



The Office of the National Coordinator for
Health Information Technology

Accelerating Application Programming Interfaces for Scientific Discovery: Researcher Perspectives

PREPARED BY

Clinovations Government + Health for the
Office of the National Coordinator for Health Information Technology

Contract No.: HHSP233201600030I/75P00120F37004

March 2021



Table of Contents

Introduction	3
Information Technology to Support Research	4
Methodology	5
Findings	7
Current Challenges for Adoption and Use	7
Standardized API Data Elements are Often Insufficient for Research.....	7
APIs Should Streamline Data Mapping and Curation Activities	8
Researchers Seek Additional Education and Experience to Facilitate Adoption of APIs	9
Characteristics and Utility of EHR Data	10
Adoption of FHIR® Bulk Data Access APIs	13
Data Gaps and Opportunities	15
Social Determinants of Health Data	16
Real-World Data and Real-World Evidence.....	16
Privacy and Security Factors	18
Summary of Key Findings.....	21
Conclusion	23
References.....	24





Introduction

Today's ecosystem of health researchersⁱ seeks improved access to a variety of rich data sets to advance scientific discovery. These data sets originate from a wide range of healthcare settings via electronic health records (EHRs), health information technology (IT) systems, and other large repositories of clinical data such as enterprise data warehouses. In addition, researchers are gaining access to a rapidly burgeoning supply of genomic data, patient-generated health data (PGHD), social determinants of health data (SDoH), medical device data, and third-party consumer data. Researchers also depend on data “facilitators” such as data aggregators and integrators to provide access to real-world data (RWD) and solutions for data exchange and interoperability between disparate systems. Recently, new regulations have enabled greater access to health data through the use of application programming interfaces (APIs) by key healthcare stakeholders including consumers,ⁱⁱ providers, payers, and researchers. The categories of APIs and patient-facing, provider-facing, and other third-party health applications (apps) such as those used for population health and research will hereafter be referred to as “APIs and apps.”

The 21st Century Cures Act (Cures Act),¹ signed into law on December 13, 2016, was designed to help accelerate medical product development and provide consumers with new innovations and advances more quickly and efficiently. The Office of the National Coordinator for Health Information Technology (ONC) Cures Act Final Rule (Cures Act Final Rule),² released in March 2020, seeks to accelerate this shift to enable consumers to store, aggregate, use, and share electronic health information (EHI) using standardized APIs and apps of their choice by establishing access to EHI “without special effort” and the ability to exchange EHI with third-parties,ⁱⁱⁱ including researchers.

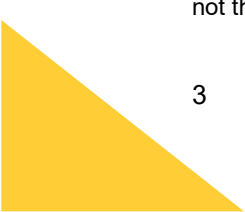
In addition, the ONC *National Health IT Priorities for Research: A Policy and Development Agenda*³ (Agenda), published in February 2020, highlights the need for 1) high quality data from health IT systems to enable discovery, and 2) a health IT infrastructure that can provide necessary functionality for research. The Agenda prioritizes effective engagement of consumers in research through the use of health IT and expanded use of APIs. Enabling the goals of the Cures Act Final Rule and the Agenda requires understanding the needs of relevant stakeholders; in particular, how researchers can benefit from the standards required by the Cures Act Final Rule.^{4,5}

This report summarizes common themes, findings, challenges, and barriers researchers face with respect to the use of standardized APIs and apps consistent with the Cures Act Final Rule and the Agenda. It also identifies data types required by researchers that are unlikely to be exchanged in the near-term by standardized APIs that define specific data elements, and thereby limit the data that is exchanged. This report does not investigate private APIs offered by health IT developers for established customers and partners that use proprietary and developer-specific mechanisms to access data.

ⁱ The term “researchers” includes academic scientists, clinical trials investigators, pharmaceutical and life sciences organizations, public health, and medical device manufacturers.

ⁱⁱ The term “consumers” includes patients and other users of personal health data.

ⁱⁱⁱ The term “third-parties” refers to app developers or users of health data (such as consumers, researchers) who are not the developers of the health IT solution that serves as the originating data source, such as an EHR.





INFORMATION TECHNOLOGY TO SUPPORT RESEARCH

While standardized APIs have the potential to alter the landscape of health data use for biomedical, informatics and health research, they must offer benefits over existing methods and mechanisms. Researchers often require access to extensive data sets tailored to the needs of each research study.^{6,7} Currently, academic medical centers and other provider organizations heavily engaged in research have developed sophisticated data warehouses that offer insights into data from not only EHR systems, but numerous health IT systems.⁸ Many organizations also integrate external data (e.g., claims, SDoH, patient-reported outcomes) into their data warehouses.^{9,10} Research teams at these organizations employ a staff with significant expertise in extracting data from the EHR or affiliated data warehouses to populate research study databases.¹¹ Often the resulting methods for data extraction are individually tailored for each study, are resource-intensive, and may incur additional costs.^{12,13,14}

Informatics and biomedical research teams utilize research databases populated from institution-specific steps to Extract, Transform, and Load (ETL) EHR and other data. ETL requires data partners or contributors to invest in costly technical resources with specialized skills in data models, terminologies, and programming.¹⁵ As reported by Priest et al., ETL processes are resource intensive and data quality assurance is an iterative process. The findings emphasize that limited resources, namely human resources, processing time, and server memory and space pose significant challenges for data contributors to participate in research data networks.¹⁶ The populated data vary depending on whether ETL is conducted from a copy of the full EHR database or if the health IT developer's data warehouse has completed basic ETL and normalization of EHR data within and across encounters.

Research studies that analyze data collected across provider organizations can utilize a common data model (CDM) to standardize the format and content of data at each collection source so that standardized applications, tools, and methods can be applied across different data sets. CDMs are based on a concept of transforming data within disparate databases into a common format (data model), and use a common representation (terminologies, vocabularies, coding schemes) to perform systematic analyses using a library of standard analytic routines written based on the common format.¹⁷ There is a variety of CDMs used for research and evidence generation; four CDMs used frequently by informatics and biomedical researchers are:

- Food and Drug Administration's (FDA) Sentinel Initiative's CDM,¹⁸
- Informatics for Integrating Biology and the Bedside (i2b2)'s CDM,¹⁹
- Observational Medical Outcomes Partnership (OMOP),¹⁷ and
- Patient-Centered Outcomes Research Network (PCORnet)'s CDM.²⁰

In practice, curating data to conform to CDMs can be time and resource intensive. Efforts to harmonize CDMs such as the Common Data Model Harmonization project²¹ have investigated the ability to reuse data and methods across research networks and leverage available health data standards. With the increasing availability of Health Level Seven International (HL7®) Fast Healthcare Interoperability Resources (FHIR®) standardized APIs from health IT developers and adoption of APIs by provider organizations, this project collected perspectives from key stakeholders in the research community to understand their current use of standardized APIs and identify opportunities and barriers to leverage APIs for research and reduce dependence on custom data extracts and complex data curation for informatics and biomedical research.





Methodology

To understand the needs of researchers, and the barriers and challenges they encounter using standardized APIs and apps, semi-structured discussions were conducted with researchers and key stakeholders from nine organizations: five academic institutions with health informatics expertise, one pharmaceutical company, one public health department, and two health solution-providers developing data products or research tools. The discussions sought perspectives on the usability of standardized APIs and apps for conducting or supporting clinical, informatics, and biomedical research, and how the standardized APIs compare to private APIs or other data extraction methods. In addition, the discussions explored the methods by which researchers access data within health IT systems (e.g., EHRs, laboratory information systems, clinical data warehouses). Discussion findings summarize data and information needs, current standards, gaps, and limitations of existing methods and tools. Discussion participants were asked to provide perspectives on:

- usability of standardized APIs and apps to meet the data needs of researchers;
- utility and value of EHR data from health IT systems for research;
- methods for obtaining robust and relevant health IT data for research;
- opportunities for accelerating consumer-directed sharing of health data for research; and
- examples of how FHIR® and FHIR Bulk Data Access APIs are used for research today.

Discussions were conducted from May 2020 to November 2020. A facilitator was joined by a note-taker who documented the discussion, organized the information collected, and assisted in the identification and analysis of key themes and findings using grounded theory techniques.²² The organizations and their respective discussion participants are listed in Table 1.

In addition to the discussions, the team reviewed industry feedback from professional organizations on how the FHIR standard could accelerate availability of clinical data for research via APIs. This feedback was compiled in response to a Request for Information (RFI)²³ from the National Institutes of Health (NIH) and provided insights into:

- the relevant research data that can be collected through the use of standardized APIs;
- how consumer-directed data exchange of EHI can accelerate the availability of and address gaps in research data; and
- the current and projected use of FHIR and FHIR Bulk Data Access API standards.





Table 1. Researcher Perspectives: Discussion Participants

Organization	Respondent Type	Discussion Participant
Cedars-Sinai	Academic	Spencer SooHoo, PhD
Datavant	Health Solution Provider	Jasmin Phua
Duke University	Academic	Erich Huang, MD, PhD
		Rachel Richesson, PhD, MPH
Georgia Tech	Academic	Mark Braunstein, MD
		Jon Duke, MD
HealthVerity	Health Solution Provider	Andrew Kress
North Carolina Department of Health and Human Services	Public Health Department	Jessie Tenenbaum, MD, PhD
Oregon Health & Science University	Academic	David Dorr, MD
		Benjamin Orwoll, MD
		William Hersh, MD
Pfizer	Pharmaceutical	Amy Cramer, MMCi, BSN, CPHQ
		Rob Goodwin, MS, MBA
Vanderbilt University	Academic	Paul Harris, PhD





Findings

CURRENT CHALLENGES FOR ADOPTION AND USE

When asked how standardized APIs, such as the FHIR® and FHIR Bulk Data Access APIs, are used to collect data for research purposes, discussion participants described experiences that varied widely, depending on the specific use case. They reported that currently, most researchers have limited knowledge of FHIR and FHIR Bulk Data Access APIs. Participants predicted that as requirements of the Cures Act Final Rule are implemented and researchers become more familiar with the use of standardized APIs, the use of standardized APIs in research will be to the benefit of the community. Specifically, discussion participants identified four motivators for adoption and use of standardized APIs:

- 1) To help streamline research efforts across organizations and health IT systems;
- 2) To reduce the need for time-consuming data scrubbing, mapping, and data curation;
- 3) To support multi-site trials and population level research; and
- 4) To enable data collection to be performed more seamlessly.

However, discussion participants also identified current challenges to greater adoption and use of APIs:

- **Standardized API data elements are often insufficient for research.** Standardized APIs successfully gather structured data elements found in common data models, but many are insufficient for research studies due to the limited data set available from EHR data;
- **APIs should streamline data mapping and curation activities.** Researchers spend considerable time and resources performing data cleaning, mapping and curation; use of standardized APIs must offer companion configuration tools to simplify and streamline these activities to demonstrate results; and
- **Researchers seek additional education and experience to facilitate adoption of APIs.** Researchers are reluctant to transition using APIs without more experience amongst their peers, known advantages, and end-user education.

Standardized API Data Elements are Often Insufficient for Research

The endless frontier of questions and hypotheses explored by research requires an ever-changing and expanding set of data to provide answers and insights. Information required is often unique to a particular research study and requires complex data mining and analysis methods. Standardized APIs successfully gather structured data elements found in common data models for clinical decision support, care coordination, and other use cases. These data elements are sufficient to provide general information, support care planning, and answer questions of population health but may not be as useful for specific areas of precision-medicine research biomedical research, or clinical trials. The nature of research studies is to seek answers to specific research questions that require specific, unique, or novel data sets or study-specific data manipulation and mapping. Standards-based APIs provide easy access to standard data sets best suited for common use cases and questions, rather than the novel perspectives often sought in research.





For clinical outcomes research, patient-level APIs are effective at gathering a well-defined subset of data elements for individual patients. However, discussion participants reported that current standardized APIs only extract patient-level data from certified EHR systems, not all health IT products. In order to supplement their data, researchers often use private APIs, HL7 interfaces, or other direct interfaces from non-EHR systems, rather than using standardized APIs.

Discussion participants said researchers often require text-based data or additional data sets not found in the EHR. While health IT developers may provide the entire clinical narrative note within the C-CDA²⁴ observation section, researchers find more benefit from parsing relevant sections from the note. In the future, several types of clinical notes will become increasingly available as part of the United States Core Data for Interoperability (USCDI).²⁵

Additionally, information contained within a given FHIR[®] resource varies based on implementation. Data are inconsistently entered into electronic health records even within the same organization, as well as across organizations. When commenting specifically on the use of the Apple Health app, one discussion participant noted, “General information on laboratory values and vital signs are available, but if one tries to drill down to access a procedure report, the FHIR resource is empty.” While health IT developer mappings from legacy data or HL7-based data structures to FHIR may vary, many EHRs allow customers to perform their own mappings and to accommodate custom data fields and workflows specific to the customer. Despite these limitations, all discussion participants supported using FHIR in research and agreed that the use of standardized APIs and apps using FHIR shows promise.

APIs Should Streamline Data Mapping and Curation Activities

Standardized APIs must add value for researchers beyond other data harmonization approaches, and reduce the burden associated with re-coding and mapping relevant data. Several discussion participants reported that, at the present, the same version of an EHR product may be implemented differently across institutions, leading to variations in mappings and data structure. Researchers must clean and curate the data before conducting analysis for research. Not all organizations have knowledgeable, experienced professionals with the skills to conduct mappings using common data models such as OMOP and i2b2’s CDM. For large-scale data analysis, many organizations are challenged with mapping legacy data from systems other than the EHR. When data are not mapped consistently across source systems, using standardized APIs to extract data from those systems may result in inconsistent or sometimes even empty data fields. Currently, the use of standardized APIs does not alleviate the burden on researchers to map or curate data; rather, they must be mapped and curated prior to data extraction using APIs in order to provide accurate data.

One discussion participant reported that consensus-driven models, where data analysts agree on the data elements to include in their mappings, are required to successfully implement mappings for a particular study. Although data element coding and mapping have improved, when investigating novel or new conditions or treatments, it is difficult to ensure that data mappings keep pace with what can be emerging definitions or understanding. Discussion participants said that, although the FHIR standards are flexible, FHIR does not always require specific terminologies, and information is difficult to collect from clinical notes and flowsheets. Additionally, a researcher may need only the most relevant piece of a narrative note, but more information than needed (or relevant) is often exchanged using the FHIR standardized APIs.





“
While there are standards, like C-CDA and FHIR to deliver the data, the challenge is that a lot of the interesting information about the patient is in a progress note or some type of free text report.”

One discussion participant said, “Researchers may want only an important narrative piece from a note, but the entire note (all sections and observations) is pasted in together within the HTML. So, the most relevant piece of a note cannot be predictably collected using FHIR®.” By overloading the FHIR resource with any narrative text within a note, the context and specificity that researchers seek from an encounter or progress note are not discernable or distinguishable from other text.

One discussion participant predicted that FHIR standards will become prevalent and ultimately replace other common data models and extraction tools currently used for research. Another discussion participant noted that FHIR

standard APIs offer more than other EHR data collection tools, and that APIs are more efficient than conventional queries or interfaces for cross-organization data collection. The excitement of using FHIR for research is “more about how to build these systems that go across networks, so we are going to be able to achieve some multi-site [study] where you write your program once and then you get access to the APIs at 10 different sites”.

“
FHIR’s contribution to making research accessible and useful may be greater than its contribution to doing the actual research...we don’t lack for technologies to go into a database and do whatever you want to do. What we do lack are technologies to make the results of research usable, useful, accessible, and interoperable.”

Researchers Seek Additional Education and Experience to Facilitate Adoption of APIs

While the use of standardized APIs shows promise for research, some discussion participants admitted to a general lack of knowledge about FHIR APIs and their capabilities related to research and offered that leveraging FHIR more broadly requires additional research community education. Discussion participants also noted that researchers such as epidemiologists and researchers performing biostatistical analysis will continue to feel more comfortable using traditional relational database models rather than utilizing new approaches.

In addition, health care systems often do not have knowledgeable, dedicated professionals to conduct clinical trials and use APIs and other technologies aimed at retrieving clinical trial study data. Pharmaceutical companies seeking to conduct clinical trials within hospitals may supplement their IT resources by applying their own personnel.

One discussion participant described their use of company resources to work directly with a provider organization’s health IT developer to ensure the appropriate data are mapped from the EHR and other health IT systems for specific clinical trials. Discussion participants recommended demonstration projects and pilot testing to evaluate the availability of research data from FHIR resources. Implementation guidance is needed for health IT and app developers to ensure FHIR resources are populated consistently across health IT systems (such as EHRs) to accelerate adoption in the ecosystem.





CHARACTERISTICS AND UTILITY OF EHR DATA

To understand the gaps in data collected by APIs for research, discussions were structured to characterize the current ways EHR data are repurposed for research. Discussion participants indicated that using EHR data for research is largely contingent upon how data are collected and entered into health IT systems. Variability of EHR implementations between organizations presents a challenge when applying common data models to extract and analyze EHR data.

Issues reported by discussion participants include:

1. Utility of EHR data is largely contingent upon tools and capabilities for mapping and extracting the data;
2. Mapping and coding of EHR data vary across organizations and health IT developers' products;
3. Data mapping can impact the integrity of the data and must be verified; and
4. Data requirements for research extend beyond information typically captured in EHRs.

“
 First, you have to map the data elements from [the] research project to the data points that are accessible via FHIR, which is a process because the end user doesn't understand FHIR or LOINC [Logical Observation Identifiers Names and Codes], which is where you get into that mapping process.
 ”

Discussion participants noted that not all health IT developers provide data extraction tools within their products or access to data within the EHR database or clinical data warehouse to all customers. Provider organizations may not purchase every EHR data extraction or report writing tool, as these solutions carry additional fees. Two different health systems with the same EHR may not expose the same data and metadata through the EHR's data warehouse or reporting solutions. While large research institutions may have the ability and resources to procure additional tools from health IT developers, less well-funded organizations do not.

Although health IT developer tools and their database products for research are improving, they are not yet sufficient to meet all researcher requirements. Health IT and informatics resources also vary across organizations. While well-funded organizations may have data analysts performing custom and complex data extraction routines, under-resourced organizations rely heavily on their health IT developer or contracted resources to extract available data to OMOP or other research data models for extraction.

“
 I don't think anybody could argue that it's more efficient within an individual organization to do things using FHIR. And so, we have to figure out how do we get to a place where it's as efficient, or not significantly less efficient, than just quickly using SQL [Structured Query Language].
 ”

Some discussion participants expressed the desire to use FHIR®-based APIs for extracting and converging the data, while others desired greater control and the ability to tailor the process of using different tools and methods.





Research organizations with considerable resources and tools typically embark on the following tasks in extracting data from EHRs and other health IT systems for research:

1. Define the data;
2. Determine which data model to use;
3. Write extract routines;
4. Define translations; and
5. Load data into research data warehouses.

Discussion participants described the disparity between different types of research organizations, and suggested that less well-resourced organizations must rely on turnkey solutions that provide data extraction in any format provided by the health IT developer. Participants recognized the value of standardized APIs may be to democratize data collection by providing more diverse organizations with open-source plug-and-play solutions. The development of the Research Electronic Data Capture (REDCap)²⁶ and other open source solutions by participating organizations aims to provide these tools. The realization that not all customers have in-house expertise has prompted enterprise health IT developers to develop and refine OMOP, i2b2, PCORnet®, Sentinel, and other CDM data extraction routines.





Initiatives Using Standardized APIs in Research

The following are several recent projects that are using standardized APIs in research, either with associated application development or large-scale data for researchers to access.

- eCare Plan Joint NIH/NIDDK AHRQ Project:** The Oregon Health & Science University (OHSU) is working on a proof-of-concept research study funded jointly through the Agency for Health Research and Quality (AHRQ), National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK), and the Assistant Secretary for Planning and Evaluation (ASPE).²⁷ The pilot eCare Plan tool developed for this project will be designed for use with patients who have chronic kidney disease (CKD), cardiovascular disease (CVD), diabetes, and/or pain with opioid use disorder (OUD). The data elements included in the USCDI show promise for supporting the project's requirements for gathering data on people with complex conditions to support clinical decisions for care planning. This project proposes using APIs to extract needed data from the EHR to assist in care planning activities and allows clinical staff to access relevant data from an app that launches directly from the EHR. Applications like these are at the forefront of how the data and API standards can fuel the research needed to impact people's lives in real clinical care.
- SUPER REDCap on FHIR®:** Several discussion participants referenced the REDCap data exchange platform, developed by Vanderbilt University.²⁶ The REDCap platform offers a secure web application for building and managing online surveys and databases and is designed to support data capture for research studies. The Weill Cornell Medicine's Information Technologies & Services department has developed the SUPER REDCap on FHIR project which uses FHIR-based APIs to access REDCap within Epic and automatically populate REDCap forms with data from Epic, eliminating copy-and-paste from Epic for research.²⁸
- Cumulus:** Boston Children's Hospital, in collaboration with Yale University and Yale-New Haven Health, aims to develop a platform based on HL7 FHIR that leverages bulk data to support an ecosystem for research and learning, called Cumulus.^{29,30} Cumulus will build from existing standards and open health IT tools to offer turnkey functions that support rapid learning within a healthcare system. Tools to be developed and tested will allow users to annotate FHIR data for analytics, de-identify data, and query cohorts. This project addresses the Leading Edge Acceleration Projects (LEAP) in Health IT fiscal year 2020 special emphasis notice area of interest 2: Cutting Edge Health IT Tools for Scaling Health Research.

“Some of the things that are going on are just fascinating, and I think they're very pertinent to the question of not just how you use APIs to do research, but how you use APIs to make research clinically useful to practitioners as well as to patients.”





- FHIR® Factories:** MedStar Health Research Institute (MHRI), in collaboration with Georgetown University's Innovation Center for Biomedical Informatics (ICBI), the American College of Emergency Physicians (ACEP), HealthLab, and Asymmetrik, aims to better understand the current state of open source, health IT tools.^{30,31} The project will develop and evaluate open health information technology (health IT) tools based on the HL7 FHIR standard. These tools will be developed and tested to support data acquisition, data transformation, and advanced analytics. This project aims to better understand the current state of open-source health IT tools through a rigorous evaluation, including an environmental scan, stakeholder interviews with key participants, and usability evaluations. Using this information, MHRI will identify and prioritize critical needs for future open-source health IT tools. This project addresses the Leading Edge Acceleration Projects (LEAP) in Health IT fiscal year 2020 special area of interest 2: Cutting Edge Health IT Tools for Scaling Health Research.

ADOPTION OF FHIR® BULK DATA ACCESS APIS

The Cures Act Final Rule requires that the FHIR Bulk Data Access API, which provides access to patient-level data across a patient population, be enabled by certified health IT developers to support population-level health research. It is anticipated that the FHIR Bulk Data Access API can support many use cases across the healthcare ecosystem, including:

- Integration of an internal clinical system with an EHR;
- Bio-surveillance, syndromic surveillance, and disease reporting;
- Post-marketing surveillance of therapeutics and devices;
- Combining claims and electronic health record data to calculate quality measures;
- Building data sets to develop and tune machine-learning algorithms; and
- Federated data sharing networks for multi-institutional research.

The FHIR Bulk Data Access API⁵ is a standard to enable easy access to population health level data. Accountable care organizations, researchers, or public health authorities should be able to, without special effort, extract a cohort of patient records from an EHR for further analysis.

Population-level research requires extracting large amounts of data from across health IT systems and most health IT developers and their customers are not yet using the FHIR Bulk Data Access API to perform this function. Unfortunately, compiling large-scale data sets using individual patient data acquired through patient-level APIs would be incredibly time-consuming and inefficient for researchers. Therefore, standardized APIs for working with population-level data, often referred to as “bulk data”, are needed to support this system.

Today, health systems perform bulk data export through proprietary pipelines. This requires teams of IT professionals to perform manual mapping to and from delimited text formats, such as comma separated values, for each reporting and analytics pipeline, of which there are many. These manual processes are extremely expensive, time-consuming, and not sustainable.





The original FHIR® and SMART³² APIs work well for accessing small amounts of data, but large exports perform poorly because they require hundreds of thousands of individual API requests. The FHIR Bulk Data Access API was rapidly defined, standardized, and piloted to address bulk data use cases, and feedback from a wide range of early adopters across the health industry is being incorporated into the standard to clarify and iterate guidance materials.

“The FHIR Bulk Data Access API is not the complete answer to putting FHIR data in a form that’s meaningful to analysis like OMOP or a SQL database. Other things are going to emerge- Bulk FHIR may in fact be a transitional technology.”

At the time discussions were conducted for this report, the FHIR Bulk Data Access API was not yet required of health IT developers or providers. According to SMART Health IT, known implementations³³ of the FHIR Bulk Data Access API are included in the examples shown in Table 2 (as of December 2020):

Table 2. FHIR® Bulk Data Access API Implementations³³

Organization	FHIR Bulk Data Access API
1upHealth	Structured Query Language ³⁴ (SQL) Analytics API https://1up.health/dev/fhir-analytics
Aetna	(in development)
Allscripts	(in development)
CMS	Data at the Point of Care https://dpc.cms.gov/ Beneficiary Claims Data API https://bcda.cms.gov/
Cerner	(in development)
EPIC	(in development)
Fire.ly	Vonk FHIR Server https://vonk.fire.ly/
Google	Google Cloud Healthcare API https://cloud.google.com/blog/topics/healthcare-life-sciences/getting-to-know-the-google-cloud-healthcare-api-part-1
Humana	(in development)
IBM	Client https://github.com/IBM/FHIR Server https://ibm.github.io/FHIR/guides/FHIRBulkOperations/ IBM Watson Health (bulk data support in development)
MEDITECH	(in development)
Microsoft	Azure API for FHIR https://github.com/microsoft/fhir-server-samples
ONC	Inferno Community https://inferno.healthit.gov/community and https://github.com/onc-healthit/inferno Inferno Program https://inferno.healthit.gov/inferno and https://github.com/onc-healthit/inferno-program (tied to 170.315(g)(10) Multi-Patient Query)
Oracle	(in development)
SMILE CDR	Bulk FHIR Server https://smilecdr.com/docs/bulk/fhir_bulk_export.html
SMART Health IT	Client https://github.com/smart-on-fhir/sample-apps-stu3/tree/master/fhir-downloader Server https://bulk-data.smarthealthit.org/





Some discussion participants are still working to determine if the FHIR® Bulk Data Access API will work with their health IT systems. These participants recognized the promise and potential benefits of this API for extraction of large amounts of data for research studies. However, implementation of the FHIR Bulk Data Access API with their supported health IT systems presents multiple challenges, including health IT developers’ readiness to support the standard and their customers’ willingness to implement the standard.

DATA GAPS AND OPPORTUNITIES

While standardized APIs may provide useful tools to extract certain data elements from the EHR, as noted by several discussion participants, researchers often seek to enrich the data available to them to make it more meaningful for research. They also want to access the metadata at a specific point in time which is not available in current FHIR-based APIs. Researchers reported often searching outside the EHR to find the information required. One discussion participant said, “EHRs may only contain five to ten percent of the data needed for research.” As a result, researchers must already use a range of techniques to gather various data types in each of the common data models, which reduces the perceived benefit of using standardized APIs which take specialized knowledge to use and collect a relatively small amount of data. Researchers accustomed to extracting data through in-house and established methods may find it easier to use existing approaches to meet their data needs.

Discussion participants from academic organizations said that, for the majority of their research projects, they also need to extract data from sources outside the enterprise EHR. Academic researchers reported that the C-CDA or USCDI may come from the EHR system, but data needs for even basic cohort studies require additional data that may be missing from EHRs, as well as additional data elements and metadata not found in EHRs. Sources to meet these data needs may not be the EHR, and the fact that needed data often reside in multiple systems highlights the need for different extraction tools and researchers’ continued reliance on data warehouses.

“ FHIR’s contribution to making research accessible and useful may be greater than its contribution to doing the actual research. We don’t lack for technologies to go into a database and do whatever you want to do. What we do lack is technologies to make the results of research usable, useful, accessible, and interoperable. ”

Discussion participants emphasized the need for additional data relevant for research studies not available within EHRs and are contained in other systems, such as:

- Financial data;
- Payer information;
- State disease registry;
- Progress notes;
- Metadata; and
- Ancillary clinical system data.

Often, additional data elements must be gathered from a native laboratory, imaging, and other ancillary systems. Metadata needed to track the sources (e.g., user, source system, other provenance data) and





timestamps associated with specific code sets, such as Logical Observation Identifiers Names and Codes (LOINC)³⁵ values, can be used to support research findings, but data mapping often presents challenges for researchers. Discussion participants also said that clinical notes presented in the summary section of a patient’s record cannot be translated easily to provide valuable data, other than through manual abstraction that adds to the difficulty of gathering useful data.

Research efforts in precision medicine, genomics, and rare diseases present use cases that demand additional, detailed, difficult-to-locate information. For example, oncology researchers would like to access information found in nursing and infusion flowsheets. However, flowsheets have not been a high priority for standards development efforts or for structured or coded data extraction by health IT developers. Precision medicine can also be highly domain-specific, making standardized electronic capture of needed data difficult.

Social Determinants of Health Data

Another opportunity for future uses of standardized APIs may be in the collection and use of SDoH data.³⁶ Discussion participants reported that most EHRs do not adequately support external (non EHR) data, such as SDoH data, which is increasingly being used in research. SDoH data are often difficult to incorporate into the EHR and are typically added from data sources outside the organization and obtained through purchases of third-party data. For example, consumer behavior data generated from socio-economic, demographic, and other data merged with large consumer data sets (e.g., credit score data, magazine subscriptions, or grocery store purchases) for specific zip codes, neighborhoods, and block groups offer insights into demographic and environmental factors for disease risk. These data are available for sale or analysis by health IT developers offering data aggregation and analytics services. Data aggregators among the discussion participants indicated that the lack of interoperability between data types is a key market challenge they meet. Researchers struggle to collect data outside core clinical systems or to integrate large sets of de-identified data to enrich commonly available data sets. Using a data integrator or a data aggregator, enables researchers to use de-identified data at high specificity and resolution, allowing them to combine their own data with additional data, such as ambulatory, specialty pharmacy, laboratory, grocery, claims, and other consumer data sets. These data integrators are building and offering their own APIs to enable access to aggregated and integrated data sets.

Real-World Data and Real-World Evidence

Real-world data (RWD)³⁷ are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. RWD originates from a variety of sources, including but not limited to:

- Electronic health records (EHRs);
- Claims and billing activities;
- Patient surveys;
- Product and disease registries;
- Patient-generated data, including in-home use settings; and
- Data gathered from other sources that can inform on health status, such as mobile devices.





Real-world evidence (RWE)³⁷ is the clinical evidence regarding the use and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses including, but not limited to, randomized trials. The use of RWD and RWE is becoming more prevalent as researchers seek to enrich their data sets and conduct high throughput research. While it is possible to extract RWD from EHR data using standardized APIs, gathering RWD from most other data sources is not possible using current standards. EHR data are insufficient to answer the questions asked by researchers conducting complex analyses trying to leverage clinical data with other RWD to develop RWE.

“ We see a lot of people creating less robust data sets that can only be used in isolation. You cannot build upon it and enrich it. We see a lot of interest in enriching data sets, but we have not seen life science customers do that. ”

One data aggregator among the discussion participants said that “the world is becoming more of a Venn diagram,” in that data sets are overlapping and the expanded use of multiple data sets used in conjunction for large research studies can produce a way to link and match patient data without having to identify patients. The research community wants to consider RWD for hypotheses testing, product efficacy analyses, and study control arms. Discussion participants reported that acceptance of this emerging practice is accelerating, but must overcome resistance to trusting data sources coming from RWD rather than real-world research studies. That trust depends on the accurate mapping of RWD contained in EHRs.

Another data aggregator described the use of RWD that exists in small data sets, which provides for more rapid, yet less robust research. These data sets do not require extensive data curation and mapping. Conversely, there are larger data sets, such as data collected during clinical trials, that are robust but time-consuming to curate and obtain. Often, such data sets cannot be merged. Some data aggregators have developed the infrastructure and open, scalable technologies to host the data and link to any additional RWD through a robust governance structure that controls access to data sets. Core RWD sets can then be joined in a privacy-preserving manner, enabling rapid research studies at scale. The rapidly evolving and expanding RWD and RWE market is blurring the lines between health IT and data products that serve traditional grant-funded academic research and industry-led activities (e.g., clinical trials, pharmaceutical and medical device funded research).

Discussion participants talked about this new era of research, where the use of clinical data and RWD increasingly enables researchers to enrich their data with de-identified data sets that they purchase. In addition, data analytics and aggregator solution providers are able to remove identifying data elements from data sets purchased from providers and other health care organizations, such as payers, laboratories, and pharmacies. The de-identified data can be combined with RWD data sets, then cross-referenced and matched to identify common patients who can be re-identified at a later date with permission if necessary. Such patient matching and linkage is a process that has become a business model and driver for integrators, because health care organizations cannot extract the data without investing in high-cost technologies. Data aggregators provide a critical function by taking data from various sources, then matching and linking data while honoring data use agreements (DUAs), identifying redundant data, managing updates to data and metadata, and working with varying data formats. The benefits of





standardized APIs may be less relevant to researchers who have established methods to obtain EHR data augmented by additional useful data, pre-cleaned and aggregated by integrators or data aggregators.

Health systems are serving as data contributors to research by supplying de-identified data sets. Two discussion participants described a growing economy from selling and purchasing large data sets by and on behalf of health care organizations.

In the future, the use of standardized APIs such as FHIR® and FHIR Bulk Data Access APIs could improve the ability to collect and disseminate RWD for research purposes, and make such data accessible to research organizations that may not have the resources to acquire it from data aggregators and integrators. Currently, the Vulcan HL7 Accelerator³⁸ is seeking to develop and refine FHIR standards to standardize the submission of RWD to the FDA for review.

“

We don't have the mechanisms to identify and stop research related data that is in our health IT systems from going out through APIs. We don't have all of the controls that we need for sensitive data, and that's one thing that makes people a little nervous.

”

PRIVACY AND SECURITY FACTORS

Privacy and security concerns remain a barrier when using health care data for research purposes. Similar to findings reported in a related ONC report focused on consumer perspectives, *Accelerating Application Programming Interfaces for Scientific Discovery: Consumer Perspectives*,³⁹ discussion participants felt that consumers are unaware of the extent to which their health data can be “de-identified”, sold, and otherwise reused without their consent. Participants added that current consent and user agreement methods are insufficient to adequately inform consumers of the responsibility they assume for their health data’s privacy and security after it is electronically transferred from an entity covered under Health Insurance Portability and Accountability Act (HIPAA)⁴⁰ to their phone or another device.

More specifically, privacy and security concerns continue to be a barrier to patient-directed sharing of health data via APIs and apps. Patients and health care organizations want to trust those with whom they share data. Comments by discussion participants emphasized that patients often do not understand what controls they have over their health data. In addition, researchers may not be able to limit the amount of data obtained through FHIR-based APIs, or withhold certain sensitive data, particularly when a study may only require a subset of data retrieved through the API. Discussion participants felt that patients should be able to determine if they want to be anonymous when contributing their data for a particular research study or if they want to remove consent when their data are to be used for subsequent studies, in a de-identified manner.

Privacy has become increasingly complex in the rapidly changing and expanding ecosystem of data analytics and large data sets available for purchase. Whether data reside in a provider’s EHR, consists of PGHD collected to address a health concern,⁴¹ or collected for a research study, data privacy is paramount to establishing trust. Several discussion participants noted they are struggling in their own organizations with privacy concerns and how to keep data safe as they receive requests for data from an increasing number of external parties. The participants reported that there are insufficient mechanisms to segment unnecessary clinical data from the data elements needed for research residing in the same EHR, and they





need more controls over sensitive data. While current standard FHIR authentication protocols, such as OAuth 2.0,⁴² effectively support API access to authorized apps and authorized recipients, restricting access to sensitive data can be challenging through standardized APIs that exchanged defined data sets. Discussion participants shared concerns about being able to “tag” certain data elements that should not be released through the API and produce enough granularity^{iv} to provide access only to data needed for the particular research purpose.

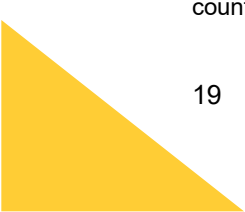
When researchers want to obtain clinical data for their studies, they are required to submit a research study protocol to an Institutional Review Board (IRB)⁴³ to be granted access to individually identifiable data sets. Some data aggregator solution providers provide the infrastructure to enable interoperability between systems, thereby reducing the barriers for researchers to gain access to larger data sets. Others have developed software tools that de-identify health data and linking them to a “patient key”. Data aggregators may sell data back to health care organizations to use for research and analysis. In this scenario, the data aggregators and integrators are not acting as “data holders”, “data participants” or “data stewards” directly. They are acting as Business Associates to covered entities under HIPAA, tasked with de-identifying data sets to enable the use of clinical data for research.

The methods for de-identification of protected health information (PHI) remains a somewhat controversial topic among researchers and data analysts, due to the relative ease by which data may be re-identified.^{44,45,46,47} As one discussion participant commented, there are “18 data elements” specified in HIPAA Privacy Rule⁴⁰ which must be removed in order to consider the data “de-identified”⁴⁴, yet genomic data is not specifically included in that list. However, genomic data are considered PHI and therefore covered under HIPAA,⁴⁸ which creates some confusion for researchers and others. Other data aggregators do play a more active role in the extraction and curation of data through DUAs and can scrub the data to strip it of identifiable data elements as identified under the HIPAA Privacy Rule. Data aggregator discussion participants acknowledged that the landscape of balancing patient privacy with the many uses for de-identified patient data can be difficult.

As described in the ONC’s report, *Key Privacy and Security Considerations for Healthcare Application Programming Interfaces (APIs)*,⁴⁹ APIs and apps are not inherently any less secure than other health IT systems. As long as healthcare APIs are implemented with appropriate privacy and security safeguards in place, they can add value to consumer-directed data sharing. Discussion participants noted that security safeguards and controls must be in place during the collection, storage, and transmission of data. In addition to safeguards -- including encryption, network security, endpoint security, access controls, and authentication -- assigning permissions to access patient identifiable data becomes even more complex when there are two or more data contributors and users. Assigning permissions is determined through a governance model and data use agreements that detail exactly who has a right to access which parts of the data and in what way the data can be used.

One discussion participant reported that there is not a “universal way to contract the ‘permissioning’ between parties” and they are “contractually reliant on the data contributor to state that they have rights to

^{iv} Granularity of data refers to the extent that data fields can be subdivided, or how detailed a single field is, such as how a location field can capture different levels of specificity, such as hospital room, address, zip code, state, or country.



the data they are contributing.” Based on their experience, data aggregators gain access to the individual patient data in two ways: 1) individuals grant access through individual consent or 2) the contributing party grants access through a DUA or Business Associate Agreement. The FHIR[®] standard API use granular “permissions” through the authorization process to determine which data are accessible. According to the SMART App Authorization Guide,⁵⁰ the most granular access scope that the guide currently permits is a FHIR[®] resource. For API developers enabling fine-grained access, the SMART App Authorization Guide provides examples that can be used to define specific OAuth 2.0 access scopes to indicate which FHIR resource types an individual or authorized user is allowed to access when requesting individual health data from an EHR system. Without these security measures, using or sharing data for research purposes can introduce liability to data contributors and data users.



Summary of Key Findings

Discussion participants provided insight into the current landscape and future promise of using standardized APIs and apps for research. Although most discussion participants reported excitement over the promise of FHIR® and FHIR Bulk Data Access APIs for research, challenges and gaps remain to accelerate researcher adoption of APIs. Once resolved, researchers will be able to repurpose their time, resources, and focus beyond the now streamlined collection of basic data elements. With easy-to-access architecture in place, the barrier to entry for research institutions previously excluded due to the resource intensity of collecting these basic elements will be lowered.

These challenges, segmented by key themes in this report, are described in Table 3.

Table 3. Challenges Identified by Discussion Participants

CHALLENGES IDENTIFIED BY DISCUSSION PARTICIPANTS	
Current Challenges for Adoption and Use	
<ul style="list-style-type: none"> Standardized APIs offer defined or limited data sets that researchers may need to supplement with additional data to support their investigation. Expanded adoption, demonstration studies, and dissemination of API used by researchers are needed to document advantages and value to facilitate clinical and informatics research. Capabilities of FHIR APIs are not widely known across the research community. 	<p><i>“the researchers...most of them don’t really understand FHIR, but they hear that this will make a lot of their life easier. That they’ll be able to run multi-site trials and do electronic data collection more seamlessly and won’t have to hire all these manual annotators to extract the data and bring it out.”</i></p> <p><i>“While the APIs work, our systems in research are maybe not as sophisticated as in the clinical care setting. So, while we can exchange data using APIs and we have programs to support that on the clinical side of the house, we don’t have that same kind of support on the research side of the house.”</i></p> <p><i>“It’s not really truly standard [data exchange across health IT systems], it’s ‘standard-ish’ which takes away from the whole concept of a standard.”</i></p>
Characteristics and Utility of EHR Data	
<ul style="list-style-type: none"> Data collection and documentation within and across EHRs varies significantly, creating challenges in retrieving data for research analysis. APIs retrieve data as-is or with very little manipulation. There is little recourse if mappings are incorrect or the wrong data is in a given field. Researchers using FHIR-based APIs must often supplement their data by using HL7 or other direct interfaces, diminishing the value and motivation to use standards-based APIs. 	<p><i>“EHR data varies and can be incomplete. Even in a given health system or practice, the variability of how that data is entered [into the EHR] is huge.”</i></p> <p><i>“One of the challenges of any standard is that it is only as good as the source data that is fed into by the contributing party.”</i></p> <p><i>“At some point in time, clinical care data and clinical research data stopped talking in the same language and that includes standards. There’s also just differences in the definitions and the data elements that we collect.”</i></p> <p><i>“The information that’s collected in health systems is probably also going to have to go through a transformation. So that will have to be a reconciliation of the information that’s collected for health purposes, and that will be needed for research if it’s going to come through those same sorts of standards and APIs then a build out of research forms in all of the systems that would be used for research...”</i></p>





CHALLENGES IDENTIFIED BY DISCUSSION PARTICIPANTS

Adoption of FHIR® Bulk Data Access APIs

- | | |
|---|---|
| <ul style="list-style-type: none"> ● Many health IT developers are still building and have not yet released their FHIR Bulk Data Access API. ● While there are emerging use cases for the FHIR Bulk Data Access API, there are limited real-world examples. | <p><i>“We don’t have Bulk FHIR turned on here. There’s no easy way to do it. I try and mimic those FHIR or Bulk FHIR calls into [our EHR] using standard queries, database queries into our enterprise data warehouse.”</i></p> <p><i>“Bulk FHIR is still not that widely used yet as far as I know.”</i></p> |
|---|---|

Data Gaps and Opportunities

- | | |
|--|--|
| <ul style="list-style-type: none"> ● Current APIs do not provide the robust data needed to populate common data models used in research. ● FHIR-based data has not included narrative data, will when the new USCDI standard is implemented. ● Researchers often have the need to enrich their EHR data with data from other sources outside the EHR. These data sources (e.g., SDoH, genomic) are not currently available using FHIR APIs. | <p><i>“While there are standards, like C-CDA and FHIR to deliver data, the challenge is that a lot of the interesting information about the patient is in a progress note or some type of free text report [that is not available currently].”</i></p> |
|--|--|

Privacy and Security Factors

- | | |
|---|--|
| <ul style="list-style-type: none"> ● Privacy and security issues remain a concern to ensure trust in how data are stored and used. ● Anonymizing data remains a controversial topic among researchers and data analysts regarding whether and how data can be de-identified. ● Informed consent language in health apps may not clearly state how consumers’ data will be used. ● Researchers are not able to limit data collected via standard APIs to only the most granular data needed for the investigation. | <p><i>“I actually do think that for the highly privacy minded among us, there should almost always be an opt-out option. If you’re very concerned about privacy you should be able to say, ‘no, I want to be off the grid, don’t want it.’”</i></p> <p><i>“Stratified [granular] consent...I know there are a few different groups working on the consent ontologies, so the ability to specify, ‘I give consent for all of my data and all of my samples to be used for any research’, or ‘only my blood or only my data but not my biospecimens’ or ‘only research in this disease, but not any others’ or ‘only in cancers, but not other types’, and enable people to be able to control their data.”</i></p> <p><i>“From a privacy standpoint, we provide them with technology that removes all the source data but does so in a way we can still track a patient privately over time...We connect data about patients to various endpoints on a de-identified or identifiably with consent basis.”</i></p> |
|---|--|





Conclusion

The discussions explored the utility and prevalence of using standardized APIs, such as the FHIR® and FHIR Bulk Data Access APIs, for research to enable scientific discovery, and gathered perspectives from stakeholder representatives in the research community actively working on various initiatives in biomedical and clinical informatics, clinical trials, and other types of health research.

The findings establish a baseline of activity and current successes using APIs for research and identified the remaining challenges and opportunities for making standardized APIs and apps work for researchers. The findings also highlight various examples of emerging APIs and apps used in the research community and for novel use cases.

Limitations on the information gathered from stakeholder interviews included the small sample size and limited number and type of stakeholder groups represented across the discussion participants. Discussion participants from academic research institutions included nationally known experts actively engaged in research using APIs. The open-ended nature of the discussions facilitated gathering valuable individual insights that may or may not be able to be aggregated and synthesized across stakeholder groups. Discussions identified potential areas for ONC to consider for future funding of studies, pilots, or policy development. Discussion participants varied in their insights into the broader goals of ONC and national priorities for research. Their viewpoints tended towards narrow focus on how their organizations use health IT data and APIs.

The priorities and needs that emerged support the strategies that ONC established in the Agenda and considerations in realizing the vision of the Cures Act Final Rule. In particular, participants underscored:

- The need for more standardization and better tools to conduct the mapping, configuration, and ETL for research data across different systems and organizations.
- The desire to leverage standardized APIs, such as the FHIR Bulk Data Access API, for clinical and biomedical informatics research, especially large-scale studies using data sets from multiple sites and data outside the EHR.
- The value of more education regarding the use of FHIR and FHIR Bulk Data Access API in the research community, so researchers may leverage the power of the APIs.
- The lack of current examples of the FHIR Bulk Data Access API currently in production for research, but development efforts continue to show promise.
- The concern for privacy and security measures to enable further adoption of APIs and apps.

While this report details findings geared towards health research, the results can support efforts aimed at accelerating the use of standardized APIs and apps for other important use cases in public health, disease registries, pharmacovigilance, and biosurveillance.





References

- 1 Federal Register: 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program [Internet]. Federalregister.gov. 2021 [cited 1 January 2021]. Available from: <https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification>
- 2 ONC Cures Act Final Rule [Internet]. Healthit.gov. 2020 [cited 1 May 2020]. Available from: <https://www.healthit.gov/curesrule/>
- 3 National Health IT Priorities for Research: A Policy and Development Agenda [Internet]. Healthit.gov. 2020 [cited 1 May 2020]. Available from: <https://www.healthit.gov/topic/scientific-initiatives/national-health-it-priorities-research-policy-and-development-agenda>
- 4 Health Level Seven International (HL7®) Fast Healthcare Interoperable Resources (FHIR®) v4.0.1 [Internet]. HL7.org. 2020 [cited 1 December 2020]. Available from: <https://hl7.org/fhir/R4/>
- 5 Health Level Seven International (HL7®) Fast Healthcare Interoperable Resources (FHIR®) Bulk Data Access (Flat FHIR) [Internet]. HL7.org. 2020 [cited 1 December 2020]. Available from: <https://hl7.org/fhir/uv/bulkdata/>
- 6 Coorevits, P., Sundgren, M., Klein, G., Bahr, A., Claerhout, B., Daniel, C., Dugas, M., Dupont, D., Schmidt, A., Singleton, P., De Moor, G. and Kalra, D., 2013. Electronic Health Records: New Opportunities for Clinical Research. *Journal of Internal Medicine*, 274(6), pp.547-560.
- 7 Cirillo, D. and Valencia, A., 2019. Big Data Analytics for Personalized Medicine. *Current Opinion in Biotechnology*, 58, pp.161-167.
- 8 Lyman, J., Scully, K. and Harrison, J., 2008. The Development of Health Care Data Warehouses to Support Data Mining. *Clinics in Laboratory Medicine*, 28(1), pp.55-71.
- 9 Ozaydin, B., Zengul, F., Oner, N., Feldman, SS. Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research. *J Med Internet Res*, 2020;22(6):e18579.
- 10 The Feasibility of Using Electronic Health Records (EHRs) and Other Electronic Health Data for Research on Small Populations. [Internet]. Hhs.gov. 2021 [cited 26 February 2021]. Available from: https://aspe.hhs.gov/system/files/pdf/107231/rpt_ehealthdata.pdf
- 11 Jensen, P., Jensen, L. and Brunak, S., 2012. Mining Electronic Health Records: Towards Better Research Applications and Clinical Care. *Nature Reviews Genetics*, 13(6), pp.395-405.
- 12 Milinovich, A. and Kattan, M., 2018. Extracting and Utilizing Electronic Health Data from Epic for Research. *Annals of Translational Medicine*, 6(3), pp.42-42.
- 13 Sendak, M., Balu, S. and Schulman, K., 2017. Barriers to Achieving Economies of Scale in Analysis of EHR Data. *Applied Clinical Informatics*, 08(03), pp.826-831.
- 14 McCord, K., Ewald, H., Ladanie, A., Briel, M., Speich, B., Bucher, H. and Hemkens, L., 2019. Current Use and Costs of Electronic Health Records for Clinical Trial Research: A Descriptive Study. *CMAJ Open*, 7(1), pp.E23-E32.
- 15 Ong, T., Pradhananga, R., Holve, E. and Kahn, M.G., 2017. A Framework for Classification of Electronic Health Data Extraction-Transformation-Loading Challenges in Data Network Participation. *eGEMs (Wash DC)*, 5(1):10.
- 16 Priest, E., Klekar, C., Cantu, G., Berryman, C., Garinger, G., Hall, L., Kouznetsova, M., Kudyakov, R. and Masica, A., 2014. Developing Electronic Data Methods Infrastructure to Participate in





Collaborative Research Networks. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, 2(1), p.18.

- 17 OHDSI: OMOP Common Data Model [Internet]. Ohdsi.org. 2020 [cited 1 May 2020]. Available from: <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- 18 FDA's Sentinel Initiative. [Internet]. Fda.gov. 2021 [cited 4 February 2021]. Available from: <https://www.fda.gov/safety/fdas-sentinel-initiative>
- 19 i2b2: Informatics for Integrating Biology & the Bedside [Internet]. i2b2.org. 2020 [cited 1 November 2020]. Available from: <https://www.i2b2.org/>
- 20 PCORNet [Internet]. Pcornet.org. 2021 [cited 3 February 2021]. Available from: <https://pcornet.org/>
- 21 Common Data Model Harmonization [Internet]. Healthit.gov. 2021 [cited 3 February 2021]. Available from: <https://www.healthit.gov/topic/scientific-initiatives/pcor/common-data-model-harmonization-cdm>
- 22 Glaser, B.G., Strauss, A.L. *Discovery of Grounded Theory: Strategies for Qualitative Research*. Somerset, NJ: Transaction Publishers; 1999.
- 23 NIH RFI Regarding the Use of FHIR® Standard for Research Purposes [Internet]. Nih.gov. 2020 [cited 1 May 2020]. Available from: <https://datascience.nih.gov/news/new-rfi-seeks-public-input-on-using-fhir>
- 24 Consolidated CDA Overview [Internet]. Healthit.gov. 2021 [cited 1 January 2021]. Available from: <https://www.healthit.gov/topic/standards-technology/consolidated-cda-overview>
- 25 United States Core Data for Interoperability (USCDI) [Internet]. Healthit.gov. 2020 [cited 1 December 2020]. Available from: <https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>
- 26 REDCap [Internet]. Project-redcap.org. 2020 [cited 1 May 2020]. Available from: <https://www.project-redcap.org/>
- 27 eCare Plan Joint NIH/NIDDK AHRQ Project [Internet]. Ahrq.gov. 2021 [cited 19 February 2021]. Available from: <https://ecareplan.ahrq.gov/>
- 28 SUPER REDCap on FHIR® now available in Epic [Internet]. Weill.cornell.edu. 2021 [cited 19 February 2021]. Available from: <https://its.weill.cornell.edu/news/super-redcap-on-fhir-now-available-in-epic>
- 29 Cumulus: A Universal Research Sidecar for a SMART Learning Healthcare System [Internet]. Healthit.gov. 2021 [cited 1 January 2021]. Available from: <https://www.healthit.gov/topic/scientific-initiatives/leap/cumulus-universal-research-sidecar>
- 30 HHS Announces Health IT Awardees Focused on Data Sharing to Support Clinical Care, Research, and Improved Outcomes [Internet]. Hhs.gov. 2021 [cited 19 February 2021]. Available from: <https://www.hhs.gov/about/news/2020/09/23/hhs-announces-health-it-awardees-focused-on-data-sharing-to-support-improved-outcomes.html>
- 31 FHIR® Factories: An Evolving Digital Architecture to Scale Health Research [Internet]. Healthit.gov. 2021 [cited 26 February 2021]. Available from: <https://www.healthit.gov/topic/scientific-initiatives/leap/fhir-factories>
- 32 SMART Health IT Project [Internet]. Smarthealthit.org. 2020 [cited 1 November 2020]. Available from: <https://smarthealthit.org>
- 33 SMART/HL7® FHIR® Bulk Data Implementations [Internet]. Smarthealthit.org. 2020 [cited 1 November 2020]. Available from: <http://docs.smarthealthit.org/bulk-data-implementations/>
- 34 Structured Query Language: ISO/IEC 9075-1:2016 Part 1: Framework [Internet]. Iso.org. 2021 [cited 19 February 2021]. Available from: <https://www.iso.org/standard/63555.html>
- 35 Logical Observation Identifiers Names and Codes (LOINC) [Internet]. Loinc.org. 2021 [cited 1 January 2021]. Available from: <https://loinc.org/>





- 36 Social Determinants of Health [Internet]. Health.gov. 2020 [cited 1 December 2020]. Available from: <https://health.gov/healthypeople/objectives-and-data/social-determinants-health>
- 37 FDA Real-World Evidence: Real-World Data and Real-World Evidence Are Playing an Increasing Role in Health Care Decisions [Internet]. Fda.gov. 2020 [cited 1 May 2020]. Available from: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>
- 38 Vulcan Accelerator [Internet]. HL7.org. 2021 [cited 1 February 2021]. Available from: <http://www.hl7.org/vulcan/>
- 39 Accelerating Application Programming Interfaces for Scientific Discovery: Consumer Perspectives [Internet]. Healthit.gov. 2021 [cited 1 January 2021]. Available from: <https://www.healthit.gov/sites/default/files/page/2020-11/Accelerating APIs Consumer Perspective.pdf>
- 40 HIPAA Privacy Rule 45 CFR Parts 160 and 164: Standards for Privacy of Individually Identifiable Health Information, Final Rule [Internet]. Hhs.gov. 2021 [cited 1 February 2021]. Available from: <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/privacyrule/prdecember2000all8parts.pdf>
- 41 Patient Generated Health Data [Internet]. Healthit.gov. 2020 [cited 1 May 2020]. Available from: [https://www.healthit.gov/topic/otherhot-topics/what-are-patient-generated-health-data#:~:text=Patient%2Dgenerated%20health%20data%20\(PGHD,help%20address%20a%20health%20concern](https://www.healthit.gov/topic/otherhot-topics/what-are-patient-generated-health-data#:~:text=Patient%2Dgenerated%20health%20data%20(PGHD,help%20address%20a%20health%20concern)
- 42 OAuth 2.0 [Internet]. Oauth.net. 2021 [cited 1 February 2021]. Available from: <https://oauth.net/2/>
- 43 Northwestern University Office for Research, Institutional Review Board Office [Internet]. Northwestern.edu. 2020 [cited 1 November 2020]. Available from: <https://www.ibr.northwestern.edu/informed-consent/>
- 44 Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act Privacy Rule [Internet]. Hhs.gov. 2021 [cited 1 February 2021]. Available from: <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>
- 45 The Debate Over “Re-Identification” of Health Information: What Do We Risk? [Internet]. Healthaffairs.org. 2021. [cited 1 February 2021]. Available from: <https://www.healthaffairs.org/doi/10.1377/hblog20120810.021952/full/>
- 46 Sweeney, L., Yoo, J., Perovich, L., Boronow, K., Brown, P. and Brody, J. Re-Identification Risks in HIPAA Safe Harbor Data: A Study of Data from One Environmental Health Study. *Technology Science*. 2017082801. August 27, 2017.
- 47 Yoo, J., Thaler, A., Sweeney, L. and Zang, J. Risks to Patient Privacy: A Re-Identification of Patients in Maine and Vermont Statewide Hospital Data. *Technology Science*. 2018100901. October 08, 2018.
- 48 Guidance Regarding Protection of Genetic Information [Internet]. Hhs.gov. 2021 [cited 1 February 2021]. Available from: <https://www.hhs.gov/hipaa/for-professionals/faq/354/does-hipaa-protect-genetic-information/index.html>
- 49 Key Privacy and Security Considerations for Healthcare Application Programming Interfaces (APIs) [Internet]. Healthit.gov. 2020 [cited 1 May 2020]. Available from: <https://www.healthit.gov/sites/default/files/privacy-security-api.pdf>
- 50 SMART App Launch: Scopes and Launch Context [Internet]. HL7.org. 2020 [cited 1 December 2020]. Available from: <http://hl7.org/fhir/smart-app-launch/scopes-and-launch-context/index.html>

