of Health (NIH), health delivery organizations, medical device manufacturers, and public health organizations. These stakeholders will play an essential role in defining the DRH's evolution and scope, designing meaningful governance, identifying barriers to accelerating research, and supporting the collection, integration, and utilization of CGM data to advance diabetes research and care.

Integrating Data for a Complete Picture

The DRH will facilitate advanced data integration, making CGMs the gateway to a larger ecosystem of diabetes-related data. Researchers will be able to leverage CGM data combined with other continuously monitored sensor data. Such data can be collected from cardiovascular monitors, exercise monitors, environmental monitors, or other types of analyte monitors to identify complex patterns and correlations that are impossible to capture with a single data source. For example, combining CGM data across multiple cohorts of patients across multiple studies over time allows exploration of how genetic predispositions affect glucose variability and response to treatment in varied patient populations.[2]

By encouraging the integration of multiple data sources across institutions, the DRH moves toward a future where researchers can use newly surfaced data with research direction decisions informed by a comprehensive understanding of multiple patients' unique data profiles.[3]

Data Sharing and Data Reuse Challenges

Data repositories play an important role in facilitating data sharing, supporting open science, enhancing reproducibility efforts, and creating new research questions. The CGM data have not been widely shared via repositories or other means. This lack of sharing is likely due to a number of challenges related to processing, cleaning, and de-identifying CGM data in order for it to be fit for deposit. When CGM data are shared, the data are typically hosted as discrete data objects (e.g., .csv files) with little metadata that can guide potential researchers as to the utility of a specific dataset. This results in researchers spending time and effort (1) evaluating datasets to see whether they are standardized to contain the same types of information, (2) transforming and normalizing data, and (3) removing extraneous information from datasets.

The DRH aims to alleviate these issues, first, by standardizing data structure and definitions so that datasets from multiple sources can be combined. Datasets may have different structures, with varying numbers and types of columns or attributes, which is known as structural heterogeneity. Even when datasets contain conceptually similar information, the format and representation of data values may differ significantly, which is known as lexical heterogeneity. Second, the DRH will clean datasets either for cloud computing or for edge computing so that they can be easily stored, analyzed, and combined. Third, the DRH's platform will provide a unified framework that maintains provenance and documentation about how datasets were incorporated, ensuring that future researchers will have access not only to the data but also to the context in which it was collected and used. This framework will empower researchers to conduct more robust and comprehensive analyses, ultimately leading to improved insights and better outcomes for diabetes care.

In addition, a significant barrier to data sharing is that content from many studies may be difficult to share because it contains identifiable information at the patient level, even if such information is not crucial to the study. The DRH will address this barrier by incorporating native deidentification and anonymization tools that help protect patient health information (PHI) by eliminating it where appropriate, while maintaining secure links to source data when necessary. Most studies do not require access to PHI, but maintaining deidentified or anonymous datasets is a challenging task—one that the DRH will be prioritized as a core feature.[4] Through the DRH, researchers will be able to take comparative effectiveness research to new levels by sharing

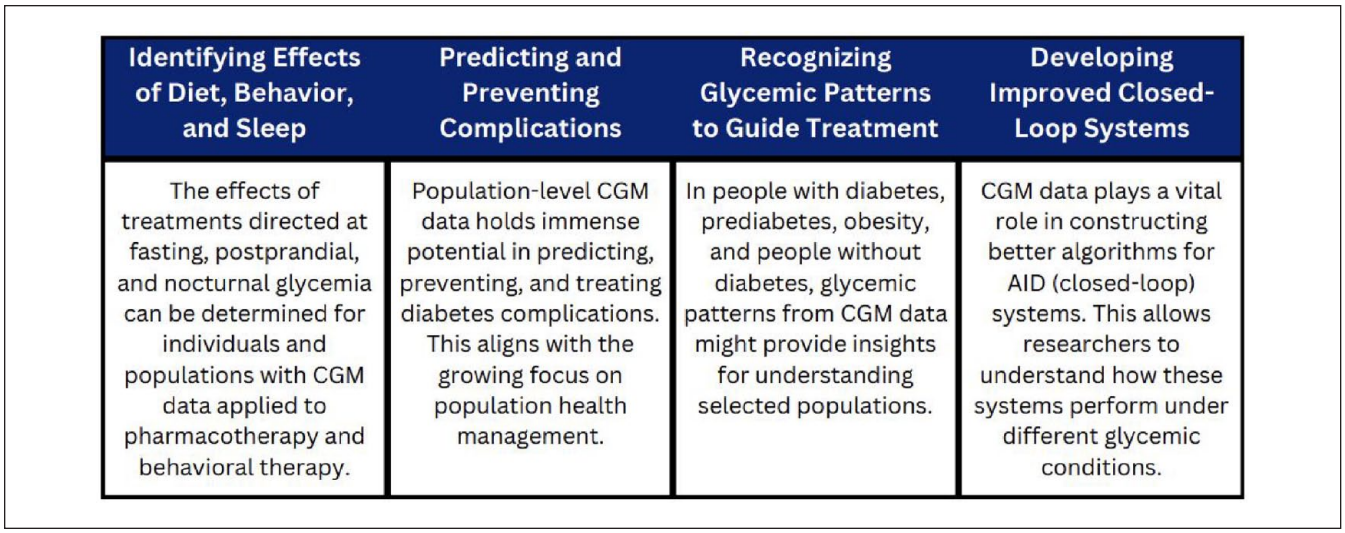

Figure 2. Four benefits of using the DRH to facilitate the use of CGM Data. Figure adapted from https://drh.diabetestechnology.org/.
Abbreviations: AID, automated insulin delivery; CGM, continuous glucose monitor.

data and leveraging integrated datasets that represent the full complexity of diabetes care.

Challenges and Solutions

Data quality, standardization, and privacy are major challenges in diabetes research—different studies lack uniformity, and patient privacy concerns hinder data sharing.[5] To address these challenges, the DRH will deliver several key features to significantly improve the research process and data utilization.

- Data Ingestion and Preprocessing: The DRH will provide enhanced data ingestion and preprocessing capabilities to streamline how CGM data from multiple sources are collected, formatted, and prepared for analysis. This feature will allow researchers to easily ingest data from various clinical trials and studies, ensuring consistency and improving efficiency in handling large datasets.[6]
- Data Cleaning and Analysis: The platform will offer advanced data cleaning and analysis tools, leveraging artificial intelligence (AI) to automatically detect and correct inconsistencies in the data. This feature will ensure that researchers have access to high-quality, reliable data that can be used for accurate and meaningful analyses, ultimately improving the robustness of their diabetes research.
- Visualization and Exploration: The DRH will provide enhanced data visualization and exploration tools, enabling researchers to generate insights more intuitively. With interactive charts and visual analytics, researchers will be able to explore CGM data trends, correlations, and patterns in real time, fostering better understanding and discovery of new research questions.
- Application Programming Interface (API) Access for Custom Applications: To facilitate custom research needs, the DRH will offer API access, allowing researchers and developers to build custom applications on top of the DRH platform. This will empower institutions to create tailored solutions for data analysis, visualization, or integration with other health care systems, enhancing the flexibility and adaptability of the platform to meet diverse research requirements.

The DRH will deliver even more in the future through advanced analytics, AI, and machine learning that help clean and harmonize disparate datasets. In addition, robust data governance and privacy frameworks will ensure compliance with health regulations, making the data both reliable and ethically sound.[7]

Benefits

The DRH will offer several significant benefits. The DRH will help researchers to understand existing studies better and help them develop new interventions and clinical insights more quickly. This accelerated research process will not only lead to timely discoveries but also will enhance the speed at which new treatment options are brought to patients, and thus contribute to improved care.

By providing a platform that integrates CGM data across multiple institutions and studies, the DRH will facilitate research that can enable data-driven personalized care, ultimately improving glycemic control, reducing complications, and enhancing quality of life for patients with diabetes. The DRH will also support innovation in medical devices and

therapeutics through tools and datasets that can encourage and facilitate the use of CGM data as a measure of performance or outcomes. This collaborative environment will encourage the development of next-generation devices that are more effective, accurate, and tailored to individual patient needs. These benefits of using CGM data, facilitated by the DRH, are summarized in Figure 2.

Future Directions

The future of diabetes research is promising with the expanding use of CGMs as well as other digital health tools, such as wearable devices and mobile apps, all enabling collection of data in real time. Platforms like the DRH will drive collaboration between health systems, researchers, and patients, making integrated care and personalized treatment plans more achievable. Future integration of real-time data and AI-driven analytics will make diabetes management even more effective. Platforms like the DRH will not only collect and organize CGM data and other data, but also generate actionable insights, to transform the landscape of diabetes care.

Getting Involved With DRH

The DRH is actively seeking collaborators interested in shaping its technologies, processes, and community. Researchers who use CGM data have several ways to participate, including:

- Participate in surveys and interviews to help us understand the pain points and technical needs of researchers;
- Download and test our free, open-source edge software and provide us feedback on features and functionality;
- Join our advisory board;
- Explore current datasets and upload your own.

More details about each of these opportunities can be found at https://drh.diabetestechnology.org/.

Abbreviations

AI, artificial intelligence; AID, automated insulin delivery; API, application programming interface; CGMs, continuous glucose monitors; DRH, Diabetes Research Hub; NIH, National Institutes of Health; PHI, patient health information.

Acknowledgment

The authors acknowledge Annamarie Sucher-Jones for her editorial expertise.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: ATA is a consultant for Liom. CNH is a consultant for Liom. EF has nothing to disclose. JE receives federal funding from FDA, NIMHD, and NCATS and is a consultant for Sanofi. SNS has nothing to disclose. DCK is a consultant for Afon, embecta, Glucotrack, Lifecare, Novo, Samsung, SynchNeuro, and Thirdwayv.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Alessandra T. Ayers https://orcid.org/0009-0000-3054-3207
Cindy N. Ho https://orcid.org/0009-0008-3067-1004
Emma Friedman https://orcid.org/0009-0009-2537-9508
Juan Espinoza https://orcid.org/0000-0003-0513-588X
Shahid N. Shah https://orcid.org/0000-0001-8481-6493
David C. Klonoff https://orcid.org/0000-0001-6394-6862

References

1. Klonoff DC, Nguyen KT, Xu NY, Gutierrez A, Espinoza JC, Vidmar AP. Use of continuous glucose monitors by people without diabetes: an idea whose time has come. *J Diabetes Sci Technol*. 2023;17(6):1686-1697. doi:10.1177/19322968221110830.
2. Aanstoot HJ, Varkevisser RDM, Mul D, et al. Cohort profile: the "Biomarkers of heterogeneity in type 1 diabetes" study-a national prospective cohort study of clinical and metabolic phenotyping of individuals with long-standing type 1 diabetes in the Netherlands. *BMJ Open*. 2024;14(6):e082453. doi:10.1136/bmjopen-2023-082453.
3. Klonoff DC, Florez JC, German M, Fleming A. The need for precision medicine to be applied to diabetes. *J Diabetes Sci Technol*. 2020;14(6):1122-1128. doi:10.1177/1932296819894295.
4. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. *J Med Internet Res*. 2019;21(5):e13484. doi:10.2196/13484.
5. Xu NY, Nguyen KT, DuBord AY, et al. The launch of the iCoDE standard project. *J Diabetes Sci Technol*. 2022;16(4):887-895. doi:10.1177/19322968221093662.
6. Rai A, Maro JC, Dutcher S, Bright P, Toh S. Transparency, reproducibility, and replicability of pharmacoepidemiology studies in a distributed network environment. *Pharmacoepidemiol Drug Saf*. 2024;33(6):e5820. doi:10.1002/pds.5820.
7. Shang T, Zhang JY, Dawson J, Klonoff DC. Benefits of conformity assessment for cybersecurity standards of diabetes devices and other medical devices. *J Diabetes Sci Technol*. 2021;15(4):727-732. doi:10.1177/19322968211018186.