

Title:

Federated Data Health Networks Hold Potential for Accelerating Emergency Research

Authors:

1. Prashant Mahajan, MD, MPH, MBA, Departments of Emergency Medicine and Pediatrics, University of Michigan, Ann Arbor, Michigan (pmahajan@med.umich.edu)
2. Charles Macias, MD, MPH, Department of Pediatrics Emergency Medicine, University Hospitals Rainbow Babies and Children's Hospital, Cleveland, Ohio (Charles.Macias@UHhospitals.org)
3. Amie Barda, PhD, Department of Pediatric Emergency Medicine, University Hospitals Rainbow Babies and Children's Hospital, Cleveland, Ohio (amie.barda@uhhospitals.org)
4. Christopher M Fung, MD, Department of Emergency Medicine, University of Michigan, Ann Arbor, Michigan (chfung@med.umich.edu)

Corresponding author: Prashant Mahajan, MD, MPH, MBA
Department of Emergency Medicine
University of Michigan
1540 E. Hospital Drive, CW 2-737
Ann Arbor, MI 48109-4260
(734) 763-9309
pmahajan@med.umich.edu

Word Count: 1946

Key words: Pediatric Emergency Medicine, Emergency Care, Data Model, Global Research

Conflicts of interest: The authors have no conflicts of interest relevant to this article to disclose.

Abstract:

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/emp2.12968](#).

This article is protected by copyright. All rights reserved.

Multi-center research networks often supported by centralized data centers are integral in generating high quality evidence needed to address the gaps in emergency care. However, there are substantial costs to maintain high functioning data centers. A novel distributed or federated data health networks (FDHN) approach has been used recently to overcome the shortcomings of centralized data approaches. A FDHN in emergency care is comprised of a series of decentralized, interconnected emergency departments (EDs) where each site's data is structured according to a common data model that allows data to be queried and/or analyzed without the data leaving the site's institutional firewall. To best leverage FDHNs for emergency care research networks, we propose a stepwise, two-level development and deployment process - creating a lower resource requiring Level I FDHN capable of basic analyses, or a more resource intense Level II FDHN capable of sophisticated analyses such as distributed machine learning. Importantly, existing electronic health records based analytical tools can be leveraged without substantial costs implications for research networks to implement a Level 1 FDHN. Fewer regulatory barriers associated with a FDHN has a potential for diverse, non-network EDs to contribute to research, foster faculty development, and improve patient outcomes in emergency care.

Abstract:

Multi-center research networks often supported by centralized data centers are integral in generating high quality evidence needed to address the gaps in emergency care. However, there are substantial costs to maintain high functioning data centers. A novel distributed or federated data health networks (FDHN) approach has been used recently to overcome the shortcomings of centralized data approaches. A FDHN in emergency care is comprised of a series of decentralized, interconnected emergency departments (EDs) where each site's data is structured according to a common data model that allows data to be queried and/or analyzed without the data leaving the site's institutional firewall. To best leverage FDHNs for emergency care research networks, we propose a stepwise, two-level development and deployment process - creating a lower resource requiring Level I FDHN capable of basic analyses, or a more resource intense Level II FDHN capable of sophisticated analyses such as distributed machine learning. Importantly, existing electronic health records based analytical tools can be leveraged without substantial costs implications for research networks to implement a Level 1 FDHN. Fewer regulatory barriers associated with a FDHN has a potential for diverse, non-network EDs to contribute to research, foster faculty development, and improve patient outcomes in emergency care.

Text:

Traditional centralized data collection limits collaboration

Successful multicenter research networks in Emergency Medicine directly address the ability to conduct high quality research on large numbers of eligible patients which is unlikely in single EDs, research gap identified by National Academies of Sciences, Engineering, and Medicine report on the Future of Emergency Care.¹ Research networks including but not limited to the Pediatric Emergency Care Applied Research Network (PECARN) and Strategies to Innovate Emergency Care Clinical Trials Network (SIREN) have generated high quality evidence, nurtured an investigator pipeline, and strengthened the overall emergency research infrastructure. This has directly contributed to improving care of ill and injured adults and children globally.² High performing emergency research networks often are supported by a centralized data center to ensure data quality, comply with institutional, national, and international regulatory standards, and curate data to conduct sophisticated analyses (Figure 1).

Sustaining centralized data centers is costly and requires substantial commitment of time and resources from researchers and specifically, health information technology experts. High costs emanate from the need to: 1) harmonize data from disparate and often highly customized electronic health record systems with varied data formats; 2) ensure compliance with cybersecurity protocols; 3) maintain patient privacy standards, especially when identifiable data is being sent outside the institution; and 4) conduct ongoing and required maintenance of databases. Furthermore, centralized data centers are often constrained by specific data collection protocols that limit conducting observational studies with up-to-date data, developing artificial intelligence models requiring many variables, and performing data exploration for future studies across network sites. Consequently, centralized data centers are not nimble, and the cost structure is a barrier for non-network sites to participate in network-based studies. This reduces the opportunity to solicit input into research study questions and designs, reduces the diversity of study populations, decreases generalizability, and limits dissemination and implementation of new evidence.

Moving towards a de-centralized approach

The COVID pandemic further underscored the ongoing need for nimble, multi-centered, multi-disciplinary, collaborative approaches to understand the rapidly evolving epidemiology of the virus, and assess the impact, efficacy, and effectiveness of various operational and therapeutic interventions. Our recent work on COVID's impact on diagnostic delays across 14 emergency departments (EDs) in

Michigan³ and a survey on the impact of COVID on provider burnout and innovation across 74 EDs in 28 countries⁴ has revealed that EDs are willing to commit site resources to contribute data for research in emergency care. To overcome the shortcomings of centralized data approaches, a distributed/de-centralized approach called federated data health networks (FDHN) has been successfully deployed in COVID research; this has opened the potential for similar approaches to emergency care network-based research.^{5,6}

A FDHN in emergency care is comprised of a series of decentralized, interconnected EDs where each site's data stays behind its institutional firewall but is structured according to a common data model (CDM) that allows data to be queried and/or analyzed without the data leaving the site.⁷ Thus, instead of data sharing and development of a central data repository, sites maintain a real-time dataset based on a common data model, which is a way of organizing data into a standard structure (Figure 2). This data is "visited" by centralized queries and algorithms.⁵ Guidelines for creation of a successful FDHN have been developed by the World Economic Forum and requires commitment of site resources, development of a robust data governance structure to address compliance with regulatory requirements, and collaboration with health information technology (HIT) personnel for data quality assurance.^{5,7}

Within the larger context of health informatics, there are existing common data models such as i2b2, PCORNet, FDA's Sentinel initiative, whose impact on improving outcomes continues to evolve.⁸⁻¹⁰ However, the single encounter-based frame of reference of the ED practice environment emphasizes the need for a simpler data model than the ones that currently used to map health systems data not relevant to emergency care delivery. It is possible for certain CDMs, such as the Observational Medical Outcomes Partnership (OMOP), whose architecture, vocabulary and accompanying analytic tools are specifically designed for federated analysis across various patient encounters (inpatient, ambulatory etc.) and across multiple institutions globally can be used for ED encounters.¹¹ However, this approach can be costly to implement, requires ongoing updating and data quality monitoring, and not tailored to the ED context.

How can we leverage the FDHN to optimize the performance of emergency care research networks?

Given the complicated nature of developing and assimilating an emergency research network based FDHN, we propose a stepwise, two-level development and deployment process. The first step is to create the FDHN itself by engaging established research networks or a consortium of EDs who are committed to its development. Each site will then obtain institutional review board approval and participate in the development of a common data model (an ED dataset based on

shared, accepted definitions of data variables) with the important caveat that the data will be retained at the site and queried in a standardized manner. Creation of this ED dataset will require the commitment of health information technology resources, clinical informatician(s), and clinical champions at each site. Basic descriptive and analytic statistics including development of regression and classification models can then be applied on each ED dataset in a standardized manner. Metanalyses of deidentified collated results can be used for hypothesis generation, health surveillance screening, development of dashboards for operational analytics, and quality improvement activities. We propose that a FDHN at this stage of development would be called a Level I FDHN: capable of performing basic descriptive analyses/models within each site and aggregating results centrally. At this stage, peer sites within the network have established a clear governance structure overseeing architecture, the common data model and vocabulary, access, and strategy utilizing primarily structured EHR data. Future states of FDHNs, which we call Level II, include automation of the data query or “visitation” process as well as inclusion of unstructured data such as text, waveforms and potentially blending a data lake strategy into the FHDN.

Once a successful Level I FDHN is established, research networks can consider a more resource-intensive Level II FDHN which would allow for sophisticated analyses such as federated machine learning on the ED datasets.¹² It is imperative to develop such a model as emergency care datasets become more complicated with the increasing availability of complex “omics” data (genome, metabolome, microbiome, transcriptome, etc.) which may need to be integrated with continuously obtained physiologic data from health monitoring devices (e.g., Fitbit, Apple Watch) into emergency care visits.¹³ Federated machine learning involves development of models locally and subsequently developing a global model by combining local model parameters in a Health Insurance Portability and Accountability Act (HIPAA) secure, cloud-based server.¹⁴ Level II FDHNs will need to follow the guidelines suggested but sites will need additional technological support/investment and regulatory oversight.

At any level, a successfully implemented FDHN could provide both timely and diverse data at scale across as many sites as willing to participate in a specific question. For instance, operational questions regarding number of left without been seen patients across participating sites during an epidemic stratified by race/ethnicity can give insights on health equity aspects of care along with syndromic surveillance. Other potential questions could range from disease, drug and medical device surveillance, to quality improvement initiatives, benchmarking, to research on rare diseases or outcomes, and even to facilitation of data collection in multicenter prospective clinical trials. Questions may range in complexity from as simple as counting the number of patients who left without being seen as a quality benchmark

to more complex analyses such as producing model weights in a federated machine learning model.¹⁵

Limitations of the federated approach

While FDHNs provide many advantages over a centralized model, there are several technical and administrative challenges that should be considered. First, from a technical perspective, a federated model places the entirety of the extract, transform and loading process within each site and thus, each site must be capable of providing the appropriate information technology support – both at inception and on an ongoing basis for maintenance and quality assurance of the dataset. Second, all participating sites within the FHDN must design, implement, maintain, and adhere to a common data model and common vocabulary which necessitates a carefully planned, potentially less flexible, data governance structure. This contrasts with centralized models where at least some of this work can be done and dictated by the coordinating center. Some of this “upfront cost” can potentially be defrayed by leveraging existing electronic health records (EHR) based analytical tools. More recently, the ongoing consolidation of EHR vendors and health systems as well as the need for ever larger datasets for research has driven many institutions to adopt either open source or proprietary common data models.⁷ Most EDs already commit to data gathering from the EHR for operational purposes or to share it as a part of their participation in federal and non-federal repositories such as the National Hospital Ambulatory Medical Care Survey, the Clinical Emergency Data Registry (CEDR), or collaborative quality groups such as the Michigan Emergency Department Improvement Collaborative (MEDIC).^{16,17} Finally, a third limitation is that a FHDN will never yield as granular data as a centralized model. By design, a limitation on the minimum level of aggregation or analysis is enforced by the FHDN and thus, requires that researchers or nodes requesting federated data access have carefully planned their analysis prior to querying individual sites. This also limits the ability to explore data across sites or repeatedly query sites for more data. Additionally, each site must be able to provide the technical resources such as data analysts and software to process incoming queries from other sites. Although this is a manual process in a level 1 framework, in future level 2 FHDNs, this work could potentially be automated with data “visitation” rules established at the network level.

How can FDHNs overcome these barriers to efficiently achieve the goals of existing emergency research networks that utilize centralized data centers?

First, having more sites committed to a common ED data model and standardizing the extraction and transformation process itself will improve data quality. Second,

both network and non-network sites that utilize the CDM will be able to contribute data in an efficient manner. Potentially, sites may also leverage the CDM for internal operational reporting, research, and other business intelligence purposes and some of each site's health information technology can be supported by extramural grant support. Third, participating in Level I FDHN will substantially lower the regulatory barrier to entry for many EDs to contribute analysis of their data by obviating the need for complex data use agreements, thus adding more diverse sites and populations into a network. This may especially be true for non-network sites that wish to participate on an ad-hoc basis in the FDHN. Network expansion allows a more rapid cultivation of a culture of research across a specialty, further expanding faculty development, recruitment, retention, and attraction of additional extramural funding. This empowers EDs that are not part of existing research networks to expand and strengthen their research mission and improve the care of emergency patients everywhere. Finally, beyond research, potential steps to advance FHDNs should include consensus building amongst stakeholders, our professional organizations and health systems regarding the optimal data architecture and vocabulary for emergency medicine. Establishment of a universal but limited CDM for emergency medicine could lower the barrier for many EDs to participate in multicenter research. Funding organizations and existing research networks should advocate for more limited, easier to adopt interoperability standards focused specifically on emergency medicine. In summary, despite some technical challenges and the need for investment by individual EDs, there is an unprecedented opportunity to leverage federated data health networks to exponentially enhance the impact of emergency care research networks.

References:

1. Institute of Medicine. 2007. Hospital-Based Emergency Care: At the Breaking Point. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11621>.
2. Papa L, Kuppermann N, Lamond K, Barsan WG, Camargo CA Jr, Ornato JP, Stiell IG, Talan DA. Structure and function of emergency care research networks: strengths, weaknesses, and challenges. *Acad Emerg Med*. 2009 Oct;16(10):995-1004.
3. Mangus C, Impact of COVID-19 on Associated Complications of High-risk Conditions in a Statewide Pediatric Emergency Network – JACEP accepted for publication
4. Mahajan P, Shu-Ling C, Gutierrez C, White E, Cher BAY, Freiheit E, Belle A; EMERGE NETWORK, Kaartinen J, Kumar VA, Middleton PM, Ng CJ, Osei-Kwame D, Roth D, Sinja TP, Galwankar S, Nypaver M, Kuppermann N, Ekelund U. A Global Survey of Emergency Department Responses to the COVID-19 Pandemic. *West J Emerg Med*. 2021 Aug 21;22(5):1037-1044. doi: 10.5811/westjem.2021.3.50358. PMID: 34546878; PMCID: PMC8463065.

5. Hallock H, Marshall SE, 't Hoen PAC, Nygård JF, Hoorne B, Fox C, Alagaratnam S. Federated Networks for Distributed Analysis of Health Data. *Front Public Health*. 2021 Sep 30;9:712569. doi: 10.3389/fpubh.2021.712569. PMID: 34660512; PMCID: PMC8514765.
6. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, Liu A, Costa AB, Wood BJ, Tsai CS, Wang CH, Hsu CN, Lee CK, Ruan P, Xu D, Wu D, Huang E, Kitamura FC, Lacey G, de Antônio Corradi GC, Nino G, Shin HH, Obinata H, Ren H, Crane JC, Tetreault J, Guan J, Garrett JW, Kaggie JD, Park JG, Dreyer K, Juluru K, Kersten K, Rockenbach MABC, Linguraru MG, Haider MA, AbdelMaseeh M, Rieke N, Damasceno PF, E Silva PMC, Wang P, Xu S, Kawano S, Sriswasdi S, Park SY, Grist TM, Buch V, Jantarabenjakul W, Wang W, Tak WY, Li X, Lin X, Kwon YJ, Quraini A, Feng A, Priest AN, Turkbey B, Glicksberg B, Bizzo B, Kim BS, Tor-Díez C, Lee CC, Hsu CJ, Lin C, Lai CL, Hess CP, Compas C, Bhatia D, Oermann EK, Leibovitz E, Sasaki H, Mori H, Yang I, Sohn JH, Murthy KNK, Fu LC, de Mendonça MRF, Fralick M, Kang MK, Adil M, Gangai N, Vateekul P, Elnajjar P, Hickman S, Majumdar S, McLeod SL, Reed S, Gräf S, Harmon S, Kodama T, Puthanakit T, Mazzulli T, de Lator VL, Rakvongthai Y, Lee YR, Wen Y, Gilbert FJ, Flores MG, Li Q. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med*. 2021 Oct;27(10):1735-1743. doi: 10.1038/s41591-021-01506-3. Epub 2021 Sep 15. PMID: 34526699; PMCID: PMC9157510.
7. Weeks, J., & Pardee, R. (2019). Learning to share health care data: A brief timeline of influential common data models and distributed health data networks in U.S. health care research. *EGEMS (Generating Evidence & Methods to improve patient outcomes)*, 7(1), 1-6. <https://doi.org/10.5334/egems.288>
8. i2b2 Informatics for Integrating Biology and the Bedside. (n.d.). Retrieved March 26, 2023, from <https://www.i2b2.org/index.html>
9. Fleurence, R. L., Curtis, L. H., Califf, R. M., Platt, R., Selby, J. V., Brown, J. S., ... & Gliklich, R. E. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, 21(4), 578-582. doi: 10.1136/amiajnl-2014-002747.
10. Sentinel Initiative. (n.d.). Retrieved March 26, 2023, from <https://www.sentinelinitiative.org/>
11. Observational Health Data Sciences and Informatics. (n.d.). Common Data Model. Retrieved March 26, 2023, from <https://ohdsi.github.io/CommonDataModel/>
12. World Economic Forum. *Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide*. Geneva (202). Available online at https://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf
13. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K. and Ourselin, S., 2020. The

future of digital health with federated learning. NPJ digital medicine, 3(1), pp.1-7.

14. Macias, C.G., Remy, K.E. and Barda, A.J., 2022. Utilizing big data from electronic health records in pediatric clinical care. Pediatric Research, pp.1-8.

15. Dayan, I., Roth, H.R., Zhong, A. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med 27, 1735–1743 (2021). <https://doi.org/10.1038/s41591-021-01506-3>

16. American College of Emergency Physicians. (n.d.). Clinical Emergency Data Registry. Retrieved March 26, 2023, from <https://www.acep.org/cedr/>

MedicQI. (n.d.). Retrieved March 26, 2023, from <https://medicqi.org/>

Figures:

Figure 1: Schematic of a Centralized Model

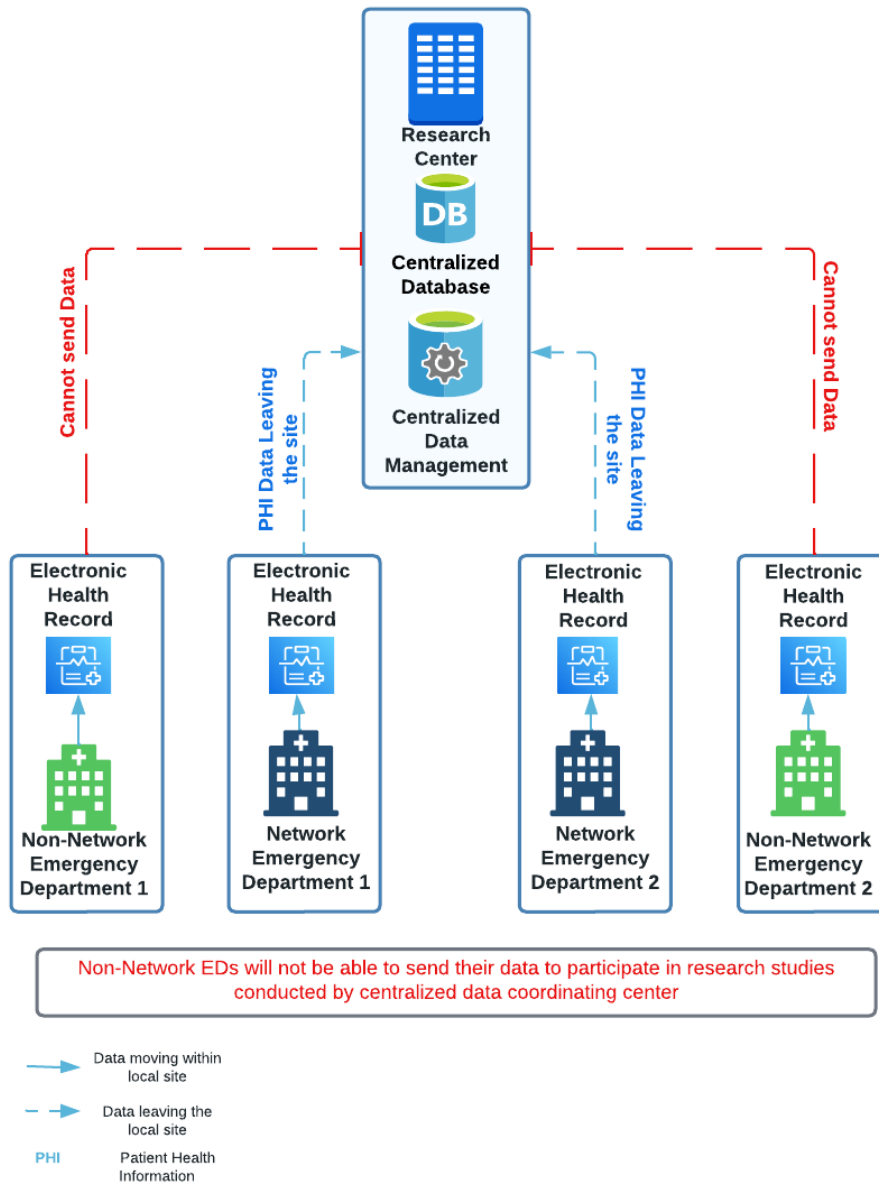


Figure 1 legend:

Diagram of centralized data collection and storage by coordinating center. Encounter level data including protected health information (PHI) leaves each participating institution and is stored by the coordinating center. Network emergency departments (EDs) can participate after completion of data use agreement (DUA). Non-network EDs generally cannot participate without completing DUA process.

Figure 2: Schematic of a Federated Model

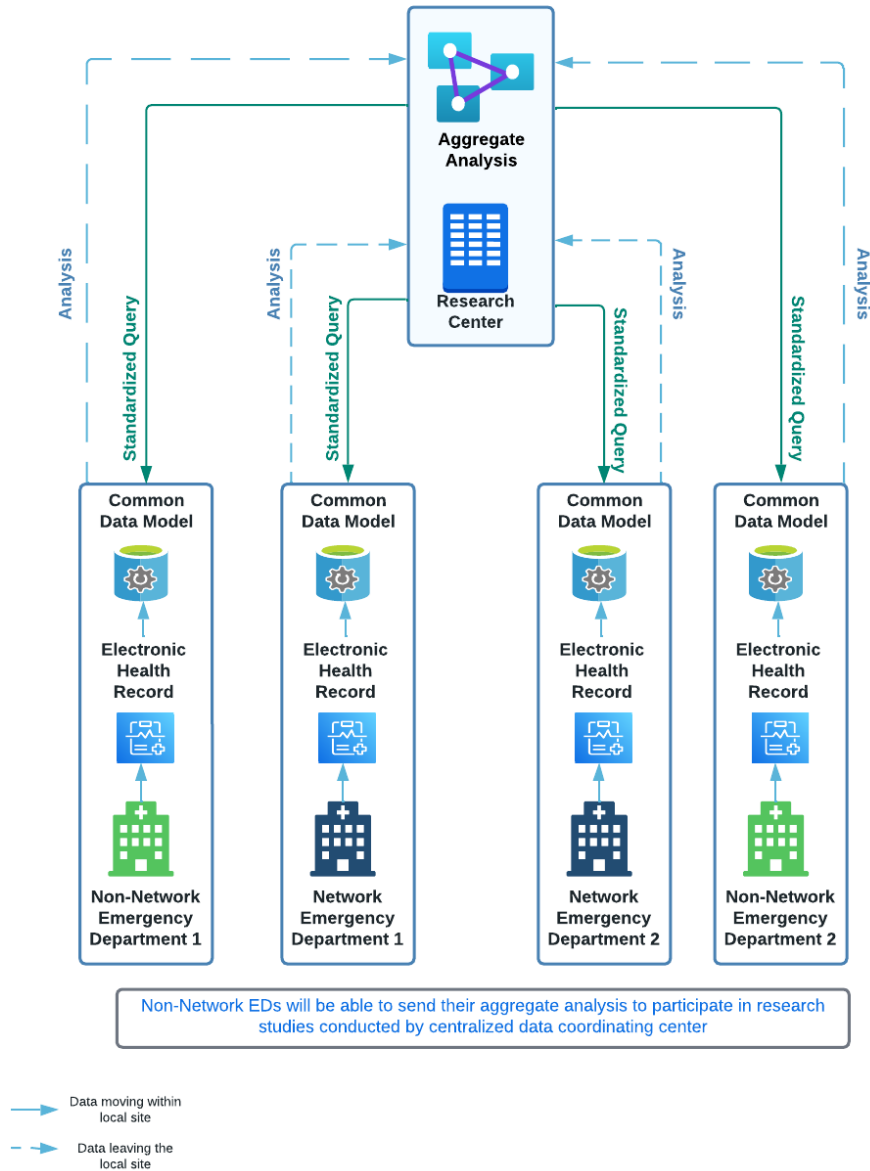


Figure 2 legend text:

Diagram of federated model with data extraction, transformation, and loading into a common data model (CDM) within the data warehouse of each ED. Standardized queries are distributed by the research center and run locally. Aggregate analysis is collected by the research center without the need for PHI to leave individual EDs and generally without the need for DUA. Non-network sites may participate if their data is mapped onto the CDM.