FHIR for Federal Health Research Studies

Policy, Data, and Technical Questions About Using Fast Healthcare Interoperability Resources (FHIR®) for Federal Health Research Studies

Health Community of Interest

Date Released: May 10, 2024

Synopsis

This paper was developed in support of the National Institutes of Health (NIH) *All of Us* Research Program and other Federal health research studies that are considering using Fast Healthcare Interoperability Resources (FHIR®) to acquire electronic health record (EHR) data and other data sources (e.g., health care claims data), as an additional method for collecting health data. This paper focuses on four different areas of FHIR generally related to Federal health research studies: (1) policy, (2) data model, (3) technology, and (4) next steps/pilot studies.

American Council for Technology-Industry Advisory Council (ACT-IAC)

ACT-IAC is a non-profit educational organization established to accelerate government mission outcomes through collaboration, leadership, and education. ACT-IAC provides a unique, objective, and trusted forum where government and industry executives are working together to improve public services and agency operations through the use of technology. ACT-IAC contributes to better communication between government and industry, collaborative and innovative problem solving, and a more professional and qualified workforce.

The information, conclusions, and recommendations contained in this publication were produced by volunteers from government and industry who share the ACT-IAC vision of a more effective and innovative government. ACT-IAC volunteers represent a wide diversity of organizations (public and private) and functions. These volunteers use the ACT-IAC collaborative process, refined over forty years of experience, to produce outcomes that are consensus-based.

To maintain the objectivity and integrity of its collaborative process, ACT-IAC welcomes the participation of all public and private organizations committed to improving the delivery of public services through the effective and efficient use of technology.

Health Community of Interest

The ACT-IAC Health Community of Interest is comprised of government and industry stakeholders advancing Federal health outcomes through collaboration and IT enhancement to include support strategy, policy-making and governance, support government on health-related projects, drive to improve health customer experience along with modernization initiatives, EHR adoption and implementations, medical device security and integration, precision medicine, data interoperability and data transparency, and health-related legislative impacts on Federal health agencies.

Disclaimer

This document has been prepared to contribute to a more effective, efficient, and innovative government. The information contained in this report is the result of a collaborative process in which several individuals participated. This document does not – nor is it intended to – endorse or recommend any specific technology, product, or vendor. Moreover, the views expressed in this document do not necessarily represent the official views of the individuals and organizations that participated in its development. Every effort has been made to present accurate and reliable information in this report. However, neither ACT-IAC nor its contributors assume any responsibility for consequences resulting from the use of the information herein.

Copyright

For further information, contact the American Council for Technology-Industry Advisory Council at (703) 208-4800 or www.actiac.org.

Table of Contents

Introduction

The ACT-IAC Health Community of Interest launched a project for the National Institutes of Health (NIH) *All of Us* Research Program to explore using **Fast Healthcare Interoperability Resources (FHIR®)** to acquire electronic health record (EHR) data as an additional method to those currently being used.[1] FHIR is a web-based standard for exchanging healthcare information developed by the standards development organization Health Level Seven (HL7®) International, Inc. A collaborative project team was assembled to assist *All of Us* in addressing specific questions related to the use of FHIR in Federal health research studies.

The NIH *All of Us* Research Program aims to gather biomedical data from over one million participants to accelerate research and improve health. Roughly 95% of enrolled participants authorize sharing of their EHR data through an authorization. However, there are several challenges with acquiring this data.

First, the most recent data release (CDR V7) contains EHR data from over 287,000 participants, representing approximately 70% of all those who consented to the study and provided a HIPAA authorization. The EHR data is acquired primarily from over 50 healthcare provider organizations (HPO) that partner with *All of Us* for enrollment. Consequently, EHR data external to HPOs is limited because it does not typically capture encounters outside of HPOs, and medical coding and record transfer can vary between organizations. This can be a source of bias for secondary research applications.

Second, one study showed that as much as 33% of *All of Us* participant EHR data lies outside of HPOs where a participant enrolled in the study.[2] This is not surprising considering the fragmentation of care and mobility patterns of people in the US.

Third, HPOs provide internal technology and informatics resources to curate EHR data and send it to the program quarterly. HPOs may leave the program in varying cycles while the participants remain enrolled. Due to this activity, alternative means for acquiring HPO-based EHR data are needed.

Finally, new HPOs entering the program may be under-resourced and unable to support EHR data curation. In parallel, as a part of the 21st Century Cures Act, the Office of the National Coordinator for Health Information Technology (ONC) mandated certified EHR vendors provide endpoints for all customers in a machine-readable format at no charge, such as FHIR, to facilitate the exchange of data. This mandate enables smaller organizations that may not have an informatics team to contribute to large data sharing initiatives.[3] Consequently, FHIR is being considered by *All of* Us as an alternative approach to acquire EHR data and overcome all these challenges.

As a result of these challenges, this paper focuses on four different areas of FHIR related to Federal health research studies. These areas are addressed as specific questions on:

- **Policy** – how do current regulations and standards support the use of FHIR for research purposes and what is required to enable the exchange of data?
- **Data Model** – how does the FHIR data model enable the exchange of information with other common data models, and specifically with the common data model that is used by *All of Us* and other large population-based research efforts?
- **Technology** – what is the state of readiness for FHIR data exchange at healthcare organizations today and in the near future?
- **Next Steps/Pilot** – what are the next steps and pilots that agencies and programs, including *All of Us*, should consider in leveraging FHIR for research studies?

Given the complexity around the implementation of FHIR and the associated standards and policies, as well as the novel use of FHIR in support of Federal health research studies, there is no straightforward implementation or single recommendation to best move forward. Rather, this paper can be used as a resource to frame the current and near-term environment, while also providing a set of next steps and pilots that can help the *All of Us* Research Program and other Federal agencies move forward on their journey.

Background

The ***All of Us* Research Program** is a historic effort to gather data from at least 1 million people living in the US, with the goal of accelerating research and medical breakthroughs and enabling individualized prevention, treatment, and care. *All of Us* is guided by a set of core values, including diversity, transparency, and accessibility. The program is creating a national research resource to inform thousands of studies, covering a wide variety of health conditions. Researchers use data from the program to learn more about how individual differences in lifestyle, environment, and biological makeup can influence health and disease.

More than 750,000 adult participants have enrolled in *All of Us* since 2017, including over 520,000 who have authorized sharing EHR data. *All of Us* comprises a partner network of more than 100 organizations supporting community outreach and engagement, participant enrollment and retention, study operations, and longitudinal cohort management. Most program participants have enrolled with support from a network of HPO partners, including medical centers, Federally Qualified Health Centers, and US Department of Veterans Affairs medical centers.

HPO partners currently extract participant data from EHRs, transform the extracted data into Observational Medical Outcomes Partnership (OMOP) tables and send it to the *All of Us* Data and Research Center (DRC) for curation and dissemination. The DRC places this data into the *All of Us* Researcher Workbench where access is granted to registered researchers.

All of Us would like to explore using FHIR to acquire EHR data as an additional method to those currently being used. Although *All of Us* would like to initially focus its exploration of FHIR for acquiring participant EHR data, the program is interested in using FHIR for other data sources, such as health care claims data, in the future. The focus on FHIR reflects its current and near-term future state implementation capability.

Methods

A collaborative ACT-IAC project team was assembled to assist *All of Us* in addressing specific questions related to the novel use of FHIR in Federal health research studies. The project team is comprised of members of industry and government who have volunteered their time to assist the *All of Us* Research Program and the community. The project began in September 2023 and concluded their work in February 2024. During this period, the team met several times to discuss the organization of the paper and the findings.

Information for this report was acquired through the expertise and personal experience of the team members, key informant discussions, review of the literature, etc. Sources referenced by the project team included specific government policies and documentation from the Office of the National Coordinator for Health Information Technology (ONC), the US Department of Health and Human Services (HHS), the White House Office of Science and Technology Policy (OSTP), the National Institute of Standards and Technology (NIST), and the Code of Federal Regulations (CFR). Non-governmental organizations that are supporting the implementation of FHIR were also referenced, including HL7, the Sequoia Project, and the CARIN (Creating Access to Real-time Information Now through Consumer-Directed Exchange) Alliance.

Individuals from ONC and the Sequoia Project provided their time to address specific items from the project team. In addition, a representative from a Qualified Health Information Exchange Network (QHIN) provided their input.

Industry expertise includes direct engagement on efforts such as the CodeX HL7 FHIR Accelerator and the HL7 Vulcan Accelerator Working Groups, FHIR Implementation Guides (IGs), real-world FHIR implementations, and technical expertise in health data exchange.

Overview of FHIR

FHIR is a web-based standard for exchanging healthcare information, developed by the standards development organization HL7. FHIR refers to a set of healthcare standards and specifications, the result of HL7's Standards Development Organization (SDO) "Fresh Look and Resources for Health" effort circa 2011. FHIR's development was in response to market needs for faster, easier, and better methods to exchange the rapidly growing amount of health data. This growth in the availability of new health data, along with the progressing "app" economy, created the need for clinicians and consumers to be able to share data in a lightweight, real-time fashion using modern internet technologies and standards. FHIR, as a standard, recently celebrated its ten (10) year anniversary.[4]

FHIR aims to facilitate the seamless and secure sharing of electronic health information (EHI) between different healthcare systems, providers, and applications. It allows healthcare information, including clinical and administrative data, to be available securely to those who have a need to access it and to those who have the right to do so for the benefit of a patient receiving care and other purposes, including research. The FHIR standard defines how healthcare information can be exchanged between different computer systems regardless of how it is stored in those systems. HL7 uses a collaborative approach to develop and upgrade FHIR, seeking stakeholder input and developing a consensus-based response to adoption and adaption changes to the standard.

While FHIR has captured the global healthcare services community's attention, it is in the earliest stages of widespread adoption. Numerous national programs across Europe, the UK, Canada, and the US have publicly identified FHIR as the future standard they will focus their efforts toward adopting. The impetus for this effort in the US has been the 21st Century Cures Act.[5]

Built upon the principles of simplicity, modularity, and extensibility, FHIR leverages widely used technologies such as Representational State Transfer (RESTful), application programming interfaces (APIs), JavaScript Object Notation (JSON), and Extensible Markup Language (XML) to represent healthcare data in a human-readable and machine-processable format. FHIR's data model is organized into modular building blocks called "resources," which can be combined to capture complex healthcare scenarios and workflows. An example of a FHIR Observation would contain all kinds of objective observation information of health and medical care, such as Vital signs, laboratory data, clinical findings, imaging results, device measurements, and clinical assessments.

FHIR profiles are essential components of the FHIR standard, providing a means to tailor and constrain base FHIR resources to address specific use cases, regional requirements, or healthcare domains. Profiles define the specific structure and rules for creating and validating healthcare data instances, ensuring they adhere to a consistent, interoperable format. This customization is achieved by specifying which data elements must be included, which can be omitted, and any additional constraints or terminology bindings required for a particular scenario.

FHIR was envisioned to support edge-based information exchange, by interacting with internal adapters capable of transforming or integrating with organizational legacy data sources (e.g., EHRs). However, there have been several implementers who have begun accepting the FHIR Resource(s), and Profiles as a representative Domain Model – with designs to persist or store FHIR information exchange in a raw FHIR native format.

The following common FHIR constructs provide context on the standards, resources, and services that are referenced throughout the paper:

- **FHIR Standard Version(s):** The current Regulatory floor for FHIR is Release 4 (R4), specifically v4.0.1. While the HL7 Argonauts FHIR Draft Standard for Trial Use 2 (DSTU2) specification is not interoperable with many of the FHIR R4 resources or FHIR Implementation Guides (IG), it should be noted that it is still being used by some organizations today; even so, it is an outlier. And while HL7 FHIR R5 was published in early 2023, FHIR R4 is expected to remain the primary FHIR Standard for many years to come. Current work on FHIR R6 is being evaluated, and any future release date is potentially very far off. Plans to remain with R4 at this point would be a solid decision based on the landscape of the US Federal Agencies and commercial vendors.
- **FHIR Capability Statement:** The FHIR standard maintains very few "required" elements. The conformance or "Capability" statement is something that every production FHIR endpoint is required to publish.[6] The Capability Statement resource documents the behaviors and functionality implemented from the HL7® FHIR® standard for a particular implementation. This remains a powerful resource if correctly leveraged and utilized. It is a way to evaluate what FHIR resources, FHIR IG, and FHIR operations are supported, it also provides an independent review of an organizations' attention to detail. Organizations with invalid, or a non-accessible Capability Statements are indicators of less sophisticated implementations and may be precursor to other conformance issues which may affect the quality of the data being exchanged.
- **FHIR Maturity Model:** FHIR maintains maturity levels (FHIR Maturity Model) for all FHIR Resources.[7] The goal as a community is to advance resources and profiles to a maturity level of "N" for normative. For resources which achieve a level of Normative, it is understood that they will not introduce "breaking changes", or if they do, it will be an exceedingly rare occurrence.
- **FHIR Implementation Guides (IGs):** Many of the base FHIR Resources and default profiles define everything to be optional or a cardinality of 0..1 (not required). It remains the responsibility of the FHIR IGs to create these implementation details (rules) with which a profile must comply for the data to be conformant or compliant. The ONC and the HL7 Accelerator, the Argonauts, have been jointly working together. With ONC being the primary driver for the US Core for Data Interoperability (USCDI) (and USCDI+[8]) to define the requirement elements, and the Argonauts taking the USCDI and creating the associated FHIR IG US Core. This work is now the responsibility of the US Realm Steering Committee and the Cross Group Projects work groups of HL7.

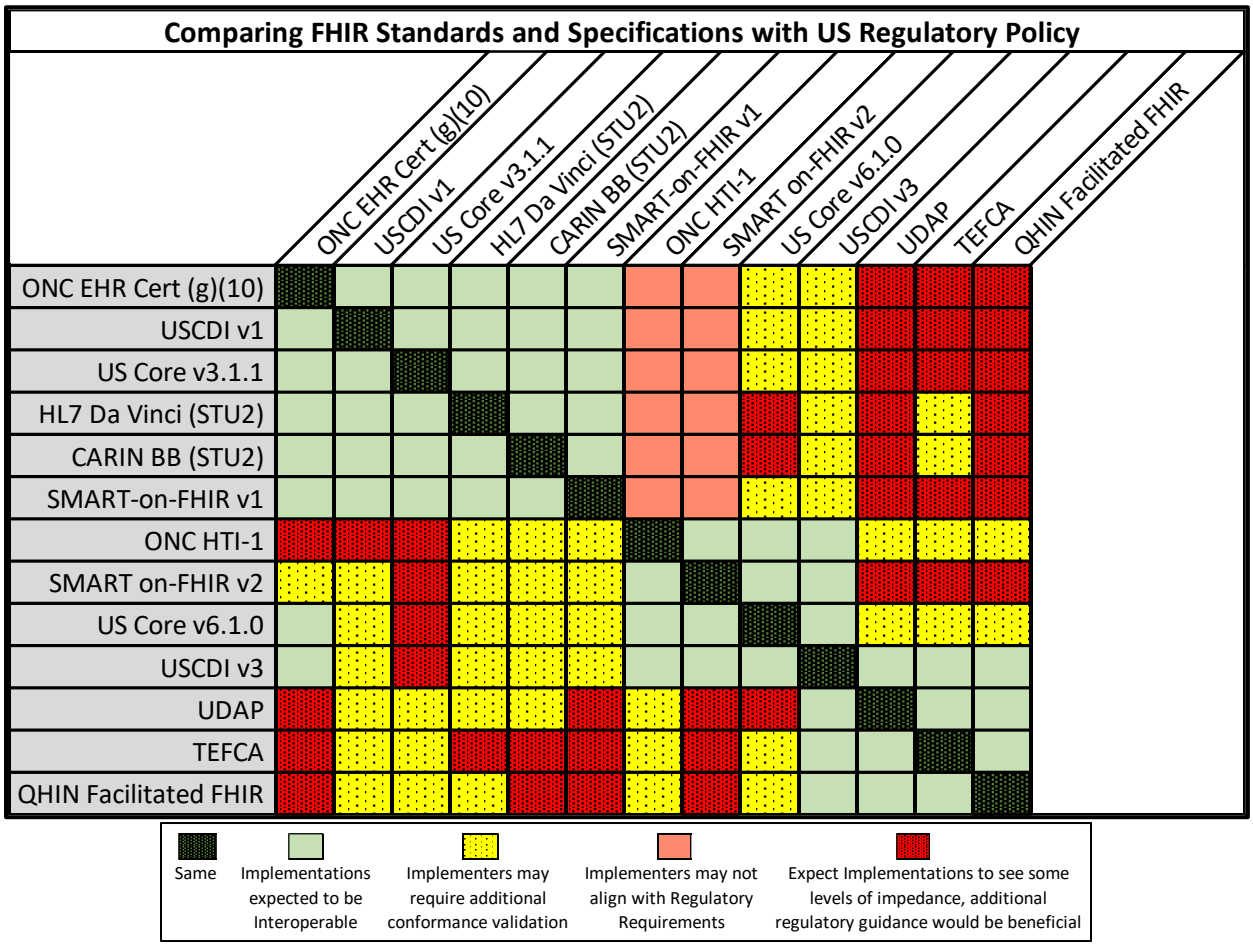
Both the ONC and CMS (along with many other US Federal healthcare initiatives) are incorporating the US Core into Regulation in support of the 21st Century Cures Act. The current regulatory floor according to the Final Rules is USCDI v1 and US Core v3.1.1. ONC published the Health Data, Technology and Interoperability: Certification Program Updates, Algorithm Transparency and Information Sharing (HTI-1) Final Rule Overview in December 2023 mandating that USCDI v3 and US Core v6.1.0 must be implemented no later than January 1, 2026.[9]

- **FHIR Extensions:** While the many FHIR Resources, and their default Profiles, provide a lot of capabilities and flexibility to implementers, there continue to be situations in which profiles require necessary extensions. Profiles and Extensions are part of the US Core,[10] and it is important that implementers remain conformant with these extensions when implementing.
- **FHIR Validation Services:** One of the most important aspects of FHIR is the FHIR Validation Services. Many organizations are seeking to improve overall data quality (e.g., AcademyHealth's Health Datapalooza and The Sequoia Project Data Usability work efforts) including the Agency for Healthcare Research and Quality (AHRQ) – and their findings on "Improving Data Collection and Standardization Across the Healthcare System."[11] The FHIR IG rules of the road are captured within an IG Package which are published into a FHIR Server Validation Service.

There is no current mechanism (other than human review) to resolve conflicting requirements across IGs, even within the scope of specific Accelerator projects such as Da Vinci. It is the role of this Validation service, and its associated IG packages, to ensure data continues to conform to rules specified in the IGs. When a FHIR message exchange happens without the benefit of being reviewed by a FHIR Validation service, either on the response side before it goes out the door, or by the requester when it receives a FHIR payload, the quality of the data may not always be fully conformant. This remains the decision of the FHIR Developers, Integrators, and Operations to evaluate these tradeoffs within their organizations.

Figure 1 provides a representation of the complexities of the current Regulatory Policies, Standards and Specifications being implemented across the US Healthcare continuum. The figure illustrates where version interaction may occur, these intersections of policy and standards will be supported across an integrated ecosystem of healthcare systems of data producers and data consumers each potentially with their own approach and implementation.

With this context, the paper addresses the specific questions posed by the *All of Us* Research Program as to if, how, and when FHIR can be leveraged to support Federal health research studies.



USCDI – The ONC publishes an annual set of standardized health data classes and constituent data elements for nationwide, interoperable health information exchange. v1 was published in May 2020, and v3 was published in July 2022.

US Core – The HL7 Accelerator Argonauts has published a FHIR IG which aligns with each version of the USCDI. USCDI v1 is implemented through US Core v3.1.1, while USCDI v3 is implemented through US Core v6.1.0.

HL7 Accelerators – Da Vinci and CARIN Blue Button® (BB) are other HL7 Accelerators which have leveraged US Core IG within their business and use cases. FHIR IGs which are not normative yet may carry the designation Standard for Trial Use (STU). Many of these IGs currently published align with or support US Core v3.1.1 only.

SMART and UDAP – SMART-on-FHIR and Unified Data Access Profiles (UDAP) represent possible security models for FHIR based exchanges. ONC (g)(10), HTI-1, and EHR Certification currently point to SMARTv1 or SMARTv2. The Trusted Exchange Framework and Common Agreement℠ (TEFCA℠) currently references its intent to operationalize UDAP.[12]

TEFCA and QHIN – Under ONC TEFCA, the QHINs will work to operationalize FHIR based information exchanges. Currently being proposed is either brokered or facilitated FHIR exchange across the QHINs.

Figure 1. Aligning FHIR standards and specifications with Federal Regulation

Addressing the Key Questions

Section I - Policy

This initial section focuses on a series of FHIR policy questions related to:

- The releasability of EHR data via FHIR for Federal health research studies, including how Trusted Exchange Framework and Common Agreement℠ (TEFCA) provides an overarching governance structure and policy that facilitates secure exchange using FHIR and its current limitations.
- The process of identity proofing and matching study participants for EHR data release via FHIR, including the identification of applicable standards, potential solutions, and assessing the role of CARIN Alliance standards in addressing the problem.
- The current set of permitted exchange purposes that can be leveraged by Federal health research studies through TEFCA for obtaining participant data.
- The requirements and timeline for including research as a permitted exchange purpose in Federal health research studies.

Question 1 - Authorization to Share EHR Data for Research Purposes

1. *Some Federal health research studies are not covered entities as defined under HIPAA and therefore do not provide a HIPAA authorization to study participants. Instead, these studies request that participants sign an authorization permitting HIPAA covered entities holding their EHRs to share those data with All of Us for research purposes (e.g., Authorization to Share My EHRs for Research). Is this sufficient, or not, to allow provider organizations to release patient EHR data on study participants via FHIR-facilitated means for Federal health research studies?*

The use of centralized Authorization for an individual to share EHR data for research purposes allows provider organizations, like the *All of Us* Research Program's network of HPO partners, to share data with the program. Access to participant EHR data can be authorized by the study participant by signing either a centralized or a provider organization's authorization form. However, use of point-to-point organizational agreements may be cumbersome for the program and its participants and could limit the scalability of the program.

TEFCA for Nationwide Health Information Interoperability was established to decrease the burden of building and maintaining costly, point-to-point interfaces for data exchange between organizations. TEFCA provides an overarching governance structure and policy that facilitates secure exchange using FHIR for improved care and welfare of populations. To date, TEFCA governance and policy supports data exchange for six purposes: treatment, payment, health care operations, public health, government benefits determination, and individual access services (IAS). Exchange for research has been identified as a future purpose, however, resources and

processes developed for other exchange purposes may lend themselves to FHIR-facilitated data exchange in support of Federal health research studies.

TEFCA-exchange is facilitated by a QHIN organization who signs the Common Agreement (CA) for Nationwide Health Information Interoperability and onboards participants and sub-participants who the agreement's clauses flow-down to. The IAS exchange purpose was intended to assist individuals in obtaining access to their health information, however, does not allow an individual to direct that personal health information (PHI) data to be sent to a third party within the confines of the TEFCA architecture. Details are provided in the Leveraging IAS Exchange Purpose discussion below (within the response to Section I, Question 3) as part of describing how IAS may support obtaining an individual's health records for use within a research study.

The Sequoia Project, which serves as the Recognized Coordinating Entity (RCE)—a neutral, stakeholder-driven, public-private collaborative whose sole mission is advancing secure, trusted, interoperable health data sharing across the US—provides an example of how data might be requested from an IAS provider at scale under TEFCA,[13] as shown in

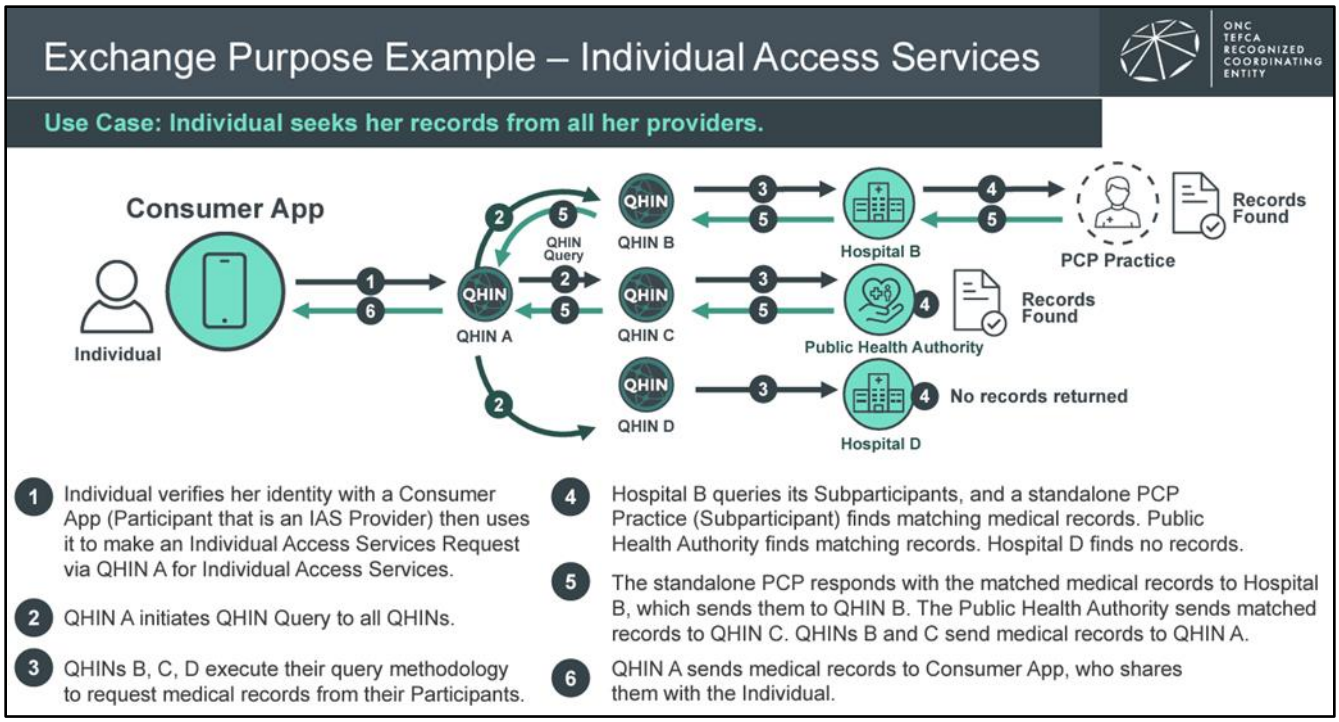


Figure **2**.

Figure 2. Example of data exchange using the IAS purpose as provided by the Sequoia Project.

Currently, no Federal program has used FHIR and a TEFCA facilitated exchange under the IAS exchange purpose to provide the participant data that would be desired for the *All of Us* research

use case. To implement such a use of the TEFCA exchange, it would require several steps, starting with the use of a consumer application by an IAS Provider. The organization would be required to verify the individual's identity and make an individual access request to a single QHIN, which would initiate a request to other QHINs and all QHIN participants and sub-participants. Data collected by the consumer application through the IAS provider organization would then be governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Rules and other applicable law. Importantly, IAS Providers must adhere to extensive privacy and security requirements.[14]

1.1. Assuming provider organizations recognize Federal health research studies authorization to share EHR data, what data restrictions (e.g., HIV data, mental health data, etc.) would apply to said release in the following jurisdictions/ domains (e.g., Federal, State, or local, HIE or HIN)?

Regardless of the data exchange mechanism and agreements, some data requests may be denied. Per 45 CFR § 164.524(a)(2), a provider organization may restrict access to some EHR data including 1) psychotherapy notes; and 2) information compiled in reasonable anticipation of, or for use in, a civil, criminal, or administrative action or proceeding. Additionally, organizations may deny individual access to records without review for use cases including, but not limited to those that may pertain to legal proceedings; jeopardize the health, safety, or wellbeing of an individual; impact an ongoing clinical trial; or if the requested records are controlled by a Federal agency under the Privacy Act. Any covered entity who receives an access request from an individual must inform that individual if the request has been approved or denied. If a request has been denied, the covered entity is required to provide a written statement indicating the basis for the denial, applicable review rights, and details of how the individual may file a complaint regarding the decision.

Restrictions on data exchanged for purposes that are not covered under TEFCA (e.g., patient authorized exchange for research) vary at the state-, local-, or organization-level. For example, laws and policies regarding state health information exchange (HIE) consent requirements to disclose mental health information for treatment, payment, and health care operations are variable across states. Additionally, conditions (e.g., mental health, HIV [human immunodeficiency virus] exposure or infection) may impact release of patient information. Some information on policy variations across states has been made available publicly;[15] however, to our knowledge, no comprehensive resource exists that provides documentation of the specific data release restrictions across jurisdictions.

Question 2 – Identity Proofing

2. What is required to identity proof (for matching and release of data for) Federal health research study participants so that provider organizations release patient EHR data via FHIR-facilitated means?

To prove the identity of research study participants, match that participant to provider organization data, and release a participant's EHR data via FHIR-facilitated means, the process typically involves the following steps:

- **Authentication and Authorization**: Secure authentication processes are necessary to prove that participants and provider organizations are who they claim to be. This includes the utilization of protected login credentials, like usernames and passwords. Additionally, an extra layer of security can be created through multi-factor authentication (MFA). The implementation of authorization processes regulates access to specific data based on a participant's role and permissions, as well as the role and permissions of a corresponding provider organization, research study, and/or study institution.
- **Consent**: Study participants must give consent for their data to be used in a research study. This involves a clear explanation of the study and its purpose, what participant data will be accessed, potential risks, and possible benefits. Consent can be acquired electronically, and participants may have the option to define the extent of data access. Note that consent forms should be accessible to all audiences by adhering to the NIH-Wide Strategic Plan for Diversity, Equity, Inclusion, and Accessibility (DEIA).
- **Identity Proofing**: Specific data will need to be collected to match participant identity. This can include verifying information such as name, date of birth, address, and other relevant details. Some participants may already have a unique identifier assigned to them within their EHR data. A unique identifier can be produced using hashing algorithms to convert existing PHI data into a code string that is not sensitive to share.
- **Security and Compliance**: To protect the confidentiality and integrity of information during data release, transfer should occur over secure channels and encrypted connections. Compliance with relevant privacy regulations, such as HIPAA, needs to be considered for data management, staff access, further data sharing, etc.
- **Audit Trail**: Audit trails allow for tracking of who accessed specific data, when it was accessed, and what occurred during data access. These trails assist monitoring and ensuring compliance with access policies.

2.1. Which standards are applicable to this problem?

Various standards from NIST as well as other non-governmental organizations are applicable to identity proof Federal health study participants. These standards contain guidelines and recommendations for digital identity management and authentication in information systems, both necessary for provider organizations releasing EHR data via FHIR-facilitated means.

Exploring the data model and related standards for applications of FHIR with regards to research purposes is discussed further in Section II, which includes approaches and limitations. Applicable standards to identity proof Federal health research study participants include:

- **NIST-800-63-3, Digital Identity Guidelines:** Technical requirements for Federal agencies implementing digital identity services and are not intended to constrain the development or use of standards outside of this purpose. The guidelines cover identity proofing and authentication of users interacting with government IT systems over open networks. They define technical requirements in each of the areas of identity proofing, registration, authenticators, management processes, authentication protocols, federation, and related assertions.[8] This standard also contains Identity Assurance Levels (IAL), Authenticator Assurance Levels (AAL), and Federation Assurance Levels (FAL), which would be considered in identity proofing, as well.
- **NIST SP 800-53 Rev. 5, Security and Privacy Controls for Information Systems and Organizations:** A catalog of security and privacy controls for information systems and organizations to protect operations and assets, individuals, and other organizations from a diverse set of threats and risks, including hostile attacks, human errors, natural disasters, and privacy risks. The flexible controls are customizable and implemented as part of an organization-wide process to manage risk. This consolidated control catalog addresses security and privacy from a functionality perspective and from an assurance perspective. Addressing functionality and assurance helps to ensure that information technology products and the systems that rely on those products are sufficiently trustworthy.[16]
- **RFC 3647:** A framework to assist the writers of certificate policies or certification practice statements for participants within public key infrastructures, such as certification authorities, policy authorities, and communities of interest that wish to rely on certificates. In particular, the framework provides a comprehensive list of topics that potentially need to be covered in a certificate policy or a certification practice statement.[17]
- **Open ID Connect (OIDC):** OpenID Connect is an interoperable authentication protocol based on the OAuth 2.0 framework of specifications (IETF RFC 6749 and 6750). It simplifies the identity verification of users based on the authentication performed by an authorization server, as well as obtaining user profile information in an interoperable way.[18]
- **OAuth 2.0:** "Open Authorization" is a standard designed to allow a website or application to access resources hosted by other web applications on behalf of a user. OAuth 2.0 provides consented access and restricts actions of what the client app can perform on resources on behalf of the user, without ever sharing the user's credentials.[19]

- **Unified Data Access Profiles (UDAP) Tiered OAuth:** To meet a broad demand for safer authentication options, data sources have started leveraging a trusted network of identity providers to authenticate users and obtain information about them to make an authorization decision. This distributed framework allows the reuse of existing user credentials and improves security by providing user data directly to the source rather than passing it through a third party. UDAP Tiered OAuth for User Authentication implements user authentication as an extension to the OAuth 2.0 authorization and OpenID Connect authentication processes, and the protocol implements tiered authorization and authentication requests.[20]

- **Security Assertion Markup Language (SAML):** SAML is an open standard used for authentication. Use of this standard enables access to multiple web applications using one set of login credentials. It works by passing authentication information in a specific format between two parties, usually an identity provider and a web application.[21]

2.2. Which solutions are applicable to solving this problem?

This report recommends leveraging a commonly adopted FHIR approach, Substitutable Medical Applications and Reusable Technologies (SMART®), as a solution to this problem. It is introduced here and expanded in this paper's proceeding sections. SMART Health IT serves as an open and standards-based technology platform, allowing for development of applications that seamlessly and securely operate throughout the healthcare system. By leveraging an EHR system or data repository that adheres to the SMART standard, patients, physicians, and healthcare professionals can access a diverse range of applications from the SMART library to enhance clinical care, facilitate research, and promote public health initiatives. Sync For Science (S4S) is a national collaboration among EHR vendors – including Allscripts, Cerner, eClinicalWorks, and Epic – and NIH, ONC, and Harvard Medical School.[22] S4S conducted a pilot project with *All of Us* in 2018 using FHIR API and a SMART server.[23]

SMART on FHIR

The SMART on FHIR specification is widely used for establishing a uniform methodology to address security and data requirements within health applications, which are the foundation necessary for identity proofing in the research domain.[24] SMART on FHIR delineates a workflow by which an application can securely request data access, obtain the requested data, and subsequently use that data. SMART on FHIR has three components:

- **Identity and Access Management:** SMART on FHIR employs the OpenID Connect (OIDC) identity management protocol to handle access to clinical data. This enables applications to request access to healthcare data, whether it is limited to read-only access for a few records or wider read/write access to an entire EHR. The SMART specifications define a customized version of OIDC tailored for use in the health, or research, context.

- **Access to Data:** SMART uses the FHIR standard for reading and/or updating data. Thus, in a SMART on FHIR architecture, a set of FHIR services are available for use by SMART applications. Access to these services is secured using the Identity and Access Management layer described in the first component.
- **Launch:** For web-based applications, SMART defines a consistent URL scheme that portals, EHR systems, and other healthcare applications can use to launch web-based applications. When launching an application, a specific context is passed to the application. This context can include information about the currently selected patient, clinical encounter details, or any relevant data needed by the application.

SMART on FHIR is explored further in Section III - FHIR Technology for Research.

There are also several FHIR IGs that could be used in support of identity proofing efforts.[25] They are as follows:

- Integrating the Healthcare Enterprise (IHE) Patient Master Identity Registry (PMIR): Supports the creating, updating, and deprecating of patient master identity information about a subject of care, as well as subscribing to these changes, using the HL7 FHIR standard and its RESTful transactions. In PMIR, "patient identity" information includes all information found in the FHIR Patient Resource such as identifier, name, phone, gender, birth date, address, marital status, photo, others to contact, preference for language, general practitioner, and links to other instances of identities. The "patient master identity" is a dominant identity managed centrally among many participating organizations (a.k.a., "Golden Patient Identity").
- Interoperable Digital Identity and Patient Matching: This FHIR IG provides guidance on leveraging Patient Matching and Digital Identity capabilities together to improve match quality and overall identity assurance in FHIR transactions. It defines methods to inform and execute cross organizational and internal patient matches via FHIR when requested for a permitted purpose or authorized by the Patient directly or by the Patient's delegate.
- Making EHR Data More available for Research and Public Health (MedMorph): This Reference Architecture enables clinical data exchange between EHR systems, public health systems/authorities, data repositories, and research organizations. This data exchange utilizes if applicable, knowledge repositories and backend services applications (e.g., FHIR APIs) to determine the triggering event(s) for the data exchange, the process for the data exchange, and validation that the data being exchanged meets a set of rules to expedite the data exchange.

2.3. How would the CARIN Alliance standards help solve this problem?

The CARIN Alliance is a bipartisan, multi-sector collaborative working to advance consumer-directed exchange of health information. They have completed, and continue to conduct, research in and around the challenge of identity proofing in healthcare data exchange. Their

research, along with collaborators, has resulted in several white papers, proofs of concepts, summits, and polices. CARIN is currently researching an open-source framework for federating trusted Identity Assurance Level 2 (IAL2) certified credentials across health care organizations using a person-centric approach which leverages modern technologies such as OpenID Connect and OAuth 2.0.[26] Much of their work in this domain falls under the umbrella of Digital Identity, as the Alliance explores best practices and standards for securely identifying, authenticating, and matching individuals to their health information across multiple health plans, providers, and HIEs in a trusted way with consumer consent.

The CARIN Alliance endorses utilization of SMART on FHIR in the healthcare ecosystem to ensure individuals have immediate access to their health information. Through their proof-of-concept project, they seek to demonstrate how individuals can voluntarily digital identity proof themselves in a trusted way without separate portal accounts with every data holder in possession of their health information. The project's objective was to scale an open framework for federating trusted NIST 800-63-3 Identity Assurance Level 2 (IAL2) certified credentials using a person-centric approach across healthcare organizations, leveraging modern technologies such as OpenID Connect and OAuth 2.0.[27] This future case would create a trusted identity proofed digital credential for accessing health information across multiple payers and providers.

This digital identity federation proof-of-concept is a "person-centric approach" that brings together the CARIN Alliance, HHS NextGen External User Management System (XMS) team, ONC, and the Centers for Medicare and Medicaid Services (CMS). By creating an automation of trust in their healthcare ecosystem use cases, CARIN Alliance has advanced the body of knowledge around identity matching and proofing. Their collaboration with offices, agencies, and programs at HHS could guide and advise *All of Us* as they explore implementing FHIR in acquiring secure, reliable, and accurate health research information. The final report presents two preferred paths toward digital identity federation: 1. leveraging HHS XMS (e.g., ID.me) as a national identity broker service, and 2. leveraging the UDAP Tiered OAuth Protocol (described in the previous section for identity proofing standards). Their recommendations may also be useful for the follow-on pilot project based on this paper's findings.

Question 3 – Permitted Exchange Purposes

3. Which permitted exchange purposes can be used by Federal health research studies through TEFCA for acquisition of study participant data? How would it be operationalized for each exchange purpose?

TEFCA provides a new record exchange mechanism to HIEs and other health record holders. Below are the currently permissible exchange purposes under the CA which are focused on treatment, workflows related to payment and eligibility, as well as public health.

The CA is the legal contract with each QHIN which defines the baseline legal and technical requirements for secure information sharing on a nationwide scale. The CA authorizes six (6)

Exchange Purposes, but initially Responses to Requests are required for Treatment and Individual Access Services since those standard operating procedures (SOPs) have been published. Responses will be required six months after the SOP for each authorized Exchange Purpose that is approved.

Currently Authorized Exchange Purposes and Definitions:[28, 29, 30]

1. **Treatment**[31, 32] **- Responses required starting June 2022**: Treatment is defined as the provision, coordination, or management of health care by one or more health care providers. It also includes the coordination or management of healthcare by a third-party provider, a consultation between providers treating a common patient, or the referral of a patient from one provider to another.
2. **Individual Access Services - Responses required starting March 16, 2023:**[33] Individual Access Services satisfies an Individual's ability to access, inspect, or obtain a copy of that Individual's Required Information that is then maintained by or for any QHIN, Participant, or Sub-participant.
3. **Healthcare Operations**[34, 35] **- Responses not required, Draft SOP released on April 2, 2023:**[36] Healthcare operations refer to certain administrative, financial, legal, and quality improvement functions of a covered entity (CE) essential to running its business and supporting treatment and payment activities.
4. **Government Benefits Determination - Responses not required:** Benefits determination refers to Federal or state government agencies deciding whether a person is eligible for Federal or state benefits for any reason other than health care.
5. **Payment**[37, 38] **- Responses not required:** Payment refers to the various activities that healthcare organizations use to obtain payment, or a reimbursement fee correlated with healthcare services. It also encompasses health plans acquiring premiums to satisfy their coverage responsibilities.
6. **Public Health - Responses not required:** The public health exchange purpose refers to any request, use, disclosure, or response authorized under HIPAA regulations or other applicable laws regulating public health activities.

Figure 3 provides the exchange purposes that are relevant to a variety of use cases, specifically public health surveillance, clinical trials, and observational longitudinal studies. Combinations of purpose and use case that cannot be utilized by the definitions of each term are left blank.

Exchange Purpose	Public Health Surveillance	Clinical Trial	Observational Longitudinal Study
Treatment	-	-	-
Individual Access Services	-	Indirectly supported	Indirectly supported
Healthcare Operations	-	-	-
Government Benefits Determination	-	-	-

Exchange Purpose	Public Health Surveillance	Clinical Trial	Observational Longitudinal Study
Payment	-	-	-
Public Health (PH)	Supported, but narrowly defined	Potentially supported if public health authority can be asserted against narrow PH definition	Potentially supported if public health authority can be asserted against narrow PH definition

Figure 3. Exchange Purposes and Possible Federal Study Utilization

There are currently no permitted exchange purposes that explicitly support government agencies for Federal health research studies allowed by TEFCA for the acquisition of study participant data. Below is a discussion of how the IAS and Public Health exchange purposes have the potential to be leveraged.

Leveraging Individual Access Service Exchange Purpose

The IAS satisfies an Individual's ability to access, inspect, or obtain a copy of that Individual's Required Information that is then maintained by or for any QHIN, Participant, or Sub-participant. The power of using TEFCA is the ability to obtain records for many if not all the study participant's health care providers (HCP) based on a single request. The downside to using the IAS purpose is the pull only nature of the request, preventing the potential to share information with any participant's HCPs, if that becomes a need for any Federal research study.

The most likely way to operationalize the IAS exchange purpose to obtain records will be through a mechanism to help study participants periodically request their own data, then share it with the research study organization. This solution builds on the example previously shown in Figure 3 from the RCE User Guide for IAS,[39] but extends that use case, as shown in Figure 4.

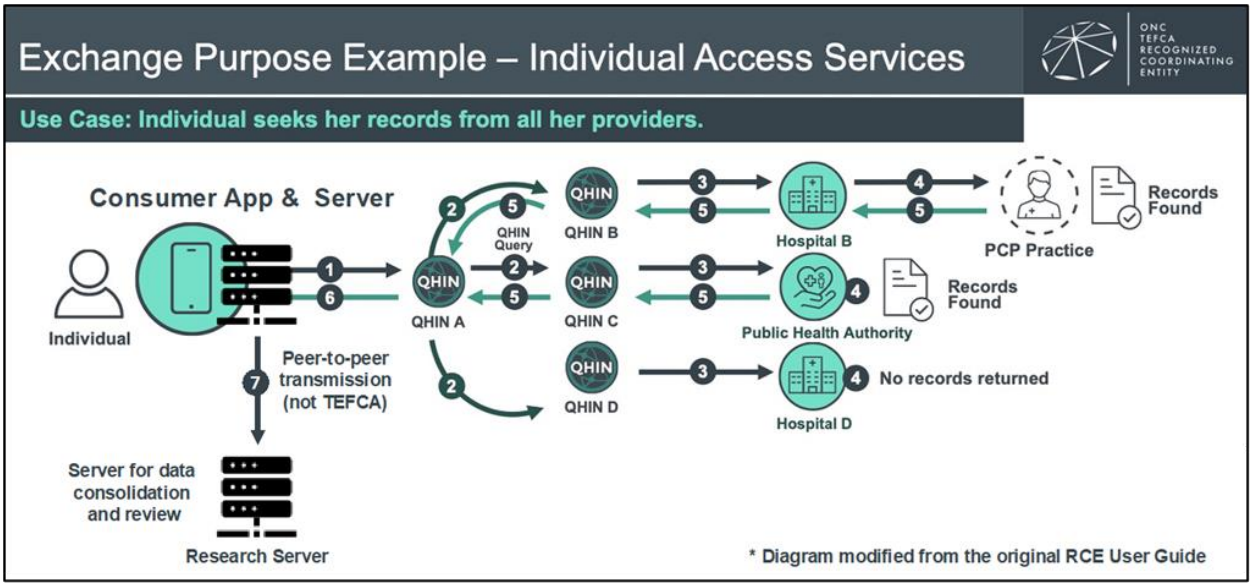


Figure 4. Example of data exchange using the IAS purpose as provided by the Sequoia Project with an extended use case

Either the Federal agency or a third party provider could establish a Consumer Application and would operate as a TEFCA Participant who is an IAS Provider. The application may request the study participant's (aka Individual) records in using the IAS exchange purpose, step 1 in the figure. The QHIN will query across the network to obtain all the participant's records. As the records are collected, the QHIN will return them to the participant, steps 2 through 6. The extension of the User Guide example is to enhance the Consumer Application and its associated server(s) to allow those records to be shared via peer-to-peer transmission with the research study organization's server(s), step 7. This transmission must be done outside of the TEFCA network since that record exchange is not currently a permitted use.

Based on feedback from ONC,[40] the proposed solution above may utilize TEFCA as long as it:

1. Contains transparent privacy and security notice in alignment with the SOP for IAS Provider Privacy and Security Notice and Practices
2. Will be used for valid exchange purposes, in this case IAS.
3. Complies with applicable law.

Directed Exchange or Patient Proxy

The final version of ONC's *The Trusted Exchange Framework (TEF): Principles for Trusted Exchange* indicates that the Individual Access Service provides patient access to their records as well as articulating that patients and their legal representatives should have the ability to direct their digital health information to any recipient they designate.[41] Specifically, HINs should support an individual's decision to access their digital health information through an API-enabled third-party application when the individual has directed the HIN to disclose a copy of that individual's health information to the application.

Based on conversations with ONC and one candidate QHIN,[42] the CA and related definitions ***do not support the IAS exchange purpose for a 3rd party to obtain records*** on behalf of a patient. The earlier principle from the TEF for directed exchange or patient proxy for records exchange is not supported currently. However, both organizations are hopeful for this type of exchange in the future. However, a patient can direct their records received via IAS to a third party if they originate the request as discussed earlier in this section.

Leveraging Public Health Exchange Purpose

There is a potential for a large scale observational, longitudinal study to leverage the Public Health exchange purpose. TEFCA includes the Public Health exchange purpose but has not developed the SOP for this use and as such does not require a response from TEFCA participants currently.

Based on Public Health Authorities, public health information may be requested from covered entities. As noted above in the Authorized Exchange Purposes and Definitions, the Public Health exchange purpose refers to any request, use, disclosure, or response authorized under HIPAA regulations or other applicable laws regulating public health activities. As such, a Public Health

Authority may request information, include records for a specific study participant, through the Public Health exchange purpose, if the request complies with applicable law, including restrictions specified in 45 CFR 164.514.[43]

The standard "minimum necessary requirement" of 45 CFR 164.514 indicates that the use and disclosure of public health information must limit any request for protected health information to that which is reasonably necessary to accomplish the purpose for which the request is made, when requesting such information. The request of complete medical records is addressed in 45 CFR 164.514(d)(5) stating that a covered entity may not use, disclose, or request an entire medical record, except when the entire medical record is specifically justified as the amount that is reasonably necessary to accomplish the purpose of the use, disclosure, or request.

To this end, the Public Health exchange purpose may be applicable, but would need to be clearly asserted by the research study organization, which is also a Public Health Authority, and comply with the specific public health purposes outlined in 45 CFR 164.512. Based on a conversation with a public health subject matter expert,[44] the use of a public health information request would typically need to be coordinated with the state health officials. Those officials commonly evaluate the information requests by weighing the urgency of the health need versus the trust of their community and how the community's data is used.

The bottom line for the Public Health exchange purpose is the need for the Federal government to be able to justify the use of their Public Health Authorities to fulfill longitudinal study objectives, and to do so in a manner consistent with 45 CFR 164.512. This will also require willing participants in each of the state and local health departments to determine whether this sort of information sharing is consistent with applicable Federal and state laws in their jurisdiction. The potential for overreach on the use of Public Health Authorities increases as the urgency for immediate action that is related to needed health data decreases. As such a longitudinal study seems unlikely to meet the threshold for urgent need.

Question 4 – Future of the Research Exchange Purpose

4. What is likely to be required from a Federal health research study to allow research as a permitted exchange purpose?

There are a number of steps required to add a research study exchange purpose, or any other new exchange purpose, under the TEFCA architecture that would be sufficient to support a Federal health research study.[45] Figure 5 provides the requirements and corresponding timeline.

The initial step is detailed in the Change Management portion of the CA and provides the steps necessary to authorize a new Exchange Purpose. While that approval process alone is expected to take three to six months, it may include at least one three-month extension. The amendment should include: 1) that research be added to the authorized Exchange Purposes, and 2) the research entities added to the list of authorized Participants and Sub-participants.

As noted in Figure 5, the addition and implementation of a new exchange purpose can take anywhere from 16 to 26 months in a best-case scenario if no additional hurdles are encountered, and no extensions are issued by any of the approving bodies.

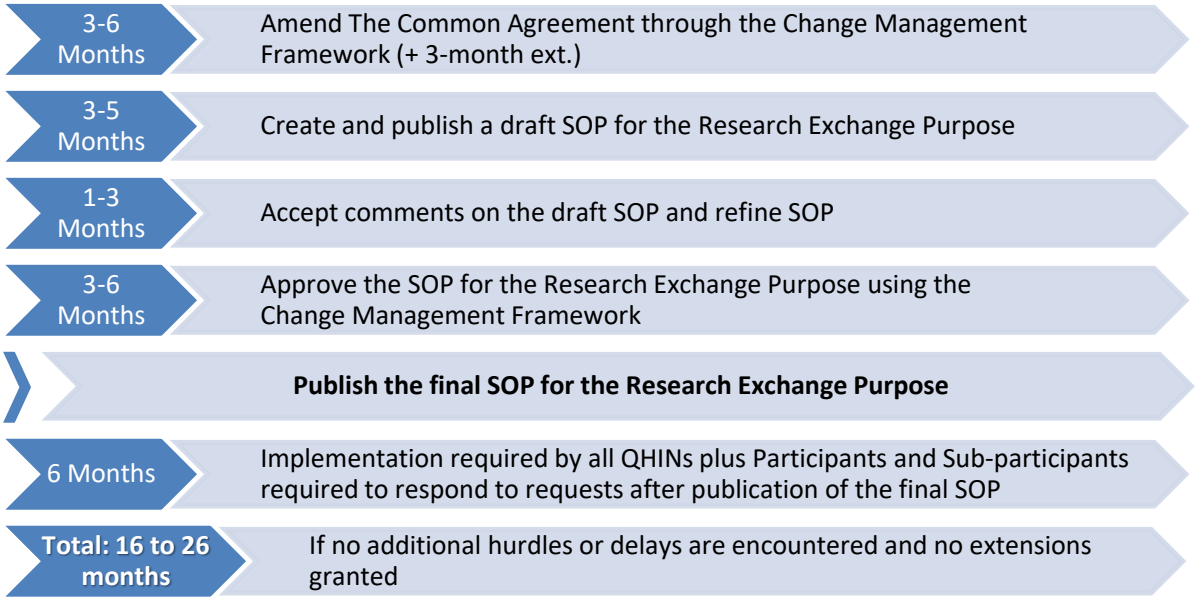


Figure 5. Timeline of requirements to add and implement a research study exchange purpose under TEFCA.

This process starts with the RCE in consultation with ONC, the Governing Council, and the QHIN and/or Participant/Sub-participant Caucuses to evaluate and provide feedback regarding the new Exchange Purpose. Once the RCE, in consultation with ONC, decides to proceed, approval requires at least two-thirds of the votes cast by the QHIN Caucus members within a 3-month period and then approval by ONC of the amendment in writing within a 3-month period,[46] which may be extended.

After approval, the CA will be the same for all QHINs. The amendment shall become effective on the effective date identified by the RCE during the amendment process and will be binding across all organizations. Any organization may terminate their participation in TEFCA within thirty days of the approved amendment if they are not willing to comply with the updated CA.

An exception to the amendment process described above is the need for an amendment to the CA that is required for the RCE to remain in compliance with applicable laws. In this case, the RCE is not required to provide QHINs the opportunity to vote on the amendment. The RCE is required to provide sixty days advanced written notice to the community unless that timeline would cause the RCE to be out of compliance with the applicable law.

The CA describes a similar process which is necessary for the SOP that will govern the implementation of a new Research exchange purpose. The primary difference from that

described above is the need for both the QHIN Caucus and the Participant/Sub-participant Caucus to approve the new SOP by at least two-thirds of the votes cast.

Future of the Research Exchange Purpose

OSTP released a "Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot" on October 28, 2022.[47] The RFI included questions regarding the inclusion of a Research exchange purpose for TEFCA. A summary of the responses was developed by the Institute for Defense Analyses' (IDA) Science and Technology Policy Institute,[48] with a companion document providing the individual RFI responses.[49] The Sequoia Project, the current TEFCA RCE, responded to the RFI in line with the process description above, indicating that they would work "on an expeditious, but deliberate pace".

Section II - Data Model for Research Purposes

This section focuses on a series of questions related to the feasibility of mapping the FHIR data model to a common data model (CDM) that is more commonly used in many Federal health research studies, the OMOP CDM. As leveraged by the *All of Us* Research Program, the OMOP CDM provides a standardized data structure that enables the systematic organization and harmonization of disparate healthcare data from various sources, such as EHRs, administrative claims, and registries. Developed by the Observational Health Data Sciences and Informatics (OHDSI) collaborative, the OMOP CDM promotes interoperability and facilitates large-scale observational research by transforming healthcare data into a consistent, unified format. Key aspects addressed in this section are:

1. The overlaps and gaps in the OMOP CDM and FHIR US Core data models.
2. Progress being made in standards and approaches to map FHIR to OMOP and the corresponding tools.
3. The limitations of the standards and approaches.
4. Potential advantages of having both models available to researchers and tools that could be used for FHIR native analytics.

Question 1 – Overlaps and Gaps in the OMOP and FHIR US Core Data Models

1. What are the overlaps and gaps in the OMOP CDM and FHIR US Core data models?

The OMOP CDM comprises a set of pre-defined tables, each representing a specific domain, such as demographics, conditions, procedures, and measurements, with well-defined relationships and standardized vocabularies for encoding clinical concepts. This structure allows researchers and healthcare institutions to conduct comparative effectiveness studies, safety surveillance, and other population-level analyses using a CDM, minimizing the challenges associated with data heterogeneity and enabling more efficient, collaborative research efforts. OMOP is used by major collaborative research studies such as *All of Us* and the National Covid Cohort Collaborative (N3C).

The US Core Profile, developed by the HL7 US Realm Steering Committee, is a set of FHIR profiles that cater to the unique requirements of the US healthcare system. By defining a common set of data elements and terminologies, the US Core Profile aims to support nationwide interoperability and facilitate the sharing of healthcare information across various healthcare organizations and systems. The US Core Profile is closely related to the USCDI, providing a FHIR-based framework to represent and exchange the common standardized health data classes used across the US healthcare system.

While FHIR and OMOP are both focused on standardizing how healthcare data is managed, they differ in several important ways as provided in Figure 6.

Area	FHIR	OMOP
Purpose	Primarily focused on enabling the interoperable exchange of healthcare information between different systems, providers, and applications and FHIR APIs are now incorporated natively into EHR systems	Designed to standardize and harmonize disparate healthcare data for large-scale observational research and population-level analyses
Scope	Covers a broad range of healthcare scenarios and workflows, including clinical, administrative, and financial aspects	Specifically tailored to support observational research and does not address the complete range of healthcare data domains
Structure	Organized into modular building blocks called "resources," which can be combined and extended to represent complex healthcare data scenarios. These resources can have complex nested structure and are often represented as JSON or XML	Essentially a set of tables (or 'domains') with specific schemas, well-defined relationships between tables, and a set of standardized vocabularies
Terminologies	Supports a variety of terminologies and code systems, depending on the specific implementation or profile	Uses a standardized vocabulary system that requires mapping source coding systems to a set of standardized vocabularies, ensuring consistency across all data sources
Data Exchange	Leverages modern web technologies, such as RESTful APIs, JSON, and XML, to support real-time data exchange and interoperability	Not designed for real-time data exchange but rather serves as a standardized data structure for the organization, storage, and analysis of healthcare data

Figure 6. Differences between FHIR and OMOP

The Vulcan FHIR Accelerator is focused on how FHIR can support the needs of clinical research and has a subproject specifically dedicated to supporting the development of FHIR to OMOP mappings. The project has attempted to align the US Core v4.1 Profiles and OMOP v5.4 domains. This alignment indicates a good degree of overlap at the domain/table level. Most of the US Core profiles (36 out of the 42 profiles considered) could reasonably be mapped onto one of the OMOP domains. Those that cannot relate to administrative aspects of patient care: *CareTeam*, *CarePlan*, *Goal*, *Provenance*, *RelatedPerson*, and *ServiceRequest*.

However, FHIR has 77 Base or Clinical resources and US Core has 26 Profile categories, whereas OMOP v5.4 has only 15 clinical data tables – reflecting the broader scope and greater flexibility of FHIR. As a result, there are often multiple FHIR profiles corresponding to a single OMOP domain, such as *Medication Request*/*Medication Profile*/*Immunization* and *drug_exposure;* or *Condition*/*Allergy Intolerance* and *condition_occurrence*. As discussed in the following section, this can lead to challenges in implementing a FHIR to OMOP mapping. The converse is more straightforward, there is a clearer one-to-one mapping from OMOP domains to FHIR resources.

The Vulcan alignment does not attempt to resolve at the field level or indicate how many of the required and optional fields in the FHIR profiles can be represented in the OMOP domain tables.

While the creation of such a field-by-field mapping is beyond the scope of this report, the following are examples of FHIR profiles that might be hard to map in their entirety:

- The FHIR *Immunization* Profile maps to the drug_exposure domain. However, Immunization contains several optional fields that do not have analogous columns in the OMOP table: *status*, *statusReason*, *site*, *reasonCode*, *subpotentReason*, *programEligibility*, *fundingSource*, and *protocolApplied.targetDisease.*
- The FHIR *AllergyIntolerance* Profile is supposed to map to *condition_occurrence*. However, there are actually significant challenges with and ambiguity about how to represent allergies in OMOP.

Finally, even the existence of a *theoretical* mapping from FHIR US Core resources and fields to OMOP tables and columns does not necessarily imply that an actual FHIR to OMOP mapping engine will fully capture the information available in a particular FHIR implementation. Differences in the implementation of FHIR at a given site might require modifications to the engine to ensure that the full range of information available at that site is captured. The challenge will be to identify whether this is the case. *All of Us* is in a unique position to be able to compare data quality and completeness metrics across sites after applying a particular FHIR to OMOP mapping implementation. This could help shed light on and address some of the issues outlined below concerning how the flexibility of FHIR complicates development of a standard mapping approach.

Question 2 – FHIR to OMOP Mapping; Standards, Approaches, and Tools

2. What progress is being made in standards and approaches to map FHIR to OMOP? What tools can be used to map FHIR data to OMOP?

There are no approved standards for mapping between FHIR and OMOP. The Vulcan FHIR to OMOP project has curated a list of projects that have attempted to create mappings between FHIR and OMOP. Figure 7 provides a curated list of these implementations and tools. The figure includes only those for which either code, mapping tables or a publication was available, as well as three additional implementations not included in the original list.

Mapping between FHIR and OMOP is not a symmetric problem. As previously noted, OMOP v5.4 has 16 clinical data tables, whereas FHIR has 77 Base or Clinical resources and US Core has 26 different profile categories. The implementations we examined that mapped from OMOP to FHIR, all take the approach of deciding which FHIR resources are analogous to a given OMOP domain and writing a one-to-one mapping between them. In contrast, constructing a single OMOP domain table from a set of FHIR resources may require combining information from multiple resources and implementing logic to reconcile them. Differing implementations of FHIR across sites could also mean that these resources are used in subtly different ways and so the

logic might need to be amended on a site-by-site basis. Mapping from FHIR to OMOP (the direction most relevant to *All of Us*) is therefore likely to be more challenging.

Project	Resource(s)	Direction	Mapping	Code	Active
NACHC's open-source implementation of FHIR to OMOP	github.com/NACHC-CAD/fhir-to-omop	FHIR → OMOP	No	Yes	Yes
GCP healthcare-data-harmonization	github.com/GoogleCloudPlatform/healthcare-data-harmonization	FHIR → OMOP	No	Yes	Last code update 2020
German Medical Informatics Initiative	https://github.com/OHDSI/ETL-German-FHIR-Core	FHIR → OMOP	Yes	Yes	Last code update Sep 2023
HL7 OMOP + FHIR Oncology Subgroup	docs.google.com/document/d/1RsIIvXO39DXoapGeMiGkWJPnZTUQMXK8BaflwNpAdmE/edit	FHIR → OMOP	Partial	No	Yes
Smile CDR / MUSC	ncbi.nlm.nih.gov/pmc/articles/PMC8243354/	FHIR → OMOP	No	No	Paper in 2021
OpenEHR to FHIR and OMOP Data Model for Microbiology	ebooks.iospress.nl/doi/10.3233/SHTI210189	FHIR → OMOP	Yes	No	Paper in 2021
Towards the Representation of Genomic Data in HL7 FHIR and OMOP CDM	ebooks.iospress.nl/doi/10.3233/SHTI210545	FHIR → OMOP	Partial	No	Paper in 2021
Justus-Liebig University	ebooks.iospress.nl/publication/54140	FHIR → OMOP	No	No	Paper in 2020
HIStream-Import (German Universities)	ebooks.iospress.nl/doi/10.3233/SHTI210053	FHIR → OMOP	No	No	Paper in 2021
CampFHIR	github.com/NCTraCSIDSci/camp-fhir	FHIR -> Relational	No	Yes	Yes
OMOP on FHIR Mappings (Georgia Tech)	github.com/gt-health/GT-FHIR	OMOP → FHIR	Partial	Yes	Yes
FHIR Ontop OHDSI Mappings (Guoqian Jiang)	github.com/BD2KOnFHIR/FHIROntopOHDSI	OMOP → FHIR	No	Yes	Last code update 2017
University of Colorado: MENDS-on-FHIR	github.com/CU-DBMI/mends-on-fhir	OMOP → FHIR	No	Yes	Yes
Simplifier OMOPinFHIR	github.com/FirelyTeam/OMOPinFHIR	OMOP → FHIR	No	Only for Person	Last code update 2021
MIRACUM	pubmed.ncbi.nlm.nih.gov/32570366/	OMOP → FHIR	No	No	Paper in 2020
FDA CDMH Mappings	build.fhir.org/ig/HL7/cdmh/	OMOP → FHIR	Yes	No	Last updated 2021

Figure 7. Curated list of mappings between FHIR and OMOP from the Vulcan FHIR Project

FHIR to OMOP Mapping Example

To understand this asymmetry through an example, consider how information about medications is captured differently in the two models.

- In OMOP, this information is stored in a single place, the *drug_exposure* domain table. This table is designed to capture each exposure of an individual to a particular drug, including the start and end date of that exposure and additional information, such as quantity and route of administration.
- In contrast, the medication domain in FHIR contains four distinct but related resources (i.e., *MedicationAdministration*, *MedicationRequest*, *MedicationDispense*, and *MedicationStatement*) that capture different aspects of the process of prescribing, receiving, and taking a medication. FHIR also has a separate Immunization resource.
- An implementer of a FHIR to OMOP mapping must therefore decide how to correctly combine information from these different medication resources. Appendix A presents a comparison of the only two FHIR to OMOP implementations which have open-source code, from Google Cloud Platform (GCP) and the National Association of Community Health Centers (NACHC).
    - The GCP code includes mappings for three of the medication resources (*MedicationStatement, MedicationRequest, and MedicationDispense*). These are treated as if they are independent and unrelated, so each resource will create a new row in the *drug_exposure* table. However, since these resources relate to different administrative parts of the process of prescribing and receiving medications, it seems likely that they would provide different information about the *same* medication and the same exposure for a patient. The *MedicationStatement* resource contains a field *derivedFrom* that references the Medication Request, Dispense, or Administration that was used to derive it. An ETL developer will likely need additional logic to deal with these interdependencies.
    - Furthermore, the GCP code does not consider the *MedicationAdministration* or *Immunization* resources. On the other hand, the NACHC code *only* includes the *MedicationRequest* resource in its mapping. In certain FHIR implementations this might be sufficient to capture all drug exposures. However, this is unlikely to be the case and could be misleading (e.g., if a drug is requested, but never dispensed).

The above discussion has considered only the *syntactic* mapping between the data models (i.e., how the tables/columns in OMOP and resources/fields in FHIR relate to each other). This mapping is not one-to-one, FHIR resources will map into different OMOP tables depending on context and capturing this context includes implementing both business logic and *semantic* mapping (i.e., mapping between different controlled medical vocabularies). This is because

OMOP expects data to be mapped into standard vocabularies such as RxNorm for medications or SNOMED for diagnoses and the table that a given FHIR resource ends up in will depend on the domain of the concept that it maps to. When mapping from FHIR to OMOP, there is no guarantee that source data is coded in any particular vocabulary; although, US Core does encourage the use of them. If the source data is coded in a common vocabulary, this may not be a problem as existing mappings available in the OHDSI Vocabulary Repository Athena could be used.[50] However, if the data is coded using non-standard or 'local' codes that are unique to the site, and a mapping between these local codes and a standardized vocabulary does not already exist, then significant work may be required to create such a mapping.

Finally, evaluating the quality of a particular mapping implementation can be challenging. Firstly, for any given implementation, it can be difficult or time-consuming to understand the underlying logic, since it requires digging into the details of code in a particular language (e.g., SQL vs. Whistle vs. Java) and the available documentation is of varying quality. More importantly, there is no generally available 'reference implementation' or set of test datasets against which to evaluate a particular implementation. Developing such a test set is challenging since real patient data on which to run such tests is sensitive. Consequently, research programs considering FHIR to OMOP mapping should design tests to evaluate the quality and completeness of the OMOP outputs and implement these against FHIR data coming from multiple different sources.

Question 3 – Limitations in the Standards and Approaches

3. What are the limitations of these standards and approaches?

There is currently a lack of high-quality mappings from FHIR to OMOP, in the sense of being complete, accurate, and production-ready (e.g., up to date, maintained, and well documented). The following outline some of the potential reasons why:

- **FHIR Flexibility and lack of generalizability.** The flexibility of FHIR and the variations in its implementation across different healthcare sites present challenges in developing a generic pipeline for data integration. Consider the issue of cardinality, which refers to the number of occurrences allowed for a particular data element. FHIR allows for flexibility in cardinality, meaning that certain data elements can have multiple instances, such as a list of phone numbers for a patient. However, this flexibility introduces complexity when attempting to create a standardized pipeline for data transformation and mapping. Other more complex examples of FHIR's flexibility were discussed in previous sections.

  Each implementation of FHIR to OMOP mappings has typically been developed by a team focusing on data from a single site, which is likely to limit that implementation's generalizability and applicability. As a multi-site program, *All of Us* can evaluate the heterogeneity of data retrieved from different sites' FHIR implementations and understand the implications for development of mapping approaches.

Implementing FHIR in compliance with the US Core profiles can help mitigate some of the challenges related to flexibility and variability across different sites. The US Core profiles provide a standardized set of resources and data elements that are widely used and accepted within the US healthcare system (i.e., USCDI availability was mandated by the 21st Century Cures Act Final Rule and EHR vendors often meet this requirement by providing US Core FHIR APIs). The US Core profiles define specific requirements for cardinality, data structures, and naming conventions, which can help address the challenges associated with developing a generic pipeline. Moreover, the US Core profiles promote the use of standardized terminologies, such as the Logical Observation Identifiers Names and Codes (LOINC®) and RxNorm,[51,52] further enhancing interoperability and facilitating data exchange. While compliance with US Core is not a guarantee of complete standardization across all aspects of FHIR implementation, it does provide a foundation for achieving greater consistency and compatibility, ultimately facilitating the development and maintenance of a more generic pipeline for data integration.

- **Inconsistency in the use of standard vocabularies and the need to map local codes.** As mentioned in the previous section, local codes at healthcare sites pose a significant challenge when transforming FHIR to OMOP. While standardized terminologies like LOINC for lab results and RxNorm for drugs are crucial for achieving interoperability, the adoption of these standards is not universal. In many cases, healthcare organizations rely on their own locally defined codes, which can lead to inconsistencies and difficulties in data exchange. The extent to which healthcare sites adhere to standard vocabularies can vary considerably, depending on factors such as organizational culture, resources, and technical capabilities. While US Core does encourage the use of standardized vocabularies, this does not mean that they are used in practice (particularly for historical information that might be valuable for research purposes). Consequently, implementing FHIR and OMOP at a specific site often requires substantial effort to map these local codes to the standardized vocabularies.

  This is not an issue that is specific to OMOP but is instead a general challenge with interoperability across multiple sites with different EHR implementations.

- **Sustainability.** The mappings between FHIR and OMOP listed previously in Figure 7 have been developed through open-source initiatives, volunteer efforts, and academic collaborations. While these initiatives have played a crucial role in advancing the interoperability between FHIR and OMOP, the sustainability and maintenance of these implementations may vary. Due to the voluntary nature of these projects, some implementations may not be actively maintained or may lag the latest updates in FHIR or OMOP standards, a significant challenge given the rapid evolution of FHIR and OMOP standards. Despite the growing interest in the interoperability between FHIR and OMOP, we found only one off-the-shelf commercial offering of FHIR to OMOP mapping.

- **Applicability of US Core for research.** So far, this paper has considered how OMOP relates to FHIR US Core. However, US Core contains a specific subset of FHIR Resources that may not be sufficient for research use cases focused on specific disease areas.

  There is a large list of other FHIR IGs, including examples such as the Minimal Common Oncology Data Elements (mCODE), an IG designed to facilitate exchange and re-use of data for patients with a cancer diagnosis. A subgroup of the HL7 FHIR-OMOP Initiative focused on oncology has created a draft IG for mapping mCODE to OMOP. However, the document cites open questions around how to conduct the mapping and does not have code available.

  Each therapeutic area may require its own IG to cover all elements required for deep research into a particular disease. However, a key challenge will be adoption of these by the EHR vendors and implementation across healthcare sites. Finally, it is likely that each FHIR IG will require either a new mapping to OMOP, or more likely, modifications to a consistent 'base' mapping based on US Core.

Question 4 – Advantages to Having Both Models; Tools for FHIR Native Analytics

4. Is there an advantage of having both models available to researchers? Are there tools available for FHIR native analytics?

As discussed in Section 2.1, FHIR and OMOP were developed for distinct purposes. OMOP is specifically designed to standardize observational studies for analytics, whereas FHIR was developed as a transport or messaging standard. Implementations of FHIR APIs are therefore generally designed for synchronous queries to retrieve information in real-time for a single person and retrieval of large amounts of data on many individuals simultaneously, required for the kind of analytical studies relevant to *All of Us*, an approach known as "bulk FHIR," is less mature.

The OMOP standard stores data in the traditional way that analytical users expect – flat tables with a simple relational schema linking them. Tools developed by the OHDSI community commonly assume that the data is stored in a relational SQL database. In contrast, FHIR resources are more complex nested structures that are not straightforward to store in tables and relational databases. This complicates doing analytics on data in FHIR.

One approach that has been suggested is to create a 'flattened' tabular version of a given FHIR resource. The SQL on FHIR project calls this a tabular 'view' and has defined something called a *ViewDefinition* that specifies columns to be included in the output table via *FHIRPath* expressions that pull out specific fields. However, this needs to be done on a case-by-case basis and does not guarantee that the set of tables that result are optimized for analytics. Defining the right set of 'views' to support your observational study, that can apply across multiple sites or data sources, is essentially what the OMOP data model has tried to solve for.

The following provides examples of open-source and commercially available FHIR-native analytical tooling:

- **HAPI FHIR** is an open-source Java-based FHIR server that provides support for FHIR analytics. It includes features like search, filtering, and aggregation of FHIR resources. HAPI FHIR has built-in support for the FHIR JSON and XML encoding formats. A built-in parser can be used to convert HAPI FHIR Java objects into a serialized form, and to parse serialized data into Java objects.
- **Smile Digital Health** provides the Smile Health Data Fabric (HDF) solution with enterprise-level FHIR data interoperability and data exchange capabilities, backed by the set of core capabilities within their Clinical Data Repository (CDR). Smile incorporates the incorporates the HAPI FHIR engine. The HDF enables prospective and retrospective analytics on quality measures, gap in care, clinical decision support, and population using FHIR-formatted information that flows into the Smile HDF.53.
- **Firely** offers a range of FHIR-related products, including Forge (a FHIR profiling and validation tool) and Firely Server (formerly known as Vonk, a FHIR server with built-in analytics capabilities). The Firely Server Facade is a means to use the Firely Server implementation of the FHIR RESTful API on top of an existing repository. This repository may be a relational database, a NoSQL database, or another web API. All in all, this solution is not walk-up usable and would require a data engineer to provide analytical support on top of the Façade server once it's deployed.
- **Cloud vendors** such as AWS, Microsoft, and GCP advertise solutions for making FHIR data available in analytics-ready formats that presumably leverage the idea of FHIR view-like transformations involving flattening FHIR Resources and fields into tables.

FHIR-native analytical tools are designed to simplify the analysis of FHIR data by providing native support for FHIR resources and operations. The FHIR-native analytical tools noted above are some of the products available on the market that can directly work with FHIR data in a flattened format without the need for translation to other formats like OMOP. However, such an approach may be challenging, particularly in a multi-site research setting. First, it might be difficult for a researcher unfamiliar with FHIR to inspect data in the form of FHIR resources, with a complex nested JSON/XML format, and be able to design the flattened version that includes the information that they need. Second, if you have FHIR data coming from multiple sources, it could be difficult to create a harmonized flattened version that is consistent across sites. This could be explored further through a pilot project as described in Section IV.

One possible hybrid approach would be to develop a generalized 'base' mapping pipeline that maps FHIR resources to the core OMOP tables and has been validated to capture some baseline percentage of the available data, but then also give end users the ability to augment this core dataset by creating dynamic 'views' from the original FHIR resources that pull out additional information not captured in the base mapping.

Section III – FHIR Technology Questions for Research

This section focuses on the state of FHIR technology readiness in terms of interoperability, implementation, and integration in support of Federal health research studies, such the *All of Us* Research Program. Specific areas covered include:

1. The state of readiness for FHIR exchange at healthcare organizations in the US today and one year from now.
2. The advantages and disadvantages of push vs. pull methods for FHIR exchange initiation.
3. The advantages and disadvantages of FHIR REST (aka individual) vs. FHIR Bulk Data (aka population) for FHIR exchange for a Federal health research study.
4. The ability of SMART on FHIR - Backend Services specification and FHIR security mechanism to meet Federal health research study program security and privacy requirements.

Question 1 – FHIR Exchange State of Readiness Now and in One Year

1. What is the state of readiness for FHIR exchange at healthcare organizations in the US today? One year from now?

The State of Readiness of FHIR Now

Both the ONC and CMS have worked diligently to guide the healthcare community to standardize on data elements through their work with USCDI (v1, and more recently with v3), and many other US Federal Agencies have been pushing to expand the USCDI to include "plus" elements. With each release of USCDI more elements are being added. Expanded work is being undertaken with Data Provenance, Identity, and Consent to name just a few.

Once the community has harmonized on these data elements, the FHIR Standards and Specification community work in collaboration to define the technical details of implementation through the FHIR IGs. While the FHIR IGs do not currently have a maturity model like the FHIR Resources and Profiles, an IG's maturity can be determined by reviewing the IG's history, and proposed changes to the version of the IG currently under development (in the CI Build process). One FHIR IG, the US Core IG, has been under development for more than seven (7) years, and is currently undergoing development of v7.0.0. One method used to review the maturity of an IG is to review any published "Change Log"[54] – the US Core change log demonstrates the maturity of its overall authoring and publishing process that would serve as a good model for other IG authors to replicate or follow.

ONC has further guided the healthcare community to standardize on its FHIR and US Core IG API through the ONC EHR Certification Program inclusion of §170.315(g)(10) Standardized API for patient and population services (and soon to include HTI-1). This effort has seen some early success and has affirmed for both the Federal Government and commercial implementers that FHIR is a step in the right direction.

Implementers are encouraged to maintain a "trust but verify" approach when evaluating

potential vendors' technologies and implementation strategies. The complexities of implementations frequently present themselves when integrating with legacy systems and authoritative data sources. These validation and verification activities should also be extended externally to any potential information exchange partners to ensure the highest quality data exchange possible.

In addition, the implementation of technology standards like FHIR and specifications like US Core remain open to interpretation. To mitigate the potential risk for each party (i.e., receiver and sender), use case(s) should also advance rigorous software testing methods. The community would be well served to ensure they are implementing FHIR in a manner which is consistent with being compliant and conformant to the regulatory requirement defined by USCDI, and the related FHIR IG US Core.

The State of Readiness of FHIR in One Year and Beyond

While FHIR has focused on many newer business and use cases and provides open access to easy (easier) to implement information exchange, it has seen a significant uptick in interest in adoption over the past several years. It has not, however, been positioned as a wholesale replacement to the immense legacy infrastructure currently in place across the US healthcare industry. Other standards and technologies such as HL7 v2, IHE, X12, and even the Clinical Document Architecture (CDA) will continue to play a role for years to come.

Over the next several years with continued support and sponsorship from ONC, CMS, HL7 and other Federal Agencies (such as CDC, FDA, etc.), one can anticipate FHIR will continue to advance toward "wider industry" adoption. The US Healthcare Community and Industry "FHIR flywheel" has significant momentum, which will undoubtedly continue moving forward.

While the implementation community may continue to face a significant number of unknowns and potential risks implementing FHIR, the maturity of FHIR as a Standard, the FHIR IGs, and the advancement in FHIR Business and Use cases currently does not show any signs of slowing down. A couple of absolute knowns: first, *there will be change*. Changes in FHIR IG versions, Security, FHIR Operations, and eventually even with FHIR itself (with R4B, R5 and/or R6). Another well-established known is that across the US healthcare community, integrated ecosystem organizations do not move (update/upgrade) in *lockstep*. Ensuring future releases seamlessly support and preserve backward compatibility is an essential part of the standards development process.

With ONCs release of HTI-1, which mandates starting January 1, 2026, all HTI-1 updated standards in certification criteria be in place. Included within HTI-1 is a call for USCDI v3 and US Core v6.1.0. While the implementation community could begin a process to transition toward USCDI v3, many of the other complimentary FHIR IGs supporting CMS (e.g., CARIN Alliance and the Da Vinci Project) and other Federal Agencies remain based on the current US Core v3.1.1. The *All of Us* Research Program and industry should monitor this closely over the next eighteen (18) months,

to plan for an approach to manage the information exchange during transition from US Core v3.1.1 to US Core v6.1.0.

This will present an operational and production risk during any industry transition window manage change, as the legacy version is being "sunsetted" and the newer approved version is gaining adoption. This could become increasingly concerning should the two versions be either incompatible or non-interoperable. Industry organization SEMVER seeks to guide the standards and implementation community with proven techniques to avoid version disruption.[55] SEMVER uses a simple set of rules and requirements that dictate how version numbers are assigned and incremented. That, combined with Postel's Law: Designing for Robustness, which states "Be conservative in what you send, be liberal in what you accept," helps navigate version interoperability. However, the debate continues about who bears the responsibility between the parties to have the necessary "version-awareness" to safeguard the exchange or information sharing.

Question 2 – Advantages and Disadvantage of Push vs. Pull Methods for FHIR Exchange

2. What are the advantages and disadvantages of push vs. pull methods for FHIR exchange initiation on a Federal health research study?

Which method to use for FHIR exchange is an architectural design decision that is best made based on the requirements for the data exchange. Theoretically, push and pull are both valid approaches to initiating data transfer, with push providing the control to the data source, and pull providing the control to the data recipient. The advantages and disadvantages for each depends on the situation.[56]

All of Us Needs

The *All of Us* Research Program, as an example Federal health research study, is currently operating with a push model where HPOs extract data into OMOP tables and send it to the *All of Us* DRC.[57] For this workflow, the HPOs (data source) must have knowledge of the *All of Us* participants, their program status, and what data to send.

For this use case, the data source has access to the necessary details:

1. The data source knows the full set of data for a set of participants, may keep track of when things were updated, and can keep track of when information has been sent to *All of Us*.
2. The data source knows when information is added or updated.
3. The timing can also be agreed upon ahead of time: data needs to flow to *All of Us* within some period after new or updated information for an in-scope participant is made available. The specifics of this period can be agreed upon during implementation.

Therefore, from a theoretical perspective a FHIR push model will support the most efficient data transfer for the *All of Us* Research Program's needs. The advantages and disadvantages of a FHIR push model for the *All of Us* Research Program are provided in Figure 8.

Advantages	Disadvantages
• **Real-Time Data Exchange:** Data can be transmitted as soon as it is captured. • **Event-Driven:** In a publish/subscribe model, data can be transmitted based on specific events or triggers such as when an encounter is closed. • **Reduced Burden:** *No* need to continuously poll for data; functionality and maintenance of sending data is on the sending sites.	• **Availability:** Receiving endpoint needs to be available; if down, data could be missed; typically, a robust system will include mitigations such as retrying to send data if the endpoint is down. • **Data Control:** Data sender is in control of the data; need to work with the sender to determine format, data set and frequency of transmission. Additionally, if data requirements change, then need to coordinate those changes with each of the sites and get the updated information. • **Record Management:** When receiving data from multiple locations for a single participant, there will need to be some form of participant identification and merging of data into a single record for an individual.

Figure 8. Advantages and Disadvantage of a FHIR Push Model from an All of Us Perspective

If *All of Us* needed the ability to have control over the specific data scope or wanted to gain access to data for a certain participant only for certain conditions outside the source EHR, then a pull-based approach would be advantageous. This is theoretically the appropriate approach because the relevant details on scope and/or timing would live on the data recipient side. Hybrid models are possible, such as where the data source pushes notifications that relevant updates have been made and data recipients pull the relevant data scope based on what it needs.

For *All of Us*, advantages and disadvantages of a FHIR pull model are provided in Figure 9.

Advantages	Disadvantages
• **Availability:** *Can* decide when to request data; less chance of missing data due to receiving endpoint being down and receiving endpoint does not need to be "always on." • **Data Control:** More control to request data when needed and what specific data to receive (within the limits of the query	• **Connection Management:** Need to make requests to many different systems and manage all these connections. • **Access Control:** Need to manage access control for each system which will have separate authorization mechanism/credentials. Also, if data requirements change the access control at the sites may also need to change. For example, if new resource types are needed that do not currently have access granted, those scopes will need to be granted to the application at each site. • **Participant and Site Lists:** Need to maintain a list of all sites and participants at each site from which they are requesting data. The identifier of each participant will also need to be known.

Advantages	Disadvantages
interface and what can be requested through that interface).	• **Fetching Data:** Need software to handle requesting and processing data.

Figure 9. Advantages and Disadvantage of a FHIR Pull Model from an All of Us Perspective

Theoretically, for this *All of Us* use case, a push-based model will be advantageous over a pull-based model. Push can be significantly more efficient as it does not rely on *All of Us* continually initiating hundreds of queries per patient within the cohort to every known endpoint. However, for the push-based model to be effective, several related industry standards and FHIR technologies need to be more widely adopted across the provider EHR communities. However, the current state of FHIR means that the only potential model available in the near term will be a pull-based model.

Existing Non-FHIR Implementations

The theoretical result of a push model being the most advantageous is supported by similar use cases that use push today. State-wide and regional HIEs that collect data from hospitals and clinics get data using the push model with Consolidated Clinical Document Architecture (CCDA) documents. While CCDA-based exchange supports a pull model, and some vendors early on in adoption of CCDA only supported pull (e.g., Epic originally only supported a pull but now supports push), a push model is now almost universally used for data feeds into HIEs. Cross-Community Document Reliable Interchange (XCDR) and Document Metadata Subscription (DSUB) are two IHE methods currently in use supporting the CCDA push model.

The State of FHIR

Practically, the chosen approach must depend on the functionality available. Unfortunately, the process of FHIR standardization and adoption does not yet support either push or pull in a robust enough way for it to be a recommended approach for research currently.

FHIR is being widely adopted by many EHR systems. However, accessing data from EHRs for research purposes presents many challenges. Some of which include what data elements are accessible via FHIR, the version of FHIR, and availability of FHIR services across EHR vendors and between institutions.[58] Furthermore, inconsistent data standards between clinical care and clinical research further complicates the sharing of data for research purposes.[59] It was also concluded in a scoping review that FHIR specifications for research are not mature and there are very few operational FHIR-based research projects today.[60]

Push-Based Data Acquisition

Within the FHIR ecosystem, support for push-based models is not yet robust. While the practical foundation for passing a bundle of FHIR data over a RESTful FHIR API exists, it is not required

functionality for EHRs to implement in the current regulations. In addition, the push-based model is limited in the size of the payload it can support, which is a significant limitation for large-scale data feeds. A bulk data push operation that can handle larger-scale data transfers has been proposed but does not yet exist. The lack of agreed-upon standards for a bulk push option means that wide-spread support in FHIR is likely at least five years away.

Pull-Based Data Acquisition

Pull-based models for both patient-level and large-scale data acquisition on FHIR are more mature, but not enough so to offset the challenges that come from using a pull-based model when the data source knows best when and what to send. There will be two major challenges: (1) identifying when new data is available and running a query, and (2) identifying the scope of data to request and making the request.

First, to identify when new data is available and to run a query, none of the available options are robust and widely implemented. There is an existing standard for FHIR-based subscriptions that supports notification when updates have been made, but it is not required by regulation and is not widely implemented. Older notification approaches (e.g., based on HL7v2) could be used in the short term. Further, no standard FHIR query exists to check if a patient has new data available, so data sources would need to be asked to export data periodically even if no new information is available. This would result in extra requests in the best case and extra data transfer in the worst case.

Secondly, for the client to identify the scope of data to request and to make the request, none of the existing options will work well. Single-patient FHIR-based queries can be used but would result in multiple queries per patient to gather data for each data type. Data sources are not required to support filtering by when data was updated. This means that excess data could be transferred, leading to significant inefficiencies. FHIR servers may, but are not required to, apply status-based filtering to remove inactive instances and a solution using this approach would need to be careful to not miss changes where previously sent instances are marked as inactive.

In addition, the ability for FHIR servers to handle searches for multiple patients lacks consistency and standardized specifications. Therefore, the approach for multi-patient FHIR based queries may vary across EHR vendors. Some data sources may support the export operation only on groups of patients that they define, so *All of Us* would be dependent on the data sources to maintain that patient list. In some instances, bulk FHIR may be used to achieve multi-patient query capability. The same considerations around filtering by updated date and the potential for additional status-based filtering are present here as well.

Question 3 – Advantages and Disadvantages of FHIR REST vs. FHIR Bulk Data for FHIR Exchange

3. What are the advantages and disadvantages of FHIR REST (aka individual) vs. FHIR Bulk Data (aka population) for FHIR exchange for a Federal health research study?

Within FHIR, there are two primary methods for querying data: Individual queries and Bulk queries. FHIR Individual Queries are used to access specific data for an individual patient or a small number of resources. Individual queries follow a RESTful approach, utilizing standard HTTP methods like GET, POST, PUT, and DELETE. They are designed for real-time transactional interactions, typically involving targeted data retrieval or updates. For instance, a user might request a single patient's record or a specific observation for that patient using FHIR's search functionality, which supports filtering by various criteria (e.g., patient ID, date range). The responses are immediate and specific to the query, conforming to the FHIR JSON or XML format.

FHIR Bulk Queries were introduced in FHIR Release 4, Bulk Data Access (a.k.a. "Flat FHIR"). Bulk queries are designed for scenarios where large amounts of data need to be extracted from the FHIR server, for which it is impractical to perform with individual queries. This approach is useful for data analytics, population health management, and other cases where entire datasets are needed. Bulk queries use the FHIR Asynchronous Request pattern, where the client initiates a request, and the server processes it in the background, providing a URL from which the resulting data set can be downloaded once ready. The data is typically provided in a series of files in NDJSON (Newline Delimited JSON) format, which are suitable for handling large quantities of FHIR resources.

The decision on whether to select a FHIR REST (aka individual) vs. FHIR Bulk Data (aka population) approach for FHIR exchange might not be using one solution over the other, but implementing support for both will depend on several factors. These factors include specification support (e.g., versioning), EHR Implementation & Market Deployment, Search Capabilities, Performance & Scalability, and Testing & Certification. Figure 10 provides some of the key considerations for these factors. A more detailed listing can be found in Appendix B.

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Specification Support	
There are several FHIR versions in operational use (DTSU2, STU3, R4) with R5 on the horizon. Client applications need to be aware of and handle these various capabilities at a national level.	Maturing specification with varying adoption between EHR vendors and their respective deployments. FHIR Bulk Data Access (Flat FHIR) was published as a Standard for Trial Use (STU) on 11-26-2021.
EHR Implementation & Market Deployment	
FHIR servers have widely adopted support for the individual query.	FHIR servers have varying support for Bulk Export.

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Search Capabilities	
• Individual queries offer a wide range of search parameters that can be used to filter, sort, and narrow down results, enabling complex queries that are focused on FHIR resources or sets of resources. • Individual queries target specific resource endpoints (e.g., Patient, Observation). This precision allows client applications to request exactly what is needed without extraneous data. • Requestors can specify response formats like JSON or XML, giving flexibility based on application requirements. • There are no capabilities at a national level to locate where an individual patient's EHI is located. Broadcast requests within a region or nationally would be needed at this time. The VA utilizes broadcast (IHE XCPD) on a nightly basis for all patients that have an appointment scheduled for the coming day to locate a patient's record. • Consent management needs further adoption through SMART, UDAP and TEFCA as implementation of sharing consents appears to be exceedingly low.	• Bulk has been designed to retrieve large amounts of data for multiple patients; although, the granularity of parameters for filtering in bulk queries might be limited based on the server's implementation. • Patients would need to be pre-identified with each healthcare organization (FHIR endpoint) beforehand. Particular EHR vendor implementations place limits on the number of patients included in a group. • Large response payloads might require special handling, pagination, or streaming mechanisms to process efficiently. • Responses are in Newline Delimited JSON (NDJSON) format, making it more manageable and efficient for streaming and parsing large datasets. • There are no capabilities at a national level to locate where an individual patient's EHI is located. Broadcast requests within a region or nationally would be needed at this time. The VA utilizes broadcast (IHE XCPD) on a nightly basis for all patients that have an appointment scheduled for the coming day to locate a patient's record. • Consent management for bulk requests needs to be resolved through pre-arrangement or updates to the specification.
Performance & Scalability	
• Individual queries can provide immediate responses to client requests; however, ONC and the Interoperability Networks have not defined Service Level Objectives (SLO) or Service Level Agreements (SLA) indicating the FHIR endpoint availability or response times related to individual queries. • Other considerations include lower resource utilization; although fetching large datasets accessing multiple patients across many FHIR endpoints requires multiple sequential or parallel requests can be inefficient.	• A single bulk request can retrieve large amounts of data, reducing the connection overhead seen with multiple individual queries. • Bulk queries can be processed asynchronously, allowing systems to make a request and retrieve results later, freeing up resources for other operations in the meantime; although, checkpoint recovery is lacking in the specification now, so failures may not be easily determined along with restarts at the failure point. • Large datasets require significant computational power, storage, and bandwidth to process and can put a strain on server resources. Transferring huge datasets can consume significant bandwidth and may result in longer response times. • Due to the potential for longer response times, bulk queries may not be ideal for scenarios requiring immediate data retrieval.

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Testing & Certification	
The ONC Health IT Certification Program provides a conformance testing framework along a set of conformance test tools.[61] These tools support a variety of capabilities, including Bulk, HL7 Da Vinci, and FAST IGs to allow organizations to test both FHIR client and servers.	FHIR Bulk Data is a maturing specification that has been tested at several FHIR Connectathons for various use cases.

Figure 10. Factors for assessing FHIR Rest vs. FHIR Bulk Data for the exchange of healthcare data.

Overall, the advantages of FHIR Individual Queries include real-time data access, high specificity, flexibility in search parameters, and suitability for small-scale data requests. However, they can be inefficient for large data sets and strain server resources with multiple requests.

FHIR Bulk Queries excel in handling large data sets efficiently, are ideal for data analytics and population health tasks, and reduce server load by processing requests asynchronously. The downside is the lack of real-time data access, potential complexity in handling bulk data formats, and the need for additional infrastructure to manage asynchronous requests and data retrieval.

The use of FHIR REST (aka individual) provides the most flexibility and maturity. It enables a client application to identify FHIR Resources based on specific search criteria to be retrieved. This would allow the system to pull a baseline for a patient from a FHIR endpoint and later request only new or updated information. The FHIR Bulk Data (aka population) approach is a still maturing specification. It can request a group of pre-identified patients to be exported along with their associated FHIR Resources.

Question 4 – SMART on FHIR

4. Will the SMART on FHIR - Backend Services specification and FHIR security mechanism meet Federal health research study program security and privacy requirements?

The FHIR Standard and Specification continues to invest in a common security model which can be implemented across the entire FHIR implementation community. The SMART initiative dates to early 2010[62] with Harvard Medical School and Boston Children's Hospital jointly advancing a web standard for API transport, authorization, and user interface, and standard medical terminologies for coded data. In 2013, they updated SMART to take advantage of the clinical data models and the API described in HL7 draft standard FHIR.

SMART on FHIR provides a standard, universal security layer (OAuth2) API for accessing Electronic Health Information (EHI). The end goal is to ensure patients can access their EHI securely from any app and device of their choosing.

SMART Standard for Trial Use (STU) Version 2 (SMART App Launch v1 IG) was published November 2021, and improved the scope definition for more granular permissions. To name a few of the enhancements, this publication introduced improved security requirements with Proof Key Code Exchange (PKCE), and profile token introspection. SMART App Launch and oAuth2 are required elements of the current ONC EHR Certification §170.315(g)(10). ONC HTI-1 introduces SMART App Launch IG Release 2.0.0 (SMARTv2 Guide) as the next major release of the SMART App Launch IG that will be enforced by the EHR Certification Program.

SMART with its well-known query capability has quickly seen wide-industry adoption, and many organizations participating at the HL7 Connectathon have continued to advance the maturity and best-practices of implementing SMART App Launch, SMART Security including with Bulk FHIR operations.

ONC FHIR At Scale Taskforce (*FAST*) and now an HL7 Accelerator *FAST* have been advancing an alternative Security for Scalable Registration, Authentication and Authorization model or UDAP for short.[63] ONC under TEFCA along with the RCE have identified UDAP as the required security model. It is not known how SMART and UDAP implementations will be coordinated. Currently, no regulatory requirement exists which have mandated UDAP adoption or implementation (including the recent ONC HTI-1 or CMS Burden Reduction Prior Auth). Both the FAST Security IG, and the RCE TEFCA IG have participated in HL7 FHIR Connectathons. Adoption of UDAP remains low now at the time of this publication. Unlike the SMART IG, the ability to define "scope" remains to be incorporated into the HL7 FAST Security IG - UDAP.

Compliance with the Federal Information Security Management Act (FISMA), HIPAA, and other NIH security and privacy guidelines will depend on the implementation of these standards within the agency and any healthcare IT system that NIH exchanges EHI. Compliance will involve having a robust information security program, conducting regular risk assessments, implementing appropriate security measures, and undergoing periodic security audits. It should be noted that the underlying SMART on FHIR standards OAuth 2.0 and OpenID are widely used open standards for authorization and authentication, respectively. However, it is important to clarify that these standards themselves are not inherently "FISMA compliant" or "non-compliant." FISMA is a US Federal law that mandates a program to protect government information, operations, and assets against natural or man-made threats. FISMA compliance is about how an organization manages its information security systems and processes, rather than about specific technologies or standards. A complete risk assessment of the intended NIH implementation will need to be performed to ensure that these capabilities meet Federal security and privacy guidelines.

Section IV – Next Steps/Pilots

Given the information provided in the prior sections, the final set of questions focuses on how the *All of Us* Research Program and other agencies engaged in Federal health research studies can best move forward with implementing FHIR. Specific areas covered include:

- The next steps Federal health research studies should take to get ready for a FHIR future.
- Potential pilots that Federal health research studies could conduct to advance the knowledge, understanding, and experience with FHIR and at the same time demonstrate the validity of the answers provided by in this paper.
- Any additional strengths, weaknesses, opportunities, and threats with respect to FHIR for Federal health research studies

Question 1 – Next Steps to Get Ready for FHIR

1. What are the next steps Federal health research studies should take to get ready for a FHIR future?

As a result of the information captured in the prior sections, there are several recommended next steps for the *All of Us* Research Program and other Federal agencies to consider. These next steps include suggested actions that relate to both policy and technology considerations.

Policy-Related Recommended Next Steps:

a. Further investigate if a study participant-facing "consumer app" can be created that will enable study participants to pull their EHR data via the IAS exchange purpose with intent to enable those records to be transferred to study organization after the end user receives them. The Consumer App creator will need to be a TEFCA Participant that offers IAS service, such as Health Gorilla (which does have an existing patient-facing application). That organization will need to structure their service offering in such a way as to ensure study participants can transmit their EHR data to a research organization after the records have been received. This is discussed further in the Pilots (Question 2) section.

b. Conduct a survey with providers to determine the proportion of providers who are QHIN participants, and which of those can support and implement IAS.

c. Engage a multistakeholder community as recommended by the TEFCA RCE (the Sequoia Project), potentially in collaboration with ARPA-H efforts, to begin the task of amending the CA to include the Research exchange purpose. Since it is known that the basic process will likely take between 16 months and more likely longer than two years, the earlier this effort is undertaken, the sooner the capability will be available.

Technology-Related Recommended Next Steps:

a. The research community would be well served to conduct a detailed analysis and review of the current USCDI v3 (which is required by the ONC HTI-1 final rule to be implemented

by January 2026) defined healthcare elements with a critical eye toward data deemed important for future Research Programs, cohorts, and prospective studies. Ensuring the Research Community participates in future ONC USCDI v4 and v5 element review and comment opportunities, and with HL7 Accelerators, such as Argonaut, as they advance US Core v7.0.0, which intend to reflect later versions of USCDI. It would be invaluable to both the standards community, and future Researchers to ensure "future proofing" of these documents and standards.

b. As the research community establishes future quality guidelines and requirements for participant organizations' information exchange using FHIR, there should be a strong program understanding of the common FHIR constructs provided above will offer to significantly improve overall data quality of the Program, and ensure the data provides years of reuse across a multitude of research studies. By ensuring participants organizations have and maintain a conformant capabilities statement, the *All of Us Research Program* will be able to maintain consistent data exchange.

   Given that most EHRs do not always run FHIR validators on outbound data on their FHIR servers, the research community will need to assess data quality for inbound FHIR data for each bespoke research project, by measuring data quality and profile conformance against FHIR IGs relevant to their research project.

c. The research community may wish to evaluate their approach, and architecture to ensure they are able to manage transition between versions of FHIR, FHIR IGs, Profiles and Regulatory Standards. This could be incorporated into future pilot efforts as a baseline "requirement."

d. The research community will benefit from investment in format-agnostic ETL tools that allow for field-level mapping between different formats (e.g., OMOP, PCORnet, and others) and specific FHIR profiles. A full market analysis of available ETL tools would be an appropriate next step to identify options, as well as gaps in those options.

e. The All of Us Research Program and other research programs should continue to develop requirements associated with cohort scopes, identity, and any future patient consent requirements. These requirements could be tested during future All of Us Research Program Pilot efforts.

f. Conduct a series of pilots to test the implementation of FHIR profiles and IGs that are useful for research purposes and test the quality of the data that is returned against those FHIR profiles. This is discussed further in the next subsection, Pilots.

Question 2 – Pilots

2. What are the two or three pilots Federal health research studies could conduct to advance the knowledge, understanding, and experience with FHIR and at the same time demonstrate the validity of the answers provided by ACT-IAC?

Beyond the next steps detailed above, several pilots are recommended for consideration by the *All of Us Research Program* to support the findings in this paper and advance the knowledge, understanding, and experience with FHIR:

a. **Conduct a pilot with a QHIN that has launched IAS support** such as Health Gorilla, Epic, or eHealth Exchange. The *All of Us Research Program* can leverage an existing IAS implementer to forward data from consenting participants enrolled in the study, utilizing a study participant-facing application provided by the IAS implementer. This pilot is most relevant to research programs that have a consented patient population that can authorize the retrieval of EHR data prospectively and retrospectively using identifiable querying of QHINs. The participant application should minimize participant burden while also protecting against data compromise.

   i. Expected outcomes: intended outcomes would be to test IAS record retrieval and data quality assessment for at least 100 individuals, across one to two IAS providers using a QHIN's existing IAS application.

   ii. Note this would result in ingesting CCDA from the TEFCA QHIN participants and converting to FHIR using established CCDA to FHIR mapping developed with QHINs (eHx and Health Gorilla) at the QHIN hub, and then transferring FHIR bundle data to *All of Us*.

   iii. Raw CCDA can also be ingested by *All of Us* to compare the completeness and quality of CCDA data and the FHIR-converted data.

   iv. Based on this, a large language model (LLM)-assisted (and human-reviewed) mapping can also be constructed between the FHIR data that results from IAS, and an extended-version of OMOP. If the LLM is used, it would be prudent to leverage retrieval-augmented generation or similar method to prevent LLM hallucinations. It is also important to note that an LLM-assisted method would only provide templates or a starting point for clinical data experts to validate, thus accelerating the mapping effort.

   v. Compare coverage to SMART on FHIR patient mediated method of data aggregation.

b. **Conduct a pilot for facilitated FHIR exchange across TEFCA QHINs who are ready to exchange FHIR data directly** (as opposed to converting CCDA to FHIR), utilizing the TEFCA Facilitated FHIR IG with 5-10 healthcare organizations who are members of a QHIN and ready to exchange FHIR.[64] This pilot is most relevant to research programs that have a consented patient population that can authorize the retrieval of EHR data prospectively and retrospectively using identifiable querying of QHINs.

   i. Expected outcomes: intended outcomes would be to test FHIR record retrieval for at least 100 individuals, across one to two QHIN sub-participants providers using existing FHIR APIs, and assess whether they meet or exceed data elements in USCDI V1.

c. **Test the implementation of FHIR IGs across a research network of 5-10 data providers**, ideally spread across 2-3 different EHR vendors to assess whether data can be harmonized when implementing FHIR IGs across different EHR vendors. This would quantify both intra-EHR and Inter-EHR variability. This pilot would be most relevant to research studies that are collecting bulk data for observational research on an unconsented de-identified population (i.e., not a prospective study).

   i. Expected outcomes: intended outcomes would be to test the FHIR IGs below across at least two healthcare providers across two different EHR vendors (e.g., Cerner and Epic). The intent is to measure the quality of outbound data from EHR's FHIR APIs against a FHIR validator, for each of the IGs listed below, for IG-specific profile-level conformance. We would also test conformance to expected value sets expressed in standardized vocabularies and terminologies and quantify how often local non-standard codes are used in outputted FHIR data.

   ii. First, it is key to evaluate the FHIR US Core v3.1.1 (which is aligned to USCDI V1) and conduct a full gap analysis to determine if data elements captured by a sample across EHR vendors who have implemented US Core v3.1.1 are comparable in completeness and quality.

      (1) Of the gaps identified in the above analysis, it would be important to differentiate which of those gaps can be addressed by USCDI+ and USCDI V3 (which will be required for EHR certification by January 2026).

      (2) This analysis would be repeated, when USCDI V3 is more widely adopted, to evaluate whether EHR vendor endpoints expose elements that meet USCDI V3.

   iii. Test the implementation of the MedMorph IG.[65]

   iv. Test the implementation of the Vulcan Real World Data IG.[66]

v. CodeX for capturing oncological and/or cardiological data.

(1) mCODE IG under the CodeX accelerator can be tested for standardized capture of oncological data.[67]

vi. FHIR FAST testing of the IG for Patient Identity Matching.[68]

d. **Conduct a pilot to develop an implementation-agnostic field-level mapping from US Core to OMOP, build an open-source implementation of this mapping, and test it across multiple real-life FHIR data sources**. This pilot would be relevant to any research studies that are collecting FHIR data for observational research, and wish to convert their FHIR data to OMOP, like the *All of Us Research Program*. The expected outcomes of this pilot would include:

i. Developing an implementation agnostic many-to-many field-level mapping of FHIR US Core fields to OMOP tables including business logic and pseudo-code for combining or transforming data within and between resources.

ii. Designing a set of tests for evaluating the quality of a mapping when applied to a particular FHIR data source (e.g., to understand the completeness of the FHIR data captured in the OMOP output). These tests will measure whether mappings result in fundamentally unabstracted data from the source to target mapping, and if not, clearly show provenance and abstraction steps taken.

iii. Using the mapping to develop an open-source pipeline that can be run against specific FHIR data sources.

iv. Run this mapping pipeline against a set of ~2-3 different sources of FHIR data and evaluating the test metrics. The goals are to better understand: 1) how heterogeneity of FHIR sources affects mapping to OMOP, 2) the reusability of a given implementation, 3) quantify how often local non-standard codes are used in outputted FHIR data, and, 4) assess the feasibility of mapping those non-standard codes to standardized vocabularies using semi-automated human-in-the-loop tools.

Question 3 – FHIR Strengths, Weaknesses, Opportunities, and Threats

3. What additional strengths, weaknesses, opportunities, and threats are there with respect to FHIR for Federal health research studies?

Strengths

There has been tremendous progress made in the last five years with the adoption of FHIR and related healthcare standards across healthcare with thanks to the 21st Century Cures act rules released by ONC and CMS, requiring ONC-certified EHRs to expose FHIR APIs for patient access, as well as allowing for third-party applications to have a standardized mechanism to access FHIR APIs exposed by provider EHRs, as well as with the designation of TEFCA QHINs in December

2023. Additionally, the FHIR standard comes with a vibrantly active community with extensive network of expertise, ready-to-deploy open-source tools and infrastructure with a plethora of documentation and in-built support networks for implementers. This community and policy-backing has resulted in a robust standard for data sharing and exchange that is widely inclusive of data elements for clinical and non-clinical use cases.

Advancements to USCDI made with USCDI v3, which will now be required for EHR certification under the ONC's HTI-1 final rule released December 2023,[69,70] advance the data elements that will be useful for research including elements promoting equity, reducing disparities, and supporting public health data interoperability which will be required as of January 1, 2026. Additionally, USCDI+ will advance data elements relevant to public health surveillance and research, such as more granular data elements related to mortality, maternal health, oncology, and clinical quality.[71]

Additionally, TEFCA has published a roadmap for FHIR API adoption across their network, which will begin QHIN facilitated FHIR API exchange in 2024 (with all QHINs supporting FHIR exchange as of their December 2023 launch).[72] **It is expected that the TEFCA QHINs will be required to exchange data in FHIR format in 2024, as part of the second version of the CA, as seen in the latest draft released**.[73] This provides the ability for consented patients recruited into studies to share their full comprehensive clinical records in FHIR format to a research project with a single query to a QHIN, using the Individual Access Services purpose of use.

Weaknesses

While advancements in FHIR have been well-documented in the breadth of data elements captured, FHIR adoption is not necessarily uniform across healthcare providers, technology vendors, and other data producers. Currently, there are some notable challenges with FHIR adoption, including:

1. The lack of uniform adoption of USCDI elements beyond V1 across providers and EHR vendors. This also results in lack of standardized data quality assurance at the data source level within EHRs, because of varying levels of standardization in terminology implementation, and varying levels of deviation from USCDI V1 in EHRs. A lack of standardized data quality validation in FHIR servers means that researchers must be extra vigilant to conduct data quality validation when ingesting data into their own research FHIR servers and conducting conformance statement tests against FHIR profiles relevant to their research use case.
2. Unstructured data is not uniformly captured in FHIR, where metadata surrounding unstructured data is largely standardized but the unstructured data itself such as clinical notes, radiology reports, imaging reports are not standardized in its captured and is often exchanged in its original document form.

3. Largely, FHIR and TEFCA exchange have not been tested yet at large scale for research, while they hold promise theoretically to be able to perform at scale. Additionally, there is reliance on SMART OAuth for FHIR data exchange, however currently TEFCA relies on FAST UDAP for authentication.
4. Most HIE data currently is exchanged with CCDA using IHE profiles. While CCDA to FHIR mappings exist for centralized mapping of CCDA to FHIR at the QHIN level, there may be variation across CCDA formats and their mapping to FHIR. While TEFCA holds the promise for exchanging data in FHIR in 2024, it does not currently require FHIR exchange in the TEF. However, it may become required in future iterations of the CA and the TEF.
5. Additional specific standards-related challenges are detailed in Section III, Question 3. Notable challenges include:
    a. The flexibility of the FHIR standard results in an overabundance of more specific FHIR Profiles in IGs, some of which have overlaps for similar use cases.
    b. While not a unique challenge of FHIR, there is variance in the common vocabularies and terminologies. For example, LOINC is not universally used for all lab results (while it is the most prevalent), and RxNorm is not universally used for all medications (while it is the most prevalent).

Opportunities

While these advancements certainly hold promise for advancing research use cases, it is imperative to test these implementations in real-world applications to fully stress-test the ability for recent advancements in data standards to be applied across research use cases. The above section discusses suggested pilots to be scoped and undertaken as next steps. These suggested pilots should be scoped and prioritized with more detailed timelines, milestones, and success criteria as a next step. A prioritization framework should be developed with input from research programs that would benefit from each pilot, with assigned weighted scores on criterial relevant to research programs that will help assign the feasibility, cost/complexity, and benefits of each pilot. Such criteria can include:

- Readiness for the research program to recruit health systems and technology partners to conduct a pilot (feasibility analysis).
- Complexity of the pilot, and the availability of funds to convene a pilot (cost analysis).
- Value gained by the research program if the pilot is successful (benefits analysis).

Additionally, stewards of research programs like *All of Us* should continue to participate in forums for progressing data standards and policy for research, such as USCDI+, various HL7 groups, the TEFCA RCE, among other policymaking groups within ONC. Additionally, research programs should continue to implement and test FHIR IGS relevant to their research use cases and consolidate IGs where possible.

Threats

The biggest threats to FHIR are two-fold:

1. Within the FHIR community itself, there is often a lack of consensus with many conflicting desires within workgroups that control the progress of specific resources, which can slow down the progress of standards and lead to lack of adoption. This has led to a high level of flexibility in the base specifications of FHIR, which results in many different IGs for specific use cases. An effort could be made to consolidate similar IGs to coalesce around common use cases.
2. Outside of the FHIR community, there can be differing implementations of the standard within technology offerings across software and cloud infrastructure that can lead to the standard becoming inherently non-standard. There is a need to ensure that when ecosystems are adopted for research, that the implemented exchange and persistence standards are consistent across the research ecosystem end-to-end from data producers to consumers.

While these are not insurmountable, it is important to note that the research community must be vigilant to engage the standards community to be aware of these challenges and work towards potential solutions.

Conclusions

The purpose of FHIR is to facilitate the seamless and secure sharing of EHR and other healthcare data between different healthcare systems, providers, and applications. Use of FHIR for research purposes is a new use case. There are several opportunities for Federal research studies emerging over the next few years to take advantage of FHIR. For example, the ONC has established TEFCA as a comprehensive governance structure and policy for facilitating secure data exchange to improve care and welfare. SMART on FHIR is a potential solution to securely access, request, and utilize EHR data for research purposes. US Core profiles and a growing list of FHIR IGs such as mCODE portend standardization in how specific disease areas will be standardized. An active FHIR community exists with open-source tools and infrastructure.

While FHIR will mature over the next year, the implementation community will encounter risks and uncertainties which may delay research integration. A significant area of focus should be on advancing TEFCA implementation of data exchange for authorization-based exchange for research purposes, and more generally research. The imprimatur of ONC, NIH, and the broader research community backing this effort, would ensure the correct approach to this effort.

A significant challenge in using healthcare data for research is mapping FHIR and OMOP CDMs. Because OMOP is used extensively, this mapping is quite important. While both models strive to standardize healthcare data, they differ in purpose, scope, structure, terminologies, and data exchange. US Core profiles do present a hopeful compromise, but this may be challenged by the inherent structure of OMOP and data transformation and mapping. Moreover, US Core contains a specific subset of FHIR Resources that may not be sufficient for research with a focus on specific disease areas. A growing list of FHIR IGs such as mCODE, may advance and alleviate this gap. However, other limitations are due to the absence of a reference implementation or test datasets, semantic mapping, and the labor-intensive and time-consuming efforts to convert to standardized vocabularies. More investment may be needed.

Considering these challenges, opportunities exist to advance the use of FHIR for Federal health research studies. This paper offers policy-related recommendations, such as exploring the creation of a study participant-facing “consumer app,” involving a multi-stakeholder community to amend the CA to include a research exchange purpose and conducting a survey with providers to assess QHIN participation and IAS implementation capabilities. Technology-related recommendations include analyzing and reviewing USCDI v3 healthcare elements; establishing quality guidelines and requirements for information exchange using FHIR; evaluating approaches, architectures, and conformance-based testing for managing FHIR version transitions; developing requirements for cohort scopes, identity, and patient consent; managing FHIR version transitions by ensuring continuous interoperability through rigorous conformance-based testing; and conducting pilot studies to test FHIR profiles and IGs for research purposes.

The suggested pilots should be scoped and prioritized with more detailed timelines, milestones, and success criteria as a next step. A prioritization framework should be developed with input from research programs that would benefit from each pilot, with assigned weighted scores on criterial relevant to research programs that will help assign the feasibility, cost/complexity, and benefits of each pilot.

Lastly, the NIH, *All of Us* Research Program, the research community, academia, and industry should closely monitor the expansion, innovation, and transition to FHIR to be ready for FHIR in the coming years. The adoption of FHIR for Federal health research studies holds significant potential for transforming the way real world data is acquired – potentially streamlining and standardizing data while making it more efficient and effective to acquire for research studies. By addressing the existing challenges and implementing the proposed policy and technology-related recommendations, stakeholders can work collaboratively to streamline healthcare data exchange, enhance research capabilities, and ultimately improve patient care and welfare. The future of FHIR in the healthcare ecosystem is promising, and its continued development and maturation will play a crucial role in shaping the landscape of health research and data interoperability.

Authors and Affiliations

This paper was developed by a consortium of government and industry. The organizational affiliations of these contributors are included for information purposes only. The views expressed in this document do not necessarily represent the official views of the individuals and organizations that participated in its development:

- *Aaiza Akhtar, Booz Allen Hamilton*
- *Ben Amor, Palantir Technologies*
- *Chandra HD Bowers, Booz Allen Hamilton*
- *Michelle Casagni, MITRE*
- *Aras Eftekhari, Bixal*
- *Dave Gordon, Palantir Technologies*
- *Shannon Harrer, Palantir Technologies*
- *Shayan Hobbi, Accenture Federal Services*
- *Mario Hyland, AEGIS.net, Inc.*
- *Jonathan Lacy, Palantir Technologies*
- *James Patterson, MITRE*
- *Marty Prahl, Global CI*

The project team is grateful to the government sponsors from the *All of Us Research Program* for being available to offer their vision and insights, and for providing direction for this paper:

- *Lew Berman, PhD, MS, Branch Chief, Digital Health Technology and Data Branch*
- *Brian Gugerty, DNS, MS, RN Health Data Standards Specialist, Digital Health Technology and Data Branch*
- *John Giannini, PhD, AAAS Policy Fellow, Digital Health Technology and Data Branch*
- *Matthew Spotnitz, MD, MS, Data Scholar, Digital Health Technology and Data Branch*
- *Ryan Lupi, Covalent, Contractor to All of Us Research Program*

The project team thanks the following industry and government reviewers who provided valuable insights and feedback:

- *Brian Ahier, MITRE*
- *Vivien Bonazzi, Deloitte*
- *Robert Carroll, Vanderbilt University*
- *Alex Cheng, Vanderbilt University*
- *Lorraine Constable, Constable Consulting*
- *Paul Harris, Vanderbilt University*
- *Nilu Kundagrami, ARPA-H*
- *Jessica Reusch, All of Us Research Program*
- *Lauren Thompson, Principal Innovation*

Appendices

Appendix A: Evaluation and Comparison of Two FHIR to OMOP Mapping Engines

Mapping 1: GCP-Healthcare-Harmonization

Language: Whistle. A domain-specific language developed by Google for mapping nested data from one schema to another.
Resources:

- Base repository: https://github.com/GoogleCloudPlatform/healthcare-data-harmonization/tree/master/wstl1/mapping_configs/fhir_omop
- Core mapping definitions: https://github.com/GoogleCloudPlatform/healthcare-data-harmonization/blob/master/wstl1/mapping_configs/fhir_omop/projector_library/resources.wstl

Last updates: September 2020
Versions supported: STU3 to v6.0
Description:

- This is a mapping from 10 different FHIR resources to 8 different OMOP domains developed by a team at Google Cloud in a custom domain-specific language called Whistle.
- The implementation assumes that all instances of a given FHIR resource (e.g., *Procedure*) map to a record in a single OMOP table (e.g., *procedure_occurrence*). Figure 11 provides a full list of the mappings between resource types and domain tables.

FHIR Resource	GCP OMOP Domain	NACHC OMOP Domain(s)
Patient	person	person
Encounter	visit_occurrence	visit_occurrence
Observation	observation	measurement/observation
Address	location	N/A
Practitioner	provider	N/A
Procedure	procedure_occurrence	procedure/measurement/observation/condition
Condition	condition_occurrence	condition_occurrence
MedicationStatement	drug_exposure	N/A
MedicationRequest	drug_exposure	drug_exposure
MedicationDispense	drug_exposure	N/A

Figure 11. Mappings between resource types and domain tables

- This approach has the advantage of simplicity, however it ignores some important details, such as:
    - Mapping all *Observation* resources to the *observation* table will result in non-compliant OMOP tables. FHIR uses the *Observation* resources to capture all forms of clinical observation including laboratory results, vitals, and clinical findings. In contrast, OMOP distinguishes between the '*measurement*' and '*observation*' domains. In practice, this is implemented in mapping pipelines by considering the 'target domain' of the code attached to a particular observation and writing the output record to the appropriate table. The GCP implementation does not do this and will therefore have an empty *measurement* table and an *observation* table with many records that should instead be in *measurement.*
    - Ignoring potential for multiple lab results in a single *Observation* resource. A single FHIR *Observation* resource can contain multiple components (e.g., systolic and diastolic component observations for blood pressure measurement). The implementation does not account for this and will only map one of the components.
    - Three different medication-related FHIR resources are mapped into the drug_exposure table. These are treated as if they are independent and unrelated (each resource will create a new row in the drug_exposure table) However, these resources relate to different administrative parts of the process of prescribing and receiving medications and it seems likely that these resources would provide different information about the *same* medication. This needs to be accounted for in the mapping logic to avoid 'double counting' in the OMOP *drug_exposure* table.
- The implementation also does not process the following clinical US Core Profiles that should in theory be mappable to OMOP domain tables: AllergyIntolerance (condition_occurrence), Immunization (drug_exposure), Implantable Device (device_exposure), QuestionnaireResponse (observation), Specimen (observation)
- The Github repository does not give any indication of evaluation results from running the mapping against a test set.

Mapping 2: NACHC-CAD

Language: Java
Resources:

- Github: https://github.com/NACHC-CAD/fhir-to-omop
- Documentation: https://nachc-cad.github.io/fhir-to-omop/index.html

Last updates: October 2023
Versions supported: STU3/R4 to v5.4
Description:

- Written in Java and takes a file-based approach where FHIR resources are downloaded from FHIR server as flat files, then processed via the Java application, and then written into the OMOP domain tables that are stored in a backing database.
- In contrast to the GCP implementation, this approach correctly maps *Observation* and *Procedure* resources to different OMOP domain tables based on the target domain of the codes.
- It handles the fact that you can have multiple observations within the same *Observation* resource.
- However, unlike the GCP implementation it only considers MedicationRequest from medication resources.
- It also does not process the following clinical US Core Profiles that should in theory be mappable to OMOP domain tables: AllergyIntolerance (condition_occurrence), Immunization (drug_exposure), Implantable Device (device_exposure), QuestionnaireResponse (observation), Specimen (observation).

Appendix B: FHIR REST (aka individual) vs. FHIR Bulk Data (aka population)

Figure 12 provides a more extensive discussion on some of the key considerations for the exchange of data via FHIR REST or FHIR Bulk data that was provided in Section III, Figure 10.

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Specification Support	
Versioning: One challenge that can impact availability and completeness of electronic health information is around versions deployed in production. Right now, there are several FHIR versions in operational use (DTSU2, STU3, R4) with R5 on the horizon. Another consideration is around maturity of FHIR resources within a particular version. A client application needs to be aware of and handle these various capabilities at a national level. Data Complexity: Data mapping and availability to the FHIR Server may show inconsistency from vendor to vendor and from site-to-site deployments. Complex searches with multiple parameters may require extensive FHIR knowledge and non-deterministic results are possible.	Versioning: FHIR Bulk Export is a maturing specification with varying adoption between EHR vendors and their respective deployments. FHIR Bulk Data Access (Flat FHIR) was published as a Standard for Trial Use (STU) on 11-26-2021.
EHR Implementation & Market Deployment	
Market Adoption: FHIR servers have widely adopted support for the individual query. It is important to consider each vendor's capability and the individual customer's deployment. Insights can be gained by accessing an endpoint's capability statement using tools like MITRE's Lantern.	Market Adoption: FHIR servers have varying support for Bulk Export. Versioning: The FHIR Bulk Export specification is marked as a Standard for Trial Use and still maturing.
Search Capabilities	
Search Parameters: Individual queries offer a wide range of search parameters that can be used to filter, sort, and narrow down results, enabling complex queries that are focused on FHIR resources or sets of resources. Standardized Operations: FHIR (Create, Read, Update, Delete) operations within the specifications work to ensure uniformity across various FHIR servers. Testing and certification from implementation to deployment and into an operational state need to be assessed to ensure FHIR servers meet the standards and supported IGs. Resource Specificity: Individual queries target specific resource endpoints (e.g., Patient, Observation). This precision allows client applications to request exactly what is needed without extraneous data.	Group ID: The specification does not define how Group Resources are created and maintained in the system. The Group Resource is at a draft maturity level. Patients within *All of Us* would need to be pre-identified with each healthcare organization (FHIR endpoint) before-hand. Particular EHR vendor implementations place limits on the number of patients included in a group. Data Retrieval: Bulk has been designed to retrieve large amounts of data for multiple patients. Data Consistency: Depending on the interval between bulk data pulls, there might be a delay in reflecting the latest changes in the dataset. Response Handling Complexity: Large response payloads might require special handling, pagination, or streaming mechanisms to process efficiently.

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Response Format: Requestors can specify response formats like JSON or XML, giving flexibility based on application requirements. Search Parameter Flexibility: Use of _include and _revinclude in searching allows users to fetch related resources in one query, effectively minimizing the number of requests. Search Combining: Individual FHIR queries can use chaining (e.g., Patient?general-practitioner.name=jones) to combine search criteria across related resources. Pagination Capability: Client applications could control pagination using _count to specify the number of resources returned per page. Note: That a FHIR server can override this parameter. Patient Location: Currently, there are no capabilities at a national level to locate where an individual patient's Electronic Health Information is located. There are different Master Patient Indexes available, but they do not provide national coverage and do not provide mappings to the associated FHIR endpoints. Without prior knowledge of where patients have been seen, broadcast requests within a region or nationally would be needed at this time. The VA utilizes broadcast (IHE XCPD) on a nightly basis for all patients that have an appointment scheduled for the coming day to locate a patient's record. Consent Management: Needs further adoption through SMART, UDAP and TEFCA as implementation of sharing consents appears to be exceedingly low.	Grouped Data Retrieval: Utilizes the $group-export operation on FHIR servers, allowing for group-level data extraction (e.g., all data related to a specific patient group).[74] NDJSON Format: Responses are in Newline Delimited JSON (NDJSON) format, making it more manageable and efficient for streaming and parsing large datasets. Patient Location: Currently, there are no capabilities at a national level to locate where an individual patient's Electronic Health Information is located. There are different Master Patient Indexes available, but they do not provide national coverage and do not provide mappings to the associated FHIR endpoints. Without prior knowledge of where patients have been seen, broadcast requests within a region or nationally would be needed at this time. The VA utilizes broadcast (IHE XCPD) on a nightly basis for all patients that have an appointment scheduled for the coming day to locate a patient's record. Consent management: Needs to be resolved through pre-arrangement or updates to the specification.
Performance & Scalability	
Responsiveness: Individual queries can provide immediate responses to client requests. At the current time, ONC and the Interoperability Networks have not defined Service Level Objectives (SLO) or Service Level Agreements (SLA) indicating the FHIR endpoint availability or response times related to individual queries. Resource Utilization: Individual queries are limited in scope and would generally consume less computational and bandwidth resources. Mass Data Retrieval Suitability: Fetching large datasets accessing multiple patients across many FHIR endpoints requires multiple sequential or parallel requests can be inefficient.	Overhead: A single bulk request can retrieve large amounts of data, reducing the connection overhead seen with multiple individual queries. Asynchronous Operations: Bulk queries can be processed asynchronously, allowing systems to make a request and retrieve results later, freeing up resources for other operations in the meantime. Checkpoint recovery is lacking in the specification now, so failures may not be easily determined along with restarts at the failure point. Export Optimization: Support for scenarios where data needs to be exported to other systems or backed up. Resource Utilization: Large datasets require significant computational power, storage, and bandwidth to

FHIR Exchange Factors	
FHIR REST	**FHIR Bulk Data**
Transactional Overhead: Individual queries require setting up and tearing down a connection, which can add overhead especially when numerous requests for individual patients with their supporting resources. Potential Throttling: Numerous requests in quick succession might trigger rate limits on certain FHIR servers or platforms. Systems may throttle requests by several methods (e.g., queries per period, time of day, query prioritization)	process and can put a strain on server resources. Transferring huge datasets can consume significant bandwidth and may result in longer response times. Real-Time Challenges: Due to the potential for longer response times, bulk queries may not be ideal for scenarios requiring immediate data retrieval. Server Load: Bulk data retrieval can put significant strain on the server, affecting its performance, especially if multiple clients initiate bulk queries simultaneously. A single bulk query depending on the number of patients and resources requested could potentially take hours to even days to satisfy before data becomes available for fetching. Data Volume Management: Applications need robust mechanisms to handle, process, and store the voluminous data received. State Management: As bulk operations are asynchronous, applications need effective state management to track request status and handle data once it is ready. Limited Query Parameters: Unlike individual queries, the granularity of parameters for filtering in bulk queries might be limited based on the server's implementation. Rate Limitations: To prevent server overload, FHIR servers might have rate limits or cooling periods between bulk requests.
Testing & Certification	
The ONC Health IT Certification Program provides a conformance testing framework called Inferno[75] that allows organizations to test their systems against FHIR specifications such as US Core adopted by the Federal government. Other testing environments support a variety of capabilities, including HL7 Da Vinci, and FAST IG's to allow organizations to provide continuous testing of client and server systems while providing CD\CI pipeline integration.	FHIR Bulk Data is a maturing specification that has been tested at several FHIR Connectathons for various use cases. Each of the Connectathon Tasks that utilized the Bulk Data Access IG 2.0.0 - STU 2[76] have indicated additional features (i.e., manifest availability before fully complete, checkpoint recovery, ETA, removal of vendor/implementation limitations, Group non-match reporting) needed to further advance the standard.

Figure 12. Expanded set of factors for assessing FHIR Rest vs. FHIR Bulk Data for the exchange of healthcare data

References

[1] ACT-IAC All of Us Research Program Brief Background https://www.actiac.org/project-activity/fhir-federal-health-research-studies.

[2] Kho, AN, et. al. Privacy-Preserving Record Linkage to Identify Fragmented Electronic Medical Records in the *All of Us* Research Program. Book chapter. Springer: New York. 2020.

[3] ONC's API Conditions and Maintenance of Certification; https://www.healthit.gov/condition-ccg/application-programming-interfaces.

[4] HL7 International 10th Birthday https://info.hl7.org/fhir-10th-birthday.

[5] H.R. 34 21st Century Cures Act https://www.congress.gov/114/bills/hr34/BILLS-114hr34enr.pdf.

[6] FHIR Resource Capability Statement https://www.hl7.org/fhir/capabilitystatement.html.

[7] The FHIR Maturity Model https://www.hl7.org/fhir/versions.html#maturity.

[8] ONC 's USCDI+ https://www.healthit.gov/topic/interoperability/uscdi-plus

[9] Health IT Certification Program Updates, Algorithm Transparency, and Information Sharing https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program.

[10] US Core Implementation Guide Profiles and Extensions http://hl7.org/fhir/us/core/STU3.1.1/profiles.html.

[11] AHRQ Standardization for Health Care Quality Improvement https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata5.html.

[12] Office of the National Coordinator for Health Information Technology Trusted Exchange Framework and Common Agreement (TEFCA); https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca.

[13] The Sequoia Project, 2023. User's Guide to the Trusted Exchange Framework and Common Agreement: 51.

[14] The Sequoia Project, 2022. Standard Operating Procedure (SOP): Individual Access Service Provider Privacy and Security Notice and Practices. https://rce.sequoiaproject.org/wp-content/uploads/2023/12/SOP-IAS-Privacy-and-Security-Notice-for-CAv1.1.pdf

[15] Office of the National Coordinator for Health Information Technology, State Health IT Privacy and Consent Laws and Policies (September 2016); https://www.healthit.gov/data/apps/state-health-it-privacy-and-consent-laws-and-policies

[16] Joint Task Force. (2020, September). *Security and Privacy Controls for Information Systems and Organizations* (NIST SP 800-53 Rev. 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-53r5.

[17] Ford, W., Chokhani, S., Wu, S., Sabett, R., & Merrill, C. (2003, November*). Internet X.509 Public Key Infrastructure* (RFC 3647). The Internet Society. https://datatracker.ietf.org/doc/html/rfc3647.

[18] How OpenID Connect Works - OpenID Foundation. (2023, February 14). https://openid.net/developers/how-connect-works/.

[19] What is OAuth 2.0 and what does it do for you? (n.d.). Auth0. https://auth0.com/intro-to-iam/what-is-oauth-2.

[20] UDAP Tiered OAuth for User Authentication STU 1. (n.d.). https://www.udap.org. Retrieved November 14, 2023, from https://www.udap.org/udap-user-auth-stu1.html.

[21] Learn about SAML SSO. (2019). OneLogin. https://www.onelogin.com/learn/saml.

[22] Sync For Science. Retrieved December 8, 2023., from https://syncfor.science/.

[23] Medicare Beneficiary Data Donation, NIH and CMS Inter-Agency Agreement, CMS Blue Button – Sync for Science (S4S) Final Project Summary (2021, November 19).

https://aspe.hhs.gov/sites/default/files/documents/2fc25ef0cd5f2220f2f0bb15b04d8e42/CMS_Medicare_Beneficiary_Data_Donation.pdf

[24] SMART on FHIR: Introduction. (n.d.). Smile CDR Documentation. Retrieved November 14, 2023, from https://smilecdr.com/docs/smart/smart_on_fhir_introduction.html.

[25] HL7 FHIR Implementation Guide Registry, from https://fhir.org/guides/registry/

[26] Digital Identity. (n.d.). CARIN Alliance. Retrieved November 14, 2023, from https://www.carinalliance.com/our-work/digitalidentity

[27] CARIN, & HHS. (2023). Healthcare Digital Identity Federation Proof of Concept. https://www.carinalliance.com/wp-content/uploads/2023/03/Proof-of-Concept_Final-Report.pdf.

[28] Office of the National Coordinator for Health Information Technology. (2023, November). *Common Agreement for Nationwide Health Information Interoperability*. U.S. Department of Health and Human Services. https://www.healthit.gov/sites/default/files/page/2023-12/Common-Agreement-v1.1.pdf

[29] The Sequoia Project. (2022). *Standard Operating Procedure (SOP): Exchange Purposes*. https://rce.sequoiaproject.org/wp-content/uploads/2022/06/SOP-Exchange-Purposes.pdf

[30] Kelly, R. (2022, January 6). *A Guide to TEFCA's Six Exchange Purposes.* Health Gorilla. https://www.healthgorilla.com/blog/a-guide-to-tefcas-six-exchange-purposes

[31] 45 CFR 164.501 "Treatment", https://www.ecfr.gov/current/title-45/part-164/section-164.501#p-164.501(Treatment)

[32] Office for Civil Rights. (2003, April 03). *Uses and Disclosures for Treatment, Payment, and Health Care Operations.* U.S. Department of Health and Human Services. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/disclosures-treatment-payment-health-care-operations/index.html

[33] The Sequoia Project. (2022, September 16). *Standard Operating Procedure (SOP): Individual Access Services (IAS) Exchange Purpose Implementation*. https://rce.sequoiaproject.org/wp-content/uploads/2022/09/Final-SOP-IAS-Exchange-Purpose-Implementation.pdf

[34] 45 CFR 164.501 "Health care operations", https://www.ecfr.gov/current/title-45/part-164/section-164.501#p-164.501(Health%20care%20operations)

[35] Office for Civil Rights. (2003, April 03). Uses and Disclosures for Treatment, Payment, and Health Care Operations. U.S. Department of Health and Human Services. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/disclosures-treatment-payment-health-care-operations/index.html

[36] The Sequoia Project. (2023, April 02). *Draft Health Care Operations: Limited Exchange Purpose Implementation SOP Feedback*. https://rce.sequoiaproject.org/draft-health-care-operations-limited-exchange-purpose-implementation-sop-feedback/

[37] 45 CFR 164.501 "Payment", https://www.ecfr.gov/current/title-45/part-164/section-164.501#p-164.501(Payment)

[38] Office for Civil Rights. (2003, April 03). Uses and Disclosures for Treatment, Payment, and Health Care Operations. U.S. Department of Health and Human Services. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/disclosures-treatment-payment-health-care-operations/index.html

[39] The Sequoia Project. (2023, March). *TEFCA User's Guide – Version 1.1*. https://rce.sequoiaproject.org/wp-content/uploads/2023/03/Common-Agreement-Users-Guide-version-1.1-dated-3.22.2023.pdf

[40] L. Haberman (personal communication, December 21, 2023)

[41] Office of the National Coordinator for Health Information Technology. (Date). *The Trusted Exchange Framework (TEF): Principles for Trusted Exchange*. U.S. Department of Health and Human Services. https://www.healthit.gov/sites/default/files/page/2022-01/Trusted_Exchange_Framework_0122.pdf

[42] C. Muir (personal communication, December 06, 2023)

[43] "45 CFR 164.514", last amended 11/27/2023, https://www.ecfr.gov/current/title-45/section-164.514

[44] C. Bezold (personal communication, December 07, 2023)

[45] Office of the National Coordinator for Health Information Technology. (2023, November). *Common Agreement for Nationwide Health Information Interoperability*. U.S. Department of Health and Human Services. https://www.healthit.gov/sites/default/files/page/2023-12/Common-Agreement-v1.1.pdf

[46] Sequoia Project, Standard Operating Procedure (SOP): TEFCA Governing Council. (2022). https://rce.sequoiaproject.org/wp-content/uploads/2022/01/SOP-Governing-Council-FINAL.pdf

[47] Office of Science and Technology Policy. (2022, October 28). *Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot*. Executive Office of the President. https://www.federalregister.gov/documents/2022/10/28/2022-23489/request-for-information-on-data-collection-for-emergency-clinical-trials-and-interoperability-pilot

[48] Office of Science and Technology Policy. (2023, September 01). *Emergency Clinical Trial Data Request for Information Analysis*. Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2023/10/Data-capture-RFI-response-summary-FINAL.pdf

[49] Office of Science and Technology Policy. (2023, October). *Responses to the White House Office of Science and Technology Policy Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot*. Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2023/10/00_ClinicalTrialsRFI_DATA_Compiled-003.pdf

[50] ATHENA – OHDSI Vocabularies Repository https://athena.ohdsi.org/

[51] Regenstrief Institute, Inc., LOINC https://loinc.org/get-started/what-loinc-is/

[52] National Library of Medicine, RxNorm Overview https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

[53] Smile Digital Health. (2023, October 26). *Smile Digital Health: A complete HL7 FHIR-based Clinical Data Repository*. Smile Digital Health: A Complete HL7 FHIR-Based Clinical Data Repository. https://www.smiledigitalhealth.com/.

[54] US Core FHIR IG Change Log http://hl7.org/fhir/us/core/changes.html.

[55] SEMVER Semantic Versioning https://semver.org/.

[56] Da Vinci Health Record Exchange Implementation Guide https://build.fhir.org/ig/HL7/davinci-ehrx/index.html.

[57] ACT-IAC All of Us Research Program Brief Background https://www.actiac.org/project-activity/fhir-federal-health-research-studies.

[58] Cheng AC, Duda SN, Taylor R, et al. REDCap on FHIR: Clinical Data Interoperability Services. *J Biomed Inform*. 2021. https://pubmed.ncbi.nlm.nih.gov/34298155/. 121:103871. doi:10.1016/j.jbi.2021.103871.

[59] Nordo AH, Levaux HP, Becnel LB, et al. Use of EHRs data for clinical research: Historical progress and current applications. *Learn Health Syst*. 2019. https://pubmed.ncbi.nlm.nih.gov/31245598/. 3(1):e10076. Published 2019 Jan 16. doi:10.1002/lrh2.10076

[60] Duda SN, Kennedy N, Conway D, et al. HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. *J Am Med Inform Assoc*. 2022. https://pubmed.ncbi.nlm.nih.gov/35818340/. 29(9):1642-1653. doi:10.1093/jamia/ocac105

[61] ONC Conformance Test Tools; https://www.healthit.gov/topic/certification-ehrs/onc-conformance-test-tools.

[62] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc*. 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4997036/. 23(5):899-908. doi:10.1093/jamia/ocv189

[63] FHIR Security for Scalable Registration, Authentication, and Authorization Implementation Guide http://hl7.org/fhir/us/udap-security/STU1/

[64] TEFCA Facilitated FHIR Implementation Guide https://rce.sequoiaproject.org/wp-content/uploads/2022/12/TEFCA-Facilitated-FHIR-Implementation-Guide-Draft-2-Pilot-Version.pdf.

[65] Making Electronic Data More available for Research and Public Health (MedMorph) https://build.fhir.org/ig/HL7/fhir-medmorph/.

[66] https://build.fhir.org/ig/HL7/vulcan-rwd/

[67] Minimal Common Oncology Data Elements (mCODE) Implementation Guide https://build.fhir.org/ig/HL7/fhir-mCODE-ig/.

[68] Interoperable Digital Identity and Patient Matching https://build.fhir.org/ig/HL7/fhir-identity-matching-ig/.

[69] The United States Core Data for Interoperability (USCDI) v3 https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi#uscdi-v3.

[70] Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program.

[71] Health IT United States Core Data for Interoperability (USCDI)+ https://uscdiplus.healthit.gov/uscdi.

[72] ONC TEFCA Recognized Coordinating Entity FHIR Roadmap for TEFCA Exchange https://rce.sequoiaproject.org/wp-content/uploads/2022/01/FHIR-Roadmap-v1.0_updated.pdf.

[73] HHS Marks Major Milestone for Nationwide Health Data Exchange https://www.hhs.gov/about/news/2023/12/12/hhs-marks-major-milestone-nationwide-health-data-exchange.html.

[74] HL7 Bulk Data Access IG v2.0.0 - STU 2, Section 7.02, Behavior: Operation Definitions; https://hl7.org/fhir/uv/bulkdata/artifacts.html#behavior-operation-definitions.

[75] ONC Inferno Framework, a standards conformance testing framework, https://inferno.healthit.gov/

[76] HL7 Bulk Data Access IG v2.0.0 - STU 2 Ballot, Section 1, Bulk Data IG Home Page; https://build.fhir.org/ig/HL7/bulk-data/index.html.