Delivering Real World Patient Data for Clinical and Translational Research: Approaches from Four Institutions

Christopher A. Harle, PhD[1], Daniella Meeker, PhD[2], Shyam Visweswaran, MD, PhD[3] Thomas R. Campion, Jr., PhD[4], Boyd M. Knosp, MS[5]

[1]University of Florida Health, Gainesville, FL; [2]Southern California Clinical and Translational Science Institute, Los Angeles, CA; [3]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; [4]Clinical & Translational Science Center, Weill Cornell Medicine, New York, NY; [5]Institute for Clinical & Translational Science, University of Iowa, Iowa City, IA

Abstract

Demand for real-world health data for research has never been higher. Moreover, the capacity to deliver real-world data services is uniquely strained by competition for labor, the changing nature of work, and information security threats. To overcome these challenges, Enterprise Data Warehouses for Research (EDW4Rs) are a critical infrastructure of Clinical and Translational Science Award (CTSA) programs, research consortia, and learning health systems. This panel presents perspectives from four institutions, describing local approaches to operating EDW4R with respect to challenges and successes, such as cloud computing, data governance, workforce development, data service design, engaging the scientific community, and sustainability.

Introduction

Enterprise Data Warehouses for Research (EDW4Rs) are a critical infrastructure that feed real world data to programs such as Clinical and Translational Science Award (CTSA) hubs, research consortia and learning health systems. For the past four years, two authors (TRC, BMK) have led a CTSA working group to explore the operational characteristics of these data infrastructures that, among other activities, enable the secondary use of data from electronic health record (EHR) systems[1,2]. Although EDW4Rs are increasingly widespread, variability in methods exists across institutions, and optimal approaches are unknown. The goal of this panel is to advance literacy about EDW4Rs through descriptions of approaches from four different academic medical centers. Along with a description of institutional approaches, panelists will discuss key EDW4R issues including cloud computing, data governance, workforce development, engaging the scientific community, and sustainability.

Demand for real-world health data for clinical and translational research has never been higher. Moreover, the capacity to deliver real-world data services for research is uniquely strained by competition for labor, the changing nature of work, and information security challenges. Therefore, this panel is urgently needed so that providers and consumers of research data services can share experiences and discuss paths forward.

CTSA consortium, as well as many healthcare systems and hospitals nationally and internationally, re-use clinical data for a wide range of applications in research. The EDW4R is an aggregation of clinical and extant data into a single data system that is used to inform a range of research functions including basic and applied informatics studies; observational studies including prospective, retrospective, and case-control; quality improvement research; participation in data-driven research networks; clinical trial cohort identification for recruitment; and population health studies. Development and maintenance of the EDW4R requires considerable resources for a number of reasons including conversion of data into a research-ready format, enablement of a secure and compliant environment for highly sensitive data, governance decisions regarding data access and sharing, and guidance for appropriate scientific interpretation of the data. The operational characteristics of the EDW4R vary across CTSA hubs and other healthcare provider organizations, an indication of the newness of this infrastructure and the rapid evolution of the EDW4R as an institutional resource.

Panelists

With moderation from Boyd Knosp, MS, each panelist will describe the EDW4R approach at their institution.

- Christopher Harle, PhD is Professor and Chief Research Information Officer (CRIO) at the University of Florida (UF) Health. As a faculty member, Dr. Harle leads a research program focused on the design,

adoption, use, and value of health information systems. As CRIO, Dr. Harle provides leadership around governance and operations of the UF Health Integrated Data Repository (IDR), an enterprise data warehouse that serves as UF's EDW4R. On this panel, Dr. Harle will present a conceptual framework being used at UF Health to support research data service strategy and operations. The framework derives from operations management approaches to successfully breaking the tradeoff between efficiency and service in the face of customer-introduced variability. He will also discuss ongoing efforts to effectively serve trainees and research teams that have a wide range of disciplinary backgrounds, data science skills, and application domain interests.

- Daniella Meeker, PhD is the co-Director of the Southern California Clinical Translational Sciences Institute and leads the Los Angeles County Department of Health Services (LACDHS) Informatics and Analytics Core Facility. The Informatics Core was formed by the LACDHS Research Oversight Board using a governance framework that balances investigator-initiated research with operational priorities of the health system[3]. By combining resource contributions and IRB reliance agreements between LACDHS and its academic partners (UCLA, USC, and Charles Drew University), informatics services have been established for the first time as part of an embedded support team. Dr. Meeker will discuss the leadership, oversight, prioritization, and policy framework, the embedded staffing model, and how enterprise data warehousing and self-service tools meet both research and operational needs.
- Shyam Visweswaran, MD, PhD, is Associate Professor of Biomedical Informatics and Director of the Informatics Core for the University of Pittsburgh Clinical and Translational Science Institute (CTSI). He is a PD/PI for the All of Us Pennsylvania research project that is a component of the NIH Precision Medicine Initiative and is the Data Harmonization lead for the Accrual of patients to Clinical Trials (ACT) network. He also directs the development and implementation of an EDW4R called Neptune. Dr. Visweswaran will discuss the atomic design of the EDW4R where data are stored at a high level of granularity as represented in source systems, how data is integrated from multiple EHR systems as well as from other sources, processes that deliver data from the EDW4R to analytics-oriented data marts and self-service tools, and how the EDW4R meets the research needs of a large academic medical center [4].
- Thomas R. Campion, Jr., PhD is Associate Professor of Population Health Sciences and CRIO at Weill Cornell Medicine in New York City. As CRIO, Dr. Campion leads efforts to support investigators with electronic patient data, especially from EHR systems made accessible through an EDW4R. Funded by the CTSA and other sources, Dr. Campion and team deliver Architecture for Research Computing in Health (ARCH), which matches investigators with informatics tools and services (e.g., i2b2, REDCap, PCORnet, natural language processing) to support science [5]. He will discuss strengths and limitations of the ARCH method for engaging investigators through people, process, and technology as well as potential generalizability of the approach to other settings.
- Moderator: Boyd Knosp, MS is Associate Dean for IT and Associate Director for Biomedical Informatics at the University of Iowa (UI) Healthcare. Boyd leads the implementation of UI's enterprise data warehouse for research (EDW4R) and directs the Iowa Health Data resource, an institutional investment in Health Science data infrastructure that extends the UI EDW4R. Boyd will moderate the panel, identifying cross-cutting concepts in the panelist presentations and will engage the audience in a dialogue about key challenges and opportunities with EDW4Rs.

Discussion

Audience members will be encouraged to engage panelists as well as respond to questions posed by the moderator, including but not limited to the following.

- How has your EDW4R impacted research?
- Which data are most important to be available?
- How does leadership (both research and IT) define/measure EDW4R success?
- How do you pay for EDW4Rs?

- How are you managing the increasing demands for data?
- What tools/techniques are you using to facilitate researcher access and address demand to EDW4R data?
- What does the ideal EDW4R user experience look like?
- What are your frustrations in EDW4R access and use?
- How can the EDW4R/informatics community make platforms easier for informaticians and non-informaticians
- What is your approach wrt EDW4R and cloud computing? Social Determinants of Health? Text analytics?

Conclusion

As panel organizer, Mr. Knosp confirms all discussants have agreed to participate in the panel.

References

1. Campion, TR, Jr., Craven CK, Dorr DA, Knosp BM, Understanding enterprise data warehouses to support clinical and translational research. J Am Med Inform Assoc, 2020. 27(9): p. 1352-1358.
2. Knosp BM, Craven CK, Dorr DA, Bernstam EV, Campion TR, Jr., Understanding enterprise data warehouses to support clinical and translational research: Enterprise information technology relationships, data governance, workforce, and cloud computing. Journal of the American Medical Informatics Association. 2021.
3. Meeker D, Fu P, Jr., Garcia G, Dyer IE, Yadav K, Fleishman R, Yee HFJr, Establishing a research informatics program in a public healthcare system: a case report with model documents, J Am Med Inform Assoc, 2021.
4. Visweswaran S, McLay B, Cappella N, Morris M, Milnes JT, Reis SE, et al., An atomic approach to the design and implementation of a research data warehouse. J Am Med Inform Assoc. 2021.
5. Campion TR, Jr., Sholle ET, Pathak J, Johnson SB, Leonard JP, Cole CL, An architecture for research computing in health to support clinical and translational investigators with electronic patient data. J Am Med Inform Assoc. 2021.