The Office of the National Coordinator for Health Information Technology

National Health IT Priorities for Research

A POLICY AND DEVELOPMENT AGENDA

Prepared on behalf of the U.S. Department of Health and Human Services (HHS), Office of the National Coordinator for Health Information Technology (ONC) under Contract: HHSP233201600021I, RTI International.

January 15, 2020

Authors

The primary authors of this document at RTI International are:

- Alison Banger, MPH
- Stephanie Rizk, MS
- Jacqueline Bagwell, MS, MMCI
- Alexa Ortiz, MSN, RN
- Sydney DeStefano, BS
- Linda Dimitropoulos, PhD

The primary authors of this document at the Office of the National Coordinator for Health Information Technology are:

- Kevin Chaney, MGS
- Teresa Zayas-Cabán, PhD

The authors would like to acknowledge the contributions of Dr. Jonathan Wald, who served as the original director of this project, as well as a number of additional subject matter experts in the field who provided input and feedback on the development of the Agenda and this document.

TABLE OF CONTENTS

LIST OF EXHIBITS

LIST OF TABLES

Executive Summary

INTRODUCTION

Widespread adoption of electronic health records and the expanded use of consumer electronics has resulted in large volumes of electronic health data. This has created a tremendous opportunity for biomedical and health services researchers. However, the field has been slow to capitalize on the value of these data for research due to difficulties with both the data and the health information technology (IT) infrastructure. Challenges include ensuring data quality and consistency, establishing the governance structures and policies that allow for access to data, the limited development of research tools and services, inconsistencies in implementation across the technical architecture, and the varying needs of individuals and organizations that contribute to and utilize data within the infrastructure. Understanding these challenges and developing actions to address them is critical to advancing biomedical and health services research.

To guide the development of a future health IT infrastructure that supports the use of electronic health data for research, the Office of the National Coordinator for Health Information Technology (ONC) undertook an effort designed to accomplish three objectives: (1) articulate a vision for an ideal health information ecosystem that supports research; (2) identify stakeholders' priorities for addressing challenges within the current health information ecosystem; and (3) propose a Policy and Development Agenda (Agenda) that will guide realization of an ideal health information ecosystem in which both the health IT infrastructure and the data it supports advance scientific discovery.

METHODOLOGY

The vision and Agenda are based on findings from a background report, discussions with key informants, and an in-person stakeholder workshop. The background report and key informant interviews provided framing for the workshop at which 30 federal and private industry stakeholders were asked to identify key health IT infrastructure gaps that should be addressed to improve scientific discovery and application. ONC presented early versions of the Agenda through panel discussions at two American Medical Informatics Association (AMIA) conferences and further refined the Agenda based on attendee feedback.

VISION

The background report, key informant interviews, and workshop informed a vision for a health IT infrastructure that supports alignment of the clinical and research ecosystems where research happens faster, better, and easier, and new knowledge is available at the point of care to improve outcomes. Researchers would be able to easily access high-quality data in standardized formats with the necessary metadata to understand where, why, how, and by whom the data were collected. The health IT infrastructure would also support widespread access to the tools and services needed for more effective and efficient research, including the ability to match data across multiple sources and aggregate those data for faster analyses. Health IT infrastructure-enabled tools would facilitate patient-centered communications and mechanisms to recruit participants across a wide spectrum of organizations and settings, improving interest and engagement. When achieved, this vision would support the pursuit of more complex research questions, the development of more rapid and reliable discoveries about health and healthcare to improve outcomes, and the engagement of a broader, more representative population in research participation.

POLICY AND DEVELOPMENT AGENDA

The Agenda describes priority areas that require action to address gaps in the health IT infrastructure and realize the outlined vision. The Agenda has two overarching goals: (1) to leverage high-quality electronic health data for research, and (2) to advance a health IT infrastructure to support research. The goals and their associated priority areas are described below.

Priorities related to leveraging high-quality electronic health data for research include:

- **Priority 1**: Improve Data Quality at the Point of Capture
  - Health data are not always captured in standard formats or with the corresponding metadata needed to ensure the integrity and fidelity of the data. Individual data points need to be captured seamlessly, completely, accurately, consistently, and in a standardized format to improve the use of electronic health data for research.

- **Priority 2**: Increase Data Harmonization to Enable Research Uses
  - For health data to be used efficiently for research, they must be extracted and aggregated in a seamless manner that allows for harmonization across different organizations and also to be available for reuse for future research inquires, in accordance with established privacy and security safeguards.

- **Priority 3**: Improve Access to Interoperable Electronic Health Data
  - In addition to being captured in standard formats, data must be accessible through standardized extraction and transmission mechanisms available to all authorized users within the infrastructure. This includes the availability of reference documentation necessary to support identification and extraction of the specific data needed to answer the research question.

Priorities related to advancing a health IT infrastructure to support research include:

- **Priority 4**: Improve Services for Efficient Data Storage and Discovery
  - Research data are often inaccessible due to localized storage. Supporting standards that ensure data are both interoperable and identifiable will allow access to data in new ways, increasing the breadth of information available for research. In addition, centralized solutions to data storage may be needed to encourage those who have collected data for research purposes to maintain and make the data available for future research.

- **Priority 5**: Integrate Emerging Health and Health-Related Data Sources
  - Integrating data collected outside of the care delivery process that may affect health outcomes—such as social determinants of health, patient-generated health data, and environmental exposures—is critical to improving clinical care and research. Work is needed to support receiving, processing, and integrating external health-related data streams into health IT systems in a standardized way, as appropriate.

- **Priority 6**: Improve Methods and Tools to Support Data Aggregation

  - Advanced data functions are needed to improve the ability to aggregate data across various sources in both the clinical and research ecosystems. These include functions to effectively and efficiently support matching and linking data, honor data use agreements (DUAs), identify redundant data, manage updates to data and metadata, and work with varying data formats.

- **Priority 7**: Develop Tools and Functions to Support Research
  - Tools are needed to more efficiently search, index, and query systems to identify patient cohorts or extract data about research participants. Additional functionality could be developed to more efficiently randomize participants to treatment and control groups in a trial. Tools that support robust de-identification and use of de-identified datasets to increase confidence in security and manage risk are also needed. Patient-centered consent tools that allow patients to control and update their data-sharing preferences if embedded into the health IT infrastructure could facilitate research participation and data sharing.

- **Priority 8**: Leverage Health IT Systems to Increase Education and Participation
  - Patients and their providers may lack clear incentives to participate in or encourage participation in research. Tools and interfaces embedded within health IT systems could be used to more effectively recruit and enroll participants by providing educational materials regarding research participation and providing information back to individuals who participate in research to increase interaction and sustained engagement.
  - The health IT infrastructure should reduce barriers to participating in research to ensure inclusion and representation of all populations. Partnerships between institutions that have developed the infrastructure and tools needed to enable research participation and institutions that lack these resources may enable broader representation of diverse patient populations in research.

- **Priority 9**: Accelerate Integration of Knowledge at the Point of Care
  - Advanced methods and solutions are needed to support aggregation of research findings within the health IT infrastructure. As these systems improve, they can be used to accelerate the rigorous but lengthy process used to integrate evidence at the point of care.

ACTIONS NEEDED TO REALIZE THE AGENDA

The Agenda outlines specific actions that may be taken to advance the priorities. Each action identifies specific tactics, relevant stakeholders, and current work that may be leveraged to promote efficiency and avoid duplication of effort. Examples of specific tactics include collaborations, demonstration projects and pilot studies, education and communication efforts, application of policy levers, improving access to tools and services, support for research and evaluation, standards development, and development of tools to support research. Stakeholders include educational institutions, federal government partners, foundations, healthcare provider organizations, health IT developers, IT sector, patient advocacy groups, payors, researchers, research funding entities, and standards development organizations.

CONCLUSION

The increased collection of electronic health data and investment in health IT infrastructure over the past decade have created unprecedented opportunities for biomedical and health services research. Although there have been notable improvements in the availability of electronic health data, challenges related to

quality, access, and management remain. As a result, research is often inhibited, and advances in discovery may be delayed because electronic health data are stored across disparate systems, data standards are inadequate or inconsistently adopted and used, and challenges in governance hinder access and use. Solutions are needed to enable the research community to more quickly and efficiently benefit from availability of those data. Stakeholders' collaboration across public and private entities will be necessary to realize the vision outlined in this report.

1. Introduction

The passage of the Health Information Technology for Economic and Clinical Health Act (HITECH Act) as part of the American Recovery and Reinvestment Act of 2009 (ARRA) provided incentives to eligible physicians and hospitals to adopt certified electronic health record (EHR) technology. As a result of the Medicare and Medicaid Electronic Health Record Incentive Programs, now known as the Promoting Interoperability Programs administered by the Centers for Medicare & Medicaid Services (CMS), 87 percent of physicians and 94 percent of hospitals now use certified EHR[1] technology. The widespread adoption of EHRs has created large volumes of electronic health data that can be used to increase our ability to generate the evidence needed to advance clinical care and improve patient outcomes faster, better, and at reduced cost. And yet, the potential to leverage these data to drive improvements in care delivery and research remains largely unrealized.

Many government-led initiatives are taking advantage of the widespread availability of electronic health data to support advancements in biomedical and health services research. The Food and Drug Administration (FDA) Sentinel initiative is a distributed research network launched in 2008 to help FDA proactively monitor the safety of medical products after they reach the market. Sentinel enables FDA to rapidly and securely access a large volume of electronic health data from EHRs, claims data, and registries through a diverse group of data partners. The Million Veteran Program (MVP), led by the Department of Veterans Affairs (VA) Office of Research and Development, aims to partner with veterans to study how genes, lifestyle, and military exposure affect health. MVP will build one of the world's largest medical databases of health information from one million veteran volunteers. More recently, the *All of Us* Research Program (*All of Us*), led by the National Institutes of Health (NIH), is a historic effort to gather longitudinal data from one million or more people living in the United States to accelerate research and improve health. *All of Us* will serve as a national research resource to inform thousands of studies, covering a wide variety of health conditions. Researchers will be able to use data from the program to learn more about how individual differences in lifestyle, environment, and biological makeup can influence health and disease. In addition, participants will be able to learn more about their own health and contribute to an effort that may improve the health of generations to come.

There are also a range of private-sector initiatives aimed at improving interoperability of electronic health data. The National Patient-Centered Clinical Research Network (PCORnet) is an integrated partnership of clinical researchers, health plans, and patients that is funded and managed by the Patient-Centered Outcomes Research Institute (PCORI). PCORnet developed a shared common data model and a robust data infrastructure that includes a broad range of organizations, from cutting-edge academic medical centers to local community health clinics. The Argonaut Project is a private sector initiative to advance industry adoption of modern, open interoperability standards. The purpose of the Argonaut Project is to rapidly develop a first-generation Health Level Seven International (HL7®) Fast Healthcare Interoperability Resources (FHIR®)-based application programming interface (API) and core data services specification to enable expanded information sharing for EHRs and other health IT based on Internet standards and architectural patterns and styles.[2] In addition, the CARIN Alliance of more than 60 health IT developers, consumer technology companies, application developers, and healthcare providers proposed a voluntary code of conduct for handling patient healthcare data shared through APIs with entities that are not covered under the Health Insurance Portability and Accountability Act (HIPAA), such as smartphone application developers. The Alliance's approach to using APIs offers a pathway for other organizations and the government to follow as APIs become more prominent in healthcare.

The broader information technology sector is also developing tools and platforms that may help researchers access and use electronic health data for research. For example, Google's Verily Life Sciences program develops tools to collect and organize health data, Microsoft's Azure platform for implementing artificial intelligence, and applications of the Amazon Web Services (AWS) such as Amazon Comprehend Medical, which is a natural language processing service that makes it easy to use machine learning to extract relevant medical information from unstructured text like clinical notes. Along with a variety of governance and standards initiatives that seek to support interoperable exchange of data for both clinical and research purposes, the landscape around capture and use of electronic health data continues to improve.

1.1 CHALLENGES WITHIN THE HEALTH INFORMATION ECOSYSTEM

Despite the progress made in advancing the use of electronic health data for research, numerous challenges remain. The limited use of mechanisms for achieving interoperability between and amongst diverse clinical and non-clinical data systems, hinders advances in both care delivery and research.[3] To overcome this challenge, industry experts have called for a "unifying software architecture for the exchange of health information" that frees data from organization and system silos.[3]

A critical challenge to achieving interoperability is to ensure use of existing data standards and also to keep current with the development of new standards needed to capture new data types. Important steps have been made in the adoption of standards and the availability of standardized clinical data for research, but the rapidly growing volume of data and number of data types is outpacing this progress.[4] To achieve interoperability, we need to develop technical services and standards for services that allow patient data to be securely linked to other data sources; develop standards, services, and policies to ensure data quality for research; create a policy framework that preserves security and privacy while improving the ability to access and query clinical data by researchers; and develop a better understanding of and methods to address the socio-legal challenges related to using patient data for research.[4]

Other key challenges to achieving an ideal health information ecosystem for researchers include linking data across sources, improving data access, ensuring privacy for research participants, and engaging patients in the process of collecting research data that matters to them. Gathering data from EHR systems outside existing research networks remains expensive and labor-intensive, and incentives are not aligned among researchers, the healthcare system, and research participants. Some have noted "the most critical step to promoting policy changes to improve the data infrastructure and data access is demonstrating the value in leveraging data to end users, including healthcare consumers, clinicians, health systems leaders, payers, and policy makers."[5]

Health care organizations currently hold the bulk of clinically relevant electronic health data. There are also diverse and expanding data sources that, when linked to clinical data, may be used to inform treatment and care delivery. Many external factors affect patients' health: diet, exercise, environmental exposures, housing, insurance coverage, and geographic location all exert significant influences on health. Data collected during clinical care are a rich source of data for research. Combining clinical data with other types of data can provide a more holistic picture of human health and lead to more productive avenues for research. Opening the research landscape to make it easier to integrate multiple sources of data about the factors that influence health is critical to improving health and healthcare.[1-3]

1.2 RECENT POLICY DEVELOPMENTS

The health IT infrastructure includes the technical architecture, standards development processes that drive quality and consistency, governance and policies that oversee access and use of data, and the individuals and organizations that contribute to and utilize data within the infrastructure. Recent policy developments that will impact the health IT infrastructure include a changing legal landscape regarding privacy and security and new proposed rules from the Department of Health and Human Services (HHS) that aim to advance interoperability.

Privacy, security, and trust are foundational to using electronic health data for research, and ideas about these topics continue to evolve, requiring an increasingly responsive policy environment. A 2019 report by the National Committee on Vital and Health Statistics (NCVHS) identifies privacy, security, and access measures to protect individually identifiable health information in an environment of electronic networking and multiple uses of data beyond HIPAA.[6] New laws and regulations, such as California's Consumer Privacy Act and the European Union's General Data Protection Regulation, give individuals new rights related to the collection and use of their data. The changes may impact how information is used for research and necessitate additional infrastructure that can support individuals' preferences regarding their health data.

In addition, there are currently few incentives for entities that hold data to share it. However, recent policy developments propose to promote electronic health data sharing through both technical and policy requirements. In 2019, ONC released a notice of proposed rulemaking that would require certified health IT products to support the export of electronic health information for both a single patient and for multiple patients in a defined population through APIs. The proposed rule also replaces the definition of the common clinical dataset standard with the United States Core Data for Interoperability (USCDI). The USCDI Version 1 has been established to include updated versions of vocabulary standard code sets; address and phone number; pediatric vital signs; provenance data elements; and clinical notes, including discharge summary, history and physical, progress, consultation, imagine narrative, laboratory report narrative, pathology report narrative, and procedures. These proposed requirements would support improvements in patient matching, the development of a longitudinal record, access to information for research purposes, and inclusion of standardized metadata information about who, when, and where data were collected.

Concurrently, CMS also issued a proposed rule, which recommends changes to the healthcare delivery system that will increase the seamless flow of health information, reduce burden on patients and providers, and foster innovation by releasing data for researchers. CMS has proposed requirements that Medicaid, the Children's Health Insurance Program, Medicare Advantage plans, and Qualified Health Plans in the federally facilitated exchanges provide enrollees with immediate electronic access to medical claims and other health information by 2020. In addition, CMS will require these healthcare providers and plans to implement open data-sharing technologies to support transitions of care as patients move between these plan types. Ensuring that patients have easy access to their information—and that information follows them on their healthcare journey—will reduce burden and help eliminate redundant procedures and testing, thus giving providers the time to focus on improving care coordination and, ultimately, health outcomes.

More recently, NIH issued a notice[7] to encourage NIH-funded investigators to explore the use of FHIR to capture, integrate, and exchange clinical data for research purposes and to enhance capabilities to share research data. In addition, NIH issued a notice[8] to small business communities that announces NIH's

special interest in supporting applications that use FHIR in the development of health IT products and services.

The policies proposed by ONC and CMS under the Promoting Interoperability Program align to advance interoperability in several important ways. CMS proposed that program participants must conform to the same advanced API standards as those proposed for certified health IT in the ONC proposed rule and include an aligned set of content and vocabulary standards for clinical data classes through the USCDI. Together, these proposed rules address both the technical and healthcare industry factors that create barriers to the interoperability of health information and that limit a patient's ability to access essential health information. In addition, in their recently released notice, NIH specified that their support for FHIR standards and APIs aligned with the objectives of both the ONC and CMS proposed rules, as well as the NIH *Strategic Plan for Data Science*.[9] Health services and biomedical researchers are also expected to benefit from improved interoperability, but additional policy and development work will be necessary to ensure maturation and use of the health information ecosystem. Aligning requirements for payors, healthcare providers, health IT developers, and researchers will help to drive an interoperable health IT infrastructure across systems, ensuring providers and patients have access to health data when and where it is needed and also making those data available, as appropriate, for scientific discovery.

1.3 ADVANCING A HEALTH IT INFRASTRUCTURE THAT SUPPORTS RESEARCH

Ensuring that the health IT infrastructure effectively supports both care delivery and biomedical and health services research is paramount to leveraging the use of electronic health data to improve outcomes for patients. To achieve this goal, ONC undertook an effort designed to accomplish three objectives: (1) articulate a vision for an ideal information ecosystem that supports research; (2) identify priorities articulated by stakeholders to overcoming challenges within the current ecosystem; and (3) propose a Policy and Development Agenda that will contribute to realizing an ideal health information ecosystem.

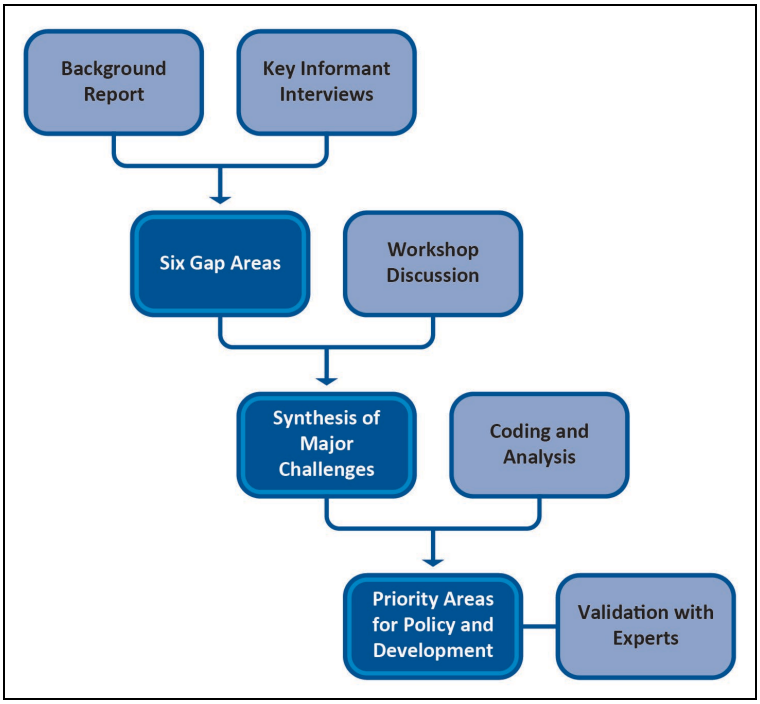
Several functions enabled by the current health IT infrastructure are shared for clinical and research purposes. These functions include data capture, data access and transport, data storage, data aggregation, and knowledge sharing. Currently, however, these shared functions are often configured differently for clinical and research purposes. Greater alignment of these functions will support better care delivery and research by providing better quality data and the ability to merge clinical data with other data sources.

At the center of this health IT infrastructure is the participation of the patient. Meaningful patient engagement is critical to both the effective delivery of clinical care and the research ecosystem. Increasing patient access to data, improving consent management processes and tools, and providing more advanced methods for participating are all areas ripe for improvement and could be better supported via the health IT infrastructure.

2. Methods

To better understand the challenges and opportunities to leverage the health IT infrastructure and the data it supports for research, we conducted a focused review of relevant literature, engaged experts in the field through key informant interviews, and held an in-person workshop. Once these activities were completed, a draft list of priorities was presented to stakeholders for feedback at two AMIA conferences in 2018 and 2019. Specific details about the approach to developing each input are noted below. **Exhibit 1** summarizes the overall approach.

Exhibit 1: Overview of project methodology



- ✓ Step 1. Gather **information** via the literature and informant interviews about initiatives utilizing electronic health data for research.
- ✓ Step 2. Identify **gaps** in the current health IT infrastructure and bring stakeholders together to discuss those gaps.
- ✓ Step 3. Synthesize discussion findings to understand major **challenges and opportunities** within and across those gaps that inhibit the use of health data in research.
- ✓ Step 4. Identify and validate **priority areas** and specific policy and development actions that will lead to improvements in the infrastructure and data needed to achieve a desired future state for biomedical and health services research.

2.1 BACKGROUND REPORT AND KEY INFORMANT INTERVIEWS

Selected peer-reviewed and grey literature, relevant research programs, and other health-related initiatives—including government (e.g., *All of Us* Research Program, the Million Veteran Program), commercial (e.g., Amazon, Google/Verily, Apple), and other initiatives—were summarized in a background report to inform preparation for the in-person workshop.

Five key informant interviews were conducted to validate and further inform the content and direction of the background report and identify potential areas for further discussion at the in-person workshop. The key informants represented experts in EHR interoperability and usability, health IT architecture requirements needed to support research, the most critically needed functionalities and services, patient protection of data and data security, data policies, and the development of core datasets for research. Interviews were conducted using a semi-structured discussion guide tailored to the informant's specific area of expertise.

The synthesis of information collected in these interviews provided insight into the gaps identified in the background report, including how EHRs are used in the collection of health data for research, interoperability and using health data for "the public good," patient protections and authorization for data used in research, specific standards (e.g., taxonomies) and organizational approaches that support research activities, data transmission standards such as FHIR, cloud services, the utility and expectations around the use of common data models, and the difficulties inherent to operationalizing evidence-based learning systems within the current EHR-based infrastructure. The information provided in the interviews was used to weight the perceived impact each topic might have in developing an action-oriented set of priorities.

Taken together, the background report and key informant interviews pointed to the following six gap areas in the health IT infrastructure that need to be addressed:

- Adaptability of the health IT infrastructure
- Ability to produce high-quality data for research
- Functionality needed for research
- Data aggregation across multiple platforms
- Advancement of patient engagement in research
- Realizing a transparent and scalable architecture

2.2 IN-PERSON WORKSHOP

A 2-day in-person workshop was conducted in July 2018 to discuss gaps in the health IT infrastructure that, if addressed, would improve scientific discovery and application. Workshop participants, including representatives from federal agencies, technology vendors, and researchers, were selected to represent a wide range of expertise and a diversity of perspectives. **Appendix A** provides a list of participants, and **Appendix B** includes the agenda for the workshop.

A keynote speaker, Dr. Kenneth D. Mandl, provided examples of high-impact health IT innovations that support advances in the access and use of health data for research. The workshop was conducted using both small-group and large-group facilitated discussions. Each day was devoted to discussion of three of the six topics identified in the background report. Participants were assigned to one of three small groups, followed by group reports and large-group discussion. A trained facilitator moderated each of the six small-group sessions, and a notetaker captured the details of the discussions.

The workshop notes were analyzed and recurring themes were identified in a high-level synthesis. **Table 1** provides an overview of the crosscutting topics that emerged from the workshop discussions.

Table 1: Critical crosscutting topics identified in workshop discussions

	Gap Areas/Workshop Breakout Sessions					
Identified Crosscutting Challenges	***Adaptability of the Health IT Infrastructure***	***Producing Data for Research***	***Health IT Functionality Needed for Research***	***Data Aggregation Across Multiple Research Platforms***	***Advancing Patient Engagement in Research***	***Realizing a Transparent and Scalable Architecture***
Lack of transparent and interoperable health-related data	●	●	●	●		
Lack of tools that allow researchers to learn how to use, interact with, and share standardized EHR data	●	●	●	●		●
Lack of support for solutions to enable aggregation across multiple, non-EHR-based data sources		●	●	●	●	●
Lack of coordination and sharing of functional solutions for patient matching and identity management	●	●	●	●	●	●
Lack of education, coordination, and technical solutions around consent management necessary for data sharing in research	●	●	●	●	●	●
Lack of research opportunities for areas and organizations that have traditionally been underserved or underutilized in research participation			●		●	
Lack of opportunities to encourage dialogue and education on the use of the health IT infrastructure for research	●	●				

2.3 IDENTIFICATION AND VALIDATION OF FINAL PRIORITIES

A draft list of Policy and Development Agenda topics emerged from the workshop synthesis. These topics were presented at the AMIA Annual Symposium in November 2018.[10] An updated draft list of topics was subsequently presented at the AMIA Informatics Summit in March 2019.[11] Feedback from attendees was used to further refine the topics.

3. Vision

The background report, key informant interviews, and workshop informed development of a vision for a health information ecosystem where research happens faster, better, easier, and new knowledge is available at the point of care to improve outcomes.

In the ideal state, electronic health data will have high reliability and validity. Further, researchers will be able to access the data they need to answer a research question with minimal effort. These data will be in standardized formats and will include the necessary metadata to understand where, why, how, and by whom they were collected. The data will also be available to researchers using common models for consolidating and securely transmitting the information.

The future health IT infrastructure will support a range of functionalities, services, and tools needed to advance research. This infrastructure will provide a platform through which researchers, providers, and patients know more about where data are stored and how they are used as well as have easy access to mechanisms that support open communication with one another. The health IT infrastructure will have services to locate data stored across various sources and tools that support needed research functions, including aggregating data from multiple internal and external sources and incorporating robust methods of identifying patients and matching them across systems to ensure adherence to the necessary privacy and security procedures. Effective communication will support efficient and patient-centered consent management. Health IT-enabled tools will enable researchers to recruit, enroll, and engage research participants and provide appropriate education about their rights. Advanced computation capabilities will be used to drive new insights and return knowledge to the point of care for actionable use.

When achieved, this vision will support the pursuit of more complex research questions and the development of more rapid and reliable discoveries about health and health outcomes, and engage a wider, more representative population in research participation.

4. Policy and Development Agenda

The Policy and Development Agenda is organized around two key goals that support the vision of a health information ecosystem where research happens faster, better, and easier, and new knowledge is available at the point of care to improve outcomes: (1) leveraging high-quality electronic health data for research and (2) advancing a health IT infrastructure to support research. Nine priorities with corresponding supporting strategies are associated with these two goals, which aim to strengthen the health IT infrastructure and increase opportunities for use of electronic health data for research over the next 3 to 5 years.

4.1 LEVERAGING HIGH-QUALITY ELECTRONIC HEALTH DATA FOR RESEARCH

4.1.1 Priority 1: Improve Data Quality at the Point of Capture

Large research datasets require a high level of validity and reliability to support meaningful analysis and comparison. However, health data are not always captured in standard formats or with the corresponding metadata needed to ensure the integrity and fidelity of the data. A provider's primary goal is to collect the information needed to treat the patient; this information may or may not be captured in a way that is complete or of suitable quality for research purposes. Data capture at the point of care imposes a certain level of burden, which may also impact the completeness of the data. Any attempt to improve data quality should avoid the possibility of adding further responsibility to the provider. The adoption and use of data standards and accompanying metadata for health data capture is critical to effectively exchange data for clinical care and to enable complete and consistent interpretation for research. Incorporating current and emerging data and metadata standards could facilitate the development of richer, higher-quality datasets.

Priority 1: Supporting Strategies

- Identify and develop metadata standards that capture more information about a given data point at the time of capture
- Promote the adoption and use of current and emerging data and metadata standards to improve data quality for care and research

Metadata are used to describe a given data element so that it is easier to understand, use, and share. Useful metadata include a range of embedded information, such as: (1) descriptive information on what identifies the data (name), (2) structural information on where the data are stored (location), (3) administrative information on when data were created and by whom (standardized date/time and provenance), (4) statistical information on how data were collected (method of administration), and (5) referential data providing other pieces of information about the systems and processes influencing the origin of the data. Electronic health data generated during a patient-provider interaction and documented in an EHR often lack the metadata required for a researcher to fully understand the context needed to answer a specific research question. The same is true for data collected outside the patient-provider interaction, through medical devices, consumer wearable devices, and other data sources. This lack of metadata prevents researchers from being able to understand the meaning of these data with the degree of reliability needed to support research.

4.1.2 Priority 2: Increase Data Harmonization to Enable Research Uses

Common data models (CDMs) enable the analysis of disparate datasets by transforming the data into a common format. CDMs allow researchers to integrate and analyze data from multiple sources. Several initiatives have developed CDMs to fit both their primary data source and intended purpose. Even with the development of a CDM, the process of extracting data and harmonizing it across data sources is time consuming, one of many factors that make researchers reluctant to share datasets once collected.

Priority 2: Supporting Strategies

- Increase support for the development and use of existing common data models to transform and analyze data for research purposes
- Identify collaborative opportunities to improve understanding regarding research data use and reuse in accordance with established privacy and security safeguards

Efforts to harmonize existing CDMs have met significant challenges.[12] A 2013 paper in *Medical Care*[13,14] reviewed four CDMs to determine how the models compared and what might be needed to align them: Observational Medical Outcomes Partnership (OMOP) (EHR data), Mini-Sentinel Common Data Model (MSCDM) (EHR and claims data), Clinical Data Interchange Standards Consortium's Analysis Data Model (CDISC-ADaM) (clinical trial data), and Biomedical Research Integrated Domain Group (BRIDG) model (clinical trial and preclinical protocol-driven research data). They extracted data based on a patient-centered outcomes of research (PCOR) study scenario from an EHR into a local clinical data warehouse for research. They then mapped the scenario data in the warehouse to the four CDMs they studied to compare the fit. Although many fields mapped easily, some field transformations required experts (data source designers and administrators) knowledgeable about the source data codes and context, which creates difficulty when trying to scale the effort. Sometimes the same data from different fields must be reconciled, which can create conflicts, such as a provider associated with a procedure (according to a billing code) versus the provider associated with a patient visit. Mechanisms to understand the context and resolve the conflicts among the CDMs are still needed.

A number of close-knit collaborative academic communities have created successful data-sharing platforms (e.g., Yale University Open Data Access and the Infectious Disease Data Observatory).[14-17] Replication of such standards and platforms will require similar coordinated efforts, which is unlikely to happen without long-term investment and federal intervention.

4.1.3 Priority 3: Improve Access to Interoperable Electronic Health Data

An API is a set of requirements that detail how one system or application communicates with another. An *open* API requires that, for example, a health IT developer share the technical specifications of its product's API to enable data exchange. Open APIs represent a key step forward in helping individuals and providers to access and combine health information from disparate sources. Requirements related to open APIs for health IT developers allow data to be accessed and extracted, but often the data received do not include enough context to make the data useful to researchers. For example, the dataset may include information about readmissions, but does not include information about how the data are stored or calculated, which makes it difficult to identify exactly where the data are found within the database of each unique implementation of the product. To truly identify the location of the desired data available through the API, health IT developers need to publish relevant schema documentation.

Priority 3: Supporting Strategy

- Ensure health IT systems provide sufficient documentation about their data models and technical specifications to develop shared tools for acquiring clinical data from those systems

Currently, many health IT developers see the information about their schemas as proprietary and often do not allow customers to publish information about their underlying architectures and pertinent database schemas for storing data in the back end of their systems. Workshop participants repeatedly noted that without knowledge of the granular details included in the schemas, it is difficult to fully leverage integration of other tools and applications used for extraction and analysis of data elements. Even those at large-scale research programs report feeling that they are "moving around vendors, not with them." At the same time, developers report feeling overwhelmed trying to keep up to date with all the requirements needed to maintain their status as a certified health IT product. Clear articulation from the research community regarding what they need to know about the schema and why, in order to be able to do research, may be needed.

4.2 ADVANCING A HEALTH IT INFRASTRUCTURE TO SUPPORT RESEARCH

4.2.1 Priority 4: Improve Services for Efficient Data Storage and Discovery

Although the volume of data generated by patients, both within and outside of the clinical care setting, continues to grow, incentives and controls on the use of these data[18] for sharing for research purposes are limited. Similarly, as research funders encourage research-related data storage and maintenance, there are concerns about long-term management of distributed data sources both within and outside of an organizational umbrella. Increasingly, requests for research funding should account for the need to convert data into interoperable formats and to maintain them for long-term use. Incentives for sharing and reusing data could reduce redundancies in data collection and reduce research costs. Cost savings could also be realized by leveraging increased computational capacity and centralized storage solutions offered by the broader IT sector.

Priority 4: Supporting Strategy

- Realize efficiencies by making advanced computational capacity and storage available to researchers to reduce redundant data collection efforts

4.2.2 Priority 5: Integrate Emerging Health and Health-Related Data Sources

Although clinical data about individuals are valuable to researchers, information about an individual's health or relevant health outcomes often exists outside the clinical environment. The current health IT infrastructure does not sufficiently support the routine and standardized collection of other data relevant to care and research, such as patient-generated health data (PGHD), social determinants of health, environmental exposures, or information about access to community services.

Priority 5: Supporting Strategies

- Support functionality within the health IT architecture to link research-relevant data sources outside the patient care setting with EHR data
- Provide support for accelerating the process of standardizing new data concepts while working to update current standards

As interest in using PGHD collected from consumer wearables, medical devices, or directly from patients increases, the specific data elements used to capture underlying concepts like "activity" must be standardized and integrated into the products and tools used to capture and share that information. Currently, medical device manufacturers have sole control of metadata standards for medical devices, without input from standards development organizations.

Social determinants of health impact health outcomes, but are not consistently captured at the point of care in a standardized way. These data are also important to many research studies that would benefit from standardized data fields and corresponding data. Efforts are currently underway to enable standardized capture and sharing of these data using health IT for clinical care and research. ONC's Advancing Standards for Precision Medicine project is working toward identifying data types (e.g., social determinants of health data, sensor data) for development through standard development organizations, and testing and piloting these data profiles into EHR systems for harmonization with research-focused activities.[19] Concurrently, the Gravity Project[20]

has convened a multi-stakeholder group to develop a systematic approach to documenting and aggregating variables such as food security, housing stability and quality, and transportation access and creating FHIR implementation guides on defined use cases. EHR data may also be used in combination with public and community data to predict social determinants of health,[21] and tools are needed to efficiently and accurately enable such data linkages.

As scientific discovery broadens and the field of medicine evolves, so does the need to include new data types in clinical care and research, such as genomic sequencing or genetic test results. The National Academy of Sciences (NAS) 2011 *Toward Precision Medicine*[22] report describes new approaches to research that leverage routine collection of data outside of a research setting that uses health IT. A new taxonomy informed by molecular medicine could reveal connections between outcomes and many other attributes, including the patient's genes (genome), other factors that influence gene expression (epigenome), the microscopic organisms that coexist inside and around the human body (microbiome), the particular physical findings they demonstrate and symptoms they experience, other reported experience (PGHD or patient-reported outcomes [PRO]), their environmental exposure (exposome), or any other type of data that relates in some way to the patient (e.g., social determinants of health) that can be measured and recorded.

Health services and biomedical researchers are increasingly looking to expand the range of research questions that can be answered by linking data from a variety of sources, including social determinants of health, health-related social needs data, genomics data, medical device data, environmental data, PGHD, and other data types. While recognition among researchers that linking multiple sources of data will advance health services research leading to improved treatment and care has grown, the importance of accelerating this potential has received a boost from the 21st Century Cures Act (Cures Act). The Cures Act provides support for accelerating and advancing scientific discovery and supports a variety of research opportunities, such as the *All of Us* Research Program led by NIH. In addition to collecting clinical and biologic data from research participants, *All of Us* plans to support collection of additional forms of data, such as PGHD from wearable devices.

When new data types or concepts emerge, care delivery organization and researchers may develop overlapping or competing standards for collecting similar data elements. The process of sharing information across care delivery organizations or collaborating between research teams to converge on a single standard for a new data concept often takes years. In addition, the governance and regulatory authority for a given standard might be governed by multiple federal departments and agencies working in conjunction with standards development organizations (SDOs), further complicating the process.

As health services and biomedical researchers identify additional data concepts needed to answer complex research questions, it will be important to establish rapid standards development procedures that are continually updated, not too dissimilar to how Logical Observation Identifiers Names and Codes (LOINC) is expanded and updated.[23] Where accepted data standards exist, the research community must work to accelerate their definition and implementation within the health IT infrastructure—a process that takes a significant amount of time and effort and is currently done on a volunteer basis.

4.2.3 Priority 6: Improve Methods and Tools to Support Data Aggregation

Advanced data functions are needed to effectively use electronic health data for clinical care and research. The capacity to accurately and efficiently match individuals' health records across organizations is critical to achieving interoperability and to creating comprehensive, longitudinal views of a person's care. Incorrect matching can result in delays of care or medical errors, compromise privacy, and increase administrative burden. Currently, organizations use a range of approaches for matching and, although match rates are often high within a single organization (more than 90 percent), they often drop significantly across organizations (less than 50 percent).[24,25] As more disparate sources of data are used to answer research questions and improve health, the ability to match individual patients with their health records is a growing challenge. A single identifier and/or additional or better patient attributes would greatly enhance researchers' ability to link disparate sources of information about a single individual.

Priority 6: Supporting Strategies

- Improve the ability to match individuals to different sources of data
- Develop tools to efficiently manage data use agreements across organizations
- Develop functionalities needed to manage data across distributed sources, including to identify redundancy; account for updates to data and metadata; and analyze data in different formats

Additional functions are needed to effectively honor DUAs between healthcare or research organizations. All large-scale research networks such as PCORnet undergo an intense process to develop a shared DUA, which outlines the parameters around which data can be shared between organizations for secondary uses such as research. Although the basic terms of these interorganizational agreements are often available, there have not been successful efforts to establish a single common agreement for research data use. Various functionalities within the health IT infrastructure could potentially allow for the management of terms of various agreements and track which institutions are willing and able to adhere to the terms of a particular DUA to enable appropriate sharing of data.

Given issues with data quality, functionality that helps identify redundant or duplicate data across fields within a record or different information systems will enable researchers to more easily distinguish a "unique" data point. Tools are also needed to manage and account for updates to data and metadata, and to aggregate and analyze data in different formats. The landscape for developing these tools has increased rapidly, requiring specific initiatives to track and monitor their development.

4.2.4 Priority 7: Develop Tools and Functions to Support Research

Additional tools and functions are needed to support research that could be embedded within the health IT infrastructure. In particular, tools are needed to more effectively index, search, and query systems to identify and recruit possible patient cohorts for a given study as well as easily extract data about participants. Such tools should facilitate not only participant recruitment, but enrollment and randomization as well.

Priority 7: Supporting Strategies

- Support easier consent management for research
- Develop additional tools to support research processes such as recruitment, enrollment, randomization, and HIPAA-compliant de-identification
- Investigate and expand tools that index, search, and query systems to identify and recruit possible patient cohorts for a given study as well as easily extract data about participants

Within the healthcare sector, the HIPAA Privacy and Security rules, state law, and organizations' own policies and practices are among the key drivers of how health information is collected, used, and shared, but how those rules and regulations are understood and implemented can vary substantially.[26] The current model for healthcare privacy emphasizes consent for disclosure and limits the ways in which healthcare organizations can use, reuse, and disclose information. However, the HIPAA Privacy Rule establishes conditions for sharing of personal health information (PHI) for research purposes, including when a research participant authorizes its use or disclosure. Automated tools are needed to provide potential research participants with easy ways to authorize such disclosures as well as to consent to research participation and to maintain a record of that consent in electronic format that could be discoverable by a larger network of researchers. This would lead to a more efficient and streamlined process for both the researcher and the study participant, which would require robust privacy protections and security practices to avoid potential misuse. Workshop participants noted that although consent is critical, it is currently used primarily to limit liability of the research organization and that consent, its function, and its implementation should be examined thoughtfully.

As functionalities around obtaining and analyzing patient data progress, so should the ways in which health IT supports researchers to collect and manage consent from participants, and to collect and manage their data. Electronic consent management services have begun to emerge through various services such as health information exchanges and private industry application developers. Along with advances in technology, proposed changes to federal regulations like CFR 42 part 2[27] and continual shifts in state laws related to consent make some form of centralized electronic mechanism to track and update an increasingly valuable proposition within the broader research ecosystem. If electronic consent management applications also provided the opportunity for patients to track their consent activities across research studies, it may also serve to decrease issues around privacy and trust.

Furthermore, tools that support robust de-identification and use of de-identified datasets to increase confidence in security and manage risk are needed. Significant reductions in the time and effort needed to advance research could be realized if datasets were shared more openly and widely among the research community. Sharing de-identified information is often presented as a solution for making sharing information for research easier. Although HIPAA specifies acceptable strategies for de-identification (removal of 18 specific identifiers or expert determination), researchers may be reluctant due to concerns

about the reliability of the process and the increasing availability of tools that support re-identification. As a result, even though HIPAA permits sharing of de-identified data for research purposes, this does not always happen in practice. A systematic literature review conducted in 2012[28] identified two key challenges to conducting research: "the absence of standardized ontologies and data collection" and "the unique data governance concerns related to the transfer, storage, de-identification, and access to electronic clinical data." Likewise, researchers are confronted with a complex array of data-sharing requirements, including restrictions on how data obtained through a local query can be shared, that prevent researchers from making the data available for reuse for other studies.

Specialized functions for researchers, such as locating specific data, searching multiple data sources, indexing data of interest, querying for matching records, and identifying consenting and randomization status are foundational components of a more efficient and effective research ecosystem. As centralized storage capacity and data aggregation functionality increase, so should the tools and functions that support the ability to search data that has been indexed and stored in a secure, searchable resource.

4.2.5 Priority 8: Leverage Health IT Systems to Increase Education and Participation

Unaware of the importance and availability of opportunities for participating in research to spur scientific discovery, patients may not be motivated to share their data or become involved in research efforts. Similarly, providers might be unaware of opportunities for their patients to participate in research. Further, while individual patients may experience a noticeable impact on their health outcomes based on involvement in a clinical trial, most research findings are part of a larger process of collecting evidence over time across multiple, independent research studies.

Priority 8: Supporting Strategies

- Develop health IT tools that deliver value for providers and patients to participate in research
- Pursue infrastructure improvements that enable participation from a diverse patient population
- Expand research opportunities beyond large health systems

There is an opportunity to develop health IT tools that could support more effective communication with patients and providers for research participation. Sophisticated interactive text alerts sent to smart phones, decision-making tools embedded into patient-facing health applications, and social media campaigns targeting specific conditions or disease groups—these are all methods of education and outreach that are already utilized by a growing number of research teams. Support for more widespread and coordinated use of these tools could support education at the population level that focuses on transparency and the role patients play through research participation could improve individuals' motivation and willingness to share their information. Patients and their families will be more likely to participate in research if they understand why their data are important and how their participation generates the evidence that drives improvements in clinical decisions. Knowledge of how their data are used will improve trust in the research process, and comprehension of their data's value and its potential impact will influence involvement.

The Federal Health IT Strategic Plan[29] calls for empowering individual, family, and caregiver health management and engagement and fostering individual, provider, and community partnerships. Achieving this goal requires taking a person-centric perspective that addresses multiple components of health and

healthcare. As research continues to investigate social determinants of health, the knowledge of what makes people healthy or alters their disease progression is changing rapidly. When communities lack resources to gather, access, analyze, and report data, there is an uneven representation of patients feeding information to research studies. This also results in uneven access to research-based interventions and decision support that are shown to improve outcomes. Better representation involves more diversity in age, gender, race, ethnicity, education, socioeconomic status, and geographic location.

Research results are not always shared with participants, resulting in little value or incentive for patients to participate in research activities. Health IT can be leveraged to more systematically engage participants throughout the duration of a study by sharing information collected for the study, as well as research study results, directly with participants. This may also serve as motivation for individuals to participate in other research studies.

Moving beyond the patient's individual-level involvement, research opportunities must also span across organizations. Smaller, less-resourced organizations may be unable to invest in needed infrastructure to engage in research. Research is usually outside the mission and scope of smaller institutions that focus primarily on patient care. Consequently, entire populations served by smaller institutions and the data collected by those institutions are not being considered. The health IT infrastructure must evolve in a way that reduces barriers to participating in research opportunities to ensure inclusion and representation of all populations with access to health IT-enabled tools, not just the largest or densest. Partnerships between institutions that have developed the infrastructure and tools needed to enable research participation and institutions that may not have the resources may enable broader-based representation of diverse patient populations and research.

4.2.6 Priority 9: Accelerate Integration of Knowledge at the Point of Care

There is consensus that it takes far too long for evidence from research findings to become embedded at the point of care.[30] The lack of a set of standardized mechanisms to digitize knowledge into computable formats and integrate it into the health IT infrastructure to be used at the point of care represents a significant portion of that delay. Without further investigation into the methods, processes, and tools needed for effective and efficient translation of research findings back to those who need it most—patients and their providers—new knowledge and discoveries will be caught between the research ecosystem and the point of care.

Priority 9: Supporting Strategies

- Advance new methods to accelerate the digitization of evidence into computable knowledge
- Develop tools to support the translation of computable knowledge at the point of care supporting providers and patients

The 2011 report by NAS envisioned opportunities for open research systems in which researchers and clinicians worked more closely with one another. It discussed the creation of an information commons supported by a broad knowledge network of researchers and considered a future in which tools supported access to established evidence as it is synthesized across multiple studies. Various collaborative functionalities embedded into the health IT infrastructure could be developed to support the type of information sharing and consensus-building that would establish a trusted communication infrastructure around the development of knowledge.

5. Actions Needed to Realize the Agenda

This chapter outlines specific steps, or actions, that can be taken to advance the priorities outlined in the preceding chapter. When appropriate, specific stakeholders are suggested to further support each activity.

Each call to action outlined below includes a set of specific activities and steps that need to be taken to address relevant data, tools, or infrastructure needs. Each of these activities has been classified into one or more of the following areas:

- Collaboration
- Demonstrations and Pilots
- Education and Communication
- Policy Levers
- Access to Tools and Services
- Research and Evaluation
- Standards
- Tool Development

The recommended actions will require collaborative partnerships with multiple stakeholders, such as:

- Educational institutions
- Federal partners
- Foundations
- Health care provider organizations
- Health IT developers
- IT sector
- Patient advocacy groups
- Payors
- Researchers
- Research funding entities
- Standards development organizations (SDOs)

Specific initiatives, agencies, or organizations that could be engaged within these stakeholder groups are noted below.

5.1 PRIORITY 1: IMPROVE DATA QUALITY AT THE POINT OF CAPTURE

5.1.1 Identify and develop metadata standards that capture more information about a given data point at the time of capture

Action Category	Description	Partners
Standards	Coordinate a multi-stakeholder effort that includes providers, payors, and researchers to develop a list of specific high-priority metadata elements for capturing high-fidelity information at the point of care and align with findable, accessible, interoperable, and reusable (FAIR) data principles.	Health care provider organizations, health IT developers, payors, researchers, research funding agencies, standards development organizations
Policy Levers	Present metadata standards to the Health IT Advisory Committee (HITAC).	Federal partners

The FAIR Data Principles[31] identify specific metadata-related requirements to establish findability, accessibility, interoperability, and reusability to ensure that data are useful to researchers, and specifically address the need for both humans and their machines to be able to use data. Although the scientific community largely supports the FAIR principles, these suggested requirements are unlikely to lead to relevant data sharing on their own.[15] The principles must be supplemented with policies and incentives to encourage their use within both the clinical and research communities and take into account normative considerations (such as transparency when reusing data, updating to the latest standards, and ensuring privacy and data protection). Further complicating the environment, no single entity is responsible for establishing agreed-upon standards for the FAIR principles—which range from establishing a persistent identifier (PID), to use of common vocabularies and communications protocols, to modifiable provenance information.

The ONC Health IT Certification Program supports the development of requirements to promote interoperability of certified health IT products. While each major update of the certification criteria includes more robust technical and interoperability requirements, these are largely focused on the development of standard formats and methods of exchanging the data. Although health IT products are required to conform to several standards to achieve more robust and consistent exchange of information, support the creation of a standardized common clinical dataset, and make that information available to patients, the requirements around individual data elements are limited. It is not within the scope of the Certification Program to dictate the use of specific standards outside of the development of the documentation currently required to demonstrate data exchange.

ONC also maintains the Interoperability Standards Advisory (ISA), which coordinates activities around interoperability standards and implementation specifications. Their focus explicitly includes topics of interest to public health and health research interoperability. Apart from the Health IT Certification Program, the ISA is designed to inform “what” standards and implementation specifications can be used to address an interoperability need.

5.1.2 Promote the adoption and use of current and emerging data and metadata standards to improve data quality for care and research

Action Category	Description	Partners
Collaboration	Work with other organizations that fund research to incent the use of advancements in new data concepts and metadata as they emerge, including FHIR-based metadata standards and metadata that support FAIR data principles.	Federal partners, foundations, research funding entities
Demonstrations and Pilots	Support demonstration projects or pilot testing of highest value emerging data and metadata elements to investigate their impact on research-quality health data.	Researchers, health IT developers, standards development organizations

The Leading Edge Acceleration Projects (LEAP) in Health IT funding opportunity is an example of ONC's ability to seek out and fund projects that seek to overcome challenges that inhibit the development, use, and/or advancement of well-designed, interoperable health IT. Past areas of focus for funding include expanding the scope, scale, and utility of population-level data-focused APIs; advancing clinical knowledge at the point of care; standardization and implementation of scalable FHIR consent resources; and design, development, and demonstration of enhanced patient engagement technologies for care and research.

NIH also recognizes the importance of and challenges to storing, managing, and publishing research data in their *Strategic Plan for Data Science*.[9] To improve data quality at the point of capture, the goal of the NIH plan is to promote and refine standards and vocabularies as well as standards for data collection. Mentioned previously, a recent call from NIH encourages researchers to use FHIR standards to capture, integrate, and exchange clinical data for research purposes. AHRQ has recently followed with a similar notice of their own, displaying a trend by research funding organizations to expect researchers to begin investigating the use of these emerging tools and standards in their work.

5.2 PRIORITY 2: INCREASE DATA HARMONIZATION TO ENABLE RESEARCH USES

5.2.1 Increase support for the development and use of existing common data models to transform and analyze data for research purposes

Action Category	Description	Partners
Collaboration; Standards	Work across agencies that fund biomedical and health services research to create incentives for researchers to use and share common data models.	Federal partners, researchers, research funding entities
Education; Communication	Work across agencies that fund biomedical and health services research to institute a mechanism to communicate updated information about new and emerging data models to study section reviewers.	Federal partners, research funding entities

There are numerous common data models that standardize EHR data for use in research. For example, *All of Us* is standardizing EHR data provided by participating healthcare provider organizations using the OMOP data model. Other large-scale research networks, such as Sentinel and PCORnet, have also

established common data models. The challenge is that each model is developed for the fit and purpose of an organization or network's research needs, making interoperability a challenge. However, the quest for CDMs that are both robust and scalable holds potential if they are designed and implemented with both data source and intended use in mind. A single CDM that maps clinical data elements across various EHR systems would be a large-scale project but is possible given the examples already developed and used in research settings.

OHDSI[32] is a consortium that leverages the OMOP CDM to study large datasets to detect drug effects observed using EHR data. Using a distributed data model and standardized queries executed in the local environment of a data holder, OHDSI participants can leverage tools and approaches designed by a community of individuals and organizations to perform healthcare data analytics using a variety of data types and sources. OMOP uses standardized clinical data, health systems data (e.g., providers, care sites), health economics data (e.g., claims data, cost data), and several OMOP-derived elements (e.g., cohort, drug era) to provide a consistent, standardized dataset.

5.2.2 Identify collaborative opportunities to improve understanding regarding research data use and reuse in accordance with established privacy and security safeguards

Action Category	Description	Partners
Access to Tools and Services	Investigate funding to support a large-scale (national, centralized) "research workbench" platform, which will allow researchers to share data elements, crowdsource and converge on data models, share tools for data extraction and cleaning, a place to promote existing services, provide peer-to-peer support, and disseminate ideas related to analysis methods and predictive models used to support research.	Federal partners, researchers, research funding entities

There are various initiatives designed to provide researchers a space to share research data as well as tools and services needed to acquire and use data. The Center for Digital Health at the Icahn School of Medicine at Mount Sinai is one such example, with others of varying size and scale emerging. A single, centralized "workbench" of resources, such as one that might conceivably develop under or in conjunction with the NIH Data Commons pilot, would provide researchers a space that is both neutral and open but also curated and peer-reviewed where information critical to developing and using various standards could be shared, along with the data, tools, and services needed to support the research ecosystem as a whole.

5.3 PRIORITY 3: IMPROVE ACCESS TO INTEROPERABLE ELECTRONIC HEALTH DATA

5.3.1 Ensure health IT systems provide sufficient documentation about their data models and technical specifications to develop shared tools for acquiring clinical data from those systems

Action Category	Description	Partners
Policy Levers	Define specific requirements for certified systems to make sure that technical specifications are publicly available so researchers can understand how data points are represented.	Federal partners
Collaboration	Work with advocacy groups to encourage health IT developers to make available APIs and accompanying specifications, and design of their EHR systems to enable access by other systems. These specifications should be open and publicly available, including transparency on what data exists and how to connect workflows for data extraction.	Health IT developers
Demonstrations and Pilots	Direct pilot studies to test the effectiveness of schema publication and API expansion toward small and midsize health IT developers.	Health IT developers, research funding agencies

To improve the usefulness of the information exchanged through the open API environment, the Argonaut Project has encouraged standardization of a subset of information being exchanged using the FHIR specification. This FHIR-based API and Core Data Services specification has enabled a rapid expansion of applications that can access and provide data with sufficient authorization and consent to end users, including patients, providers, and researchers. Additionally, the SMART project is an open-source initiative that seeks to support developers in building applications that run seamlessly across EHR instances and use the FHIR API and resource definitions.[33]

While still in the process of rulemaking, the ONC proposed regulation to improve interoperability and decrease information blocking takes important steps forward. Not only does the proposed rule include an update to the EHR certification requirement for API usage to require the FHIR standard for exchanging information, it also requires the systems to support export for bulk or population-level data extractions in addition to single-patient data extraction. The proposed criteria do allow health IT developers the flexibility to determine the export standard if the export file is computable and includes documentation that is sufficient for the end user to interpret and use the electronic health data included in the extract. The documentation must be made publicly available.

Although these activities continue to improve functional access and interoperability within the healthcare landscape, the registration process for third-party applications developed to support extraction and integration of data is not standardized between health IT developers. Third-party application developers are finding it difficult and time-consuming to meet the different requirements and workflows necessary to connect to multiple vendor systems. It will be important to continue supporting these requirements as the rule is finalized and closely monitor how they are implemented within the developer community.

5.4 PRIORITY 4: IMPROVE SERVICES FOR EFFICIENT DATA STORAGE AND DISCOVERY

5.4.1 Realize efficiencies by making advanced computational capacity and storage available to researchers to reduce redundant data collection efforts

Action Category	Description	Partners
Access to Tools and Services	Support dialogue with commercial companies to establish low-cost or free access for researchers to tools and services developed by private industry in support of increased computational capacity, data storage, and other services that enable access to data.	IT sector
Collaboration	Regular briefings of federal agencies across and outside of HHS that are developing high-capacity data storage and analysis services geared toward researchers to discuss developments and collaborate on specific opportunities.	Federal partners
Collaboration	Coordinate with NIH Data Commons Pilots focused on data sharing to incorporate current standards to facilitate FAIR principles and continue support for activities specified under the NIH *Strategic Plan for Data Science*, including standards issues related to storing, managing, and publishing data.	Federal partners
Demonstrations and Pilots	Support demonstration and pilot projects that use novels methods for identifying data and make data discoverable.	Federal partners, health IT developers, researchers

As data are shared from one source to another, it is important to understand the provenance of the data. Unless there is a concerted effort to maintain provenance as data are moved around, such information is usually lost. Currently, there is a lack of appropriately identified provenance information embedded in the data captured across various data sources. This includes data from EHR systems; data collected and housed in external databases, such as data coordinating centers; and PGHD collected through sensors, wearables, mobile applications (apps), and other mechanisms. The lack of provenance information limits the ability for data to be routinely shared and reused for research.

The National Science Foundation (NSF) recently funded researchers at Indiana University to develop a PID for research data to make digital data discoverable. The *Enhanced Robust Persistent Identification of Data* (Enhanced RPID) is a compilation of software tools to define common operations and perform digital object mapping to data that has been collected for research that will allow management of these data by assigning individual identifiers to make data FAIR, and will include technical and educational components, including a testbed with corresponding scientific use cases. When completed, the testbed aims to facilitate better access to and utilization of existing data repositories.

It is critical to encourage organizations that have made large-scale investments in computational power balance monetary gains with the potential to vastly improve research and discovery. Researchers from smaller organizations and provider settings often lack the resources to implement and maintain such solutions and could benefit from the availability of stand-alone applications to query and analyze clinical data for research purposes.

Under the recently launched the Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative, NIH partnered with Google Cloud and AWS to help reduce economic and technological barriers to accessing and computing with large biomedical datasets.[34] This initiative will allow NIH-supported projects to leverage cloud computing, machine learning, and storage capabilities. This initiative will also involve collaborations with the NIH Data Commons Pilot that focuses on data sharing to incorporate current standards to facilitate FAIR principles. The NIH Data Commons provides a cloud-based platform where investigators can store, share, access, and interact with digital objects generated from biomedical and behavioral research. The pilot developed a number of computational and digital tools focused on improving researchers' access to and analysis of data.

Additionally, the VA partnered with the Department of Energy (DoE) to establish the MVP Computational Health Analytics for Medical Precision to Improve Outcomes (MVP-CHAMPION) initiative, which will leverage supercomputers from the DoE to enable large-scale analysis of VA data and support the MVP. The VA also maintains the VA Informatics and Computing Infrastructure (VINCI) initiative, which is focused on providing researchers with access to VA data. VINCI continues to improve on tools and workspace services to support research and analysis of the data.

5.5 PRIORITY 5: INTEGRATE EMERGING HEALTH AND HEALTH-RELATED DATA SOURCES

5.5.1 Support functionality within the health IT architecture to link research-relevant data sources outside the patient care setting with EHR data

Action Category	Description	Partners
Standards; Collaboration	Collaborate across federal partners to support standardization of taxonomies and methods for collecting data across a wide variety of settings and purposes related to improved research capabilities (e.g., work with FDA on standards related to medical devices and consumer-grade wearable devices used for PGHD).	Federal partners
Policy Levers	Work to determine the feasibility of informing reimbursement rules tied to the availability of standardized real-world and patient-reported data.	Federal partners
Standards	Continue working to support the process of standardizing specific data elements related to precision medicine; revisit the 2015 Health IT Standards Committee recommendations to ensure they are kept up to date.	Federal partners
Tool Development	Support the development and/or dissemination of advanced tools and third-party applications that improve data quality for researchers (e.g., data mining and cleaning) and reduce provider burden through the Innovation Challenge and/or Leading Edge Acceleration Projects (LEAP) in Health IT programs at ONC.	Health IT developers, IT sector, researchers

Research investigating the effect of social and environmental factors on health has a long history; however, new capabilities to identify and synthesize various disparate data sources on a large scale are emerging. From medical devices to patient-reported data, genomic discovery to socioeconomic factors,

this is a rapidly moving field with a growing number of initiatives and interested stakeholders in the research and clinical communities.

In recent years, ONC, in partnership with NIH, launched Sync for Science (S4S), S4S Privacy and Security, Sync for Genes, and Advancing Standards for Precision Medicine. S4S is a collaboration among health IT developers, healthcare organizations, ONC, NIH, and researchers to develop and pilot read-only APIs that permit patients to direct the sharing of EHR data with the *All of Us* Research Program's clinical data repository for research purposes. While this work has been critical to convening activity around a number of important advancements to support the provision of data for research, the work led by ONC in this area should continue to advance, along with advancements in the architecture and our understanding of what sources of data may be relevant to the research community.

As research increasingly looks to incorporate genetic data, continuing work similar to the Electronic Medical Records and Genomics (eMERGE) Network[35] will be important. eMERGE is an NIH-funded consortium of five institutions with DNA data linked to EHRs that assessed the utility of EHRs as a consistent and reliable source of phenomic data. Their work has shown that high-quality EHR-derived phenotypes required free text in addition to clinical codes, laboratory-medicine results, and medication histories. Natural language processing of physician comments was essential to get high predictive values.

In addition, NAS supported a workshop in 2018, "Examining the Impact of Real-World Evidence on Medical Product Development,"[36] which identified several challenges when health IT used in care delivery, payment, and operations is leveraged for research. Real-world evidence derived from real-world data is especially attractive among drug and device makers because it may help streamline the regulatory approval process. There would be significant benefits in working closely with FDA to better understand the architecture requirements needed to advance the use of EHR data to support activities such as medical device surveillance and the use of real-world data to support a more rapid, safety-focused research-to-market pipeline.

The development of the technical architecture needed to routinely incorporate data generated from medical devices into the clinical and research ecosystem could provide access to a critical data source, which is currently largely separated from both. In 2018, the American Medical Association launched a challenge competition sponsored by Google focused on bringing PGHD from a device into use during a clinical care visit, and data from the visit back to the device. This type of activity provides real-world experience in understanding which components of the current health IT architecture can be leveraged and which are still needed to demonstrate data transfer in the care setting.

Once the necessary architecture and/or standards are in place, it will be important to promote the use of these standards for data sources other than the EHR to interact more seamlessly with the current architecture.

5.5.2 Provide support for accelerating the process of standardizing new data concepts while working to update current standards

Action Category	Description	Partners
Standards	Support activities to improve more rapid-consensus based standardization of data concepts, such as providing additional financial support to research consortia and other similar groups that currently carry the burden of developing consensus on standardization of rapidly developing data elements and support the lengthy process of interfacing with SDOs to achieve new standards.	IT sector, researchers, standards development organizations

ONC has consistently supported the standards development process, maintaining close ties with HL7 and other relevant international standards bodies. ONC has sought to meet the standards development needs inherent to the particular context and evolution of the health IT infrastructure. In recent years, ONC has increased their support for the standards development process, in particular the development and testing of FHIR-based implementation guides, Connectathons, and the process of taking a draft standard through the ballot process. The ISA, in particular, provides support to convene and rapidly prioritize health IT challenges and subsequently develop and harmonize standards.

Despite these continued efforts, the standards development process remains lengthy and time consuming. While the establishment of standards must be grounded in consensus and testing, providing resources to those performing these functions could lead to a noticeable improvement in the ability for SDOs to respond more rapidly to the need to identify and implement standards as they emerge.

5.6 PRIORITY 6: IMPROVE METHODS AND TOOLS TO SUPPORT DATA AGGREGATION

5.6.1 Improve the ability to match individuals to different sources of data

Action Category	Description	Partners
Policy Levers	Provide leadership and direction to develop and implement universal patient matching and identification methods.	Federal partners, IT sector
Research & Evaluation	Assess the participant enrollment and consent infrastructure developed under the *All of Us* Research Program to determine the feasibility of leveraging and expanding the functionality to other research initiatives.	IT sector, researchers
Demonstrations and Pilots	Identify pilot or demonstration funding to establish the evidence for participant matching to accelerate private industry advancements with patient matching solutions.	Health IT developers, researchers, research funding entities

Although lifting the congressionally imposed restriction on using federal funds to research the use of a unique patient identifier was proposed in June 2019, ONC has investigated alternate solutions to patient matching during the decades-long ban. These include initiatives such as the Patient Matching, Aggregating, and Linking (PMAL) project,[37] which provided support for a white paper[25,26] and real-world

application of solutions via the ONC-led Patient Matching Algorithm Challenge.[38,39] The winners of the challenge utilized referential matching, which was shown to increase matching accuracy beyond the current industry standard. The 21st Century Cures Act instructed the U.S. Government Accountability Office (GAO) to expand their patient matching work to include the costs and risks of patient mismatches. In January 2019, the GAO published a report on patient matching, *Approaches and Challenges to Electronically Matching Patients' Records across Providers,*[1] as directed by the 21st Century Cures Act. In this report, GAO describes (1) stakeholders' patient record matching approaches and related challenges; and (2) efforts to improve patient record matching identified by stakeholders.

Following this publication, ONC published a request for information (RFI) on patient matching as part of the ONC 21st Century Cures Act proposed rule published to the *Federal Register* in February 2019. In the RFI, ONC sought comment on additional opportunities that may exist in the patient matching space and ways that ONC can lead and contribute to coordination efforts with respect to patient matching. ONC noted particular interest in ways that patient matching can facilitate improved patient safety, better care coordination, the quality of care, and advanced interoperability. ONC intends to review the responses to the RFI in concert with the GAO report, to further inform continued efforts to advance patient matching building on the lessons learned from the PMAL project.[40]

Various private industry initiatives have also contributed to the development of a patient matching solution. The College of Healthcare Information Management Executives (CHIME) supported a 2-year National Patient ID Challenge in an effort to incent improvements in patient matching, but ended the program in late 2017, noting that approaches utilizing data algorithms had greater potential to move solutions forward within the healthcare industry. The Pew Charitable Trust released their report on patient matching in late 2018, which reviewed multiple options, including unique identifiers, patient-directed solutions, improved standardization of demographic data, and referential matching.[41] The report also noted the overall need for a nationwide strategy and identified the need for a stewardship entity, such as the Recognized Coordinating Entity (RCE) under the proposed Trusted Exchange Framework and Common Agreement (TEFCA). As proposed by ONC, this framework would provide a set of conditions and terms that will overcome the various governance issues that have prevented widespread nationwide exchange of health data and serves as a critical counterpart to the technical advances and standardization needed to achieve interoperability. Although the TEFCA does not currently include research as a permitted purpose, the research community may want to leverage the framework for research purposes.

5.6.2 Develop tools to efficiently manage data use agreements across organizations

Action Category	Description	Partners
Research and Evaluation	Investigate the ability to electronically manage data use agreements between parties.	Health IT developers, researchers, healthcare provider organizations

DUAs are used by institutions as a mechanism to outline the parameters of data use when data are transmitted between parties. While typically used in the context of treatment, payment, and operations, institutions collaborating in research projects or that join research networks may use DUAs to establish important safeguards regarding who can access the data and what they can be used for. While these agreements are critical to ensuring appropriate use of data, a more robust health IT infrastructure could

enable better centralized management of these agreements. Such an approach could also facilitate extending use to new partners who agree to adhere to the requirements of the initial DUA.

5.6.3 Develop functionalities needed to manage data across distributed sources, including to identify redundancy; account for updates to data and metadata; and analyze data in different formats

Action Category	Description	Partners
Tool Development	Support the development of shared services for de-duplication of records and to manage updates to data and metadata.	Health IT developers, researchers, federal partners, foundations

A future state health IT infrastructure should enable the development of and access to a variety of tools that would support the process of aggregating data in both clinical and research settings. Examples of these tools include advanced de-duplication services and electronic management of updates to data standards and metadata.

5.7 PRIORITY 7: DEVELOP TOOLS AND FUNCTIONS TO SUPPORT RESEARCH

5.7.1 Support easier consent management for research

Action Category	Description	Partners
Policy Levers; Collaboration	Monitor efforts by Office for Civil Rights (OCR) to clarify guidance around HIPAA and the right of access of data.	Federal partners
Research and Evaluation	Determine what steps are needed to identify a trusted eConsent management organization.	Federal partners, IT sector
Research and Evaluation	Support a study of patients to identify the functionalities desired in an eConsent system and the level of support for centralized management of that information.	Health IT developers, IT sector, researchers
Research and Evaluation	Investigate infrastructure requirements and standards needed for consent managements systems that allow multi-tiered, alterable consent preferences.	Health IT developers, IT sector, SDO members
Demonstrations and Pilots	Work with current eConsent application developers to assess the feasibility of a centralized solution; identify specific infrastructure requirements that need to be established to execute solution.	Health IT developers, IT sector

Within the current healthcare environment, individual-level permissions management of health information is limited, resulting in a patient's data being stored in silos across multiple organizations. Health systems often require a paper-format patient consent before data can be shared or used for research, and the wide range of institutions with customized EHRs has resulted in silos of records for patients. Patients typically can individually share portions of their health data with multiple systems, but those systems rarely have the infrastructure to extract, reuse, or match data with one another. As a result,

a patient's data are distributed across multiple EHRs and institutions. Blue Button is one initiative from ONC that allows patients to view, download, and share their whole record through structured and standardized data formats; however, this feature is only available through a certain number of federal agencies, health plans, and personal health record vendors.

Consent processes continue to grow more sophisticated as initiatives to access and share information. The *All of Us* Research Program has employed an interactive consent process that uses modules and videos to ensure the highest level of informed consent, but also offers a paper form as an alternative. ONC has also supported steps forward related to eConsent, including the recent Move Health Data Forward Challenge and the 2019 LEAP in Health IT funding opportunity priority area supporting standardization and implementation of the FHIR Consent Resource.

5.7.2 Develop additional tools to support research processes such as recruitment, enrollment, randomization, and HIPAA-compliant de-identification

Action Category	Description	Partners
Demonstrations and Pilots	Provide support for developers, researchers, and private industry to test advancements of artificial intelligence (AI)-driven or other leading-edge tools that integrate seamlessly into the health IT architecture and build confidence in de-identification, identification of research cohorts, and other tools targeted to facilitating research activities.	Federal partners, health IT developers, IT sector, researchers

The application of AI deep learning techniques to automated de-identification would be an invaluable tool to researchers. Recent work has shown promise in using deep learning for information extraction, representation learning, outcome prediction, phenotyping, and de-identification.[42] While these technologies are still in early stages of development, they provide support for identifying benchmarks that would show that automated de-identification of EHR data satisfies HIPAA requirements.

There also continues to be a rise in using advanced statistical analyses or machine learning to identify research cohorts based on criteria such as phenotype. While these tools are growing in popularity and sophistication, efforts to further develop and target availability to researchers would improve their overall impact within the research ecosystem, especially as other priorities related to standards and data availability continue to mature.

5.7.3 Investigate and expand tools that index, search, and query systems to identify and recruit possible patient cohorts for a given study as well as easily extract data about participants

Action Category	Description	Partners
Research and Evaluation	Support a study investigating the tools and services currently used to support indexing and query functions within distributed data networks; identify any gaps and challenges.	Researchers, research funding entities

Action Category	Description	Partners
Access to Tools and Services	Promote development of tools that provide solutions to identified gaps and support the distribution of indexing and query tools widely within research ecosystem.	Federal partners, health IT developers, researchers, research funding entities

An ideal health IT infrastructure would support the development of a centralized database, or similarly indexed set of databases in which patient data and its provenance could be both highly protected by advanced security techniques, such as blockchain, and available to researchers given appropriate authorizations and adherence to guidelines. There are many larger-scale research models that can provide important insight into both the challenges and the tools available to support advanced query tools within a distributed research network, such as the FDA Sentinel program and PCORnet. OHDSI's query capability and use of a common data model to generate real-world data for use by not only researchers, but many other stakeholders, is an example of an important approach that researchers and patients anticipate the health IT infrastructure will support.

It is essential to develop these tools in conjunction with the appropriate security considerations, access and authorization policies, and sufficient architectural functionalities to prevent unauthorized access to the data, which would cause certain harm to the fabric of trust within the patient population needed to generate support for a more advanced research ecosystem driven by the health IT infrastructure.

5.8 PRIORITY 8: LEVERAGE HEALTH IT SYSTEMS TO INCREASE EDUCATION AND PARTICIPATION

5.8.1 Develop health IT tools that deliver value for providers and patients to participate in research

Action Category	Description	Partners
Education and Communication	Improve education and engagement around participation in research through a coordinated communications campaign that is embedded as a component of the services and tools that patients already use to access their health data (e.g., patient portals).	Researchers, patient advocacy groups
Research and Evaluation	Support a focused study to assess effective methods to add value for patients who participate in research.	IT sector, patient advocacy groups, researchers
Research and Evaluation	Support research on evidence-based methods for delivering information to patients to inform improvements and updates to patient portals and other patient health record systems.	Researchers

The NIH-led *All of Us* Research Program promotes ongoing participant involvement throughout the research process, allowing program participants access to study results along with summaries of their data. The program also uses a "Participants as Partners" model, engaging participants to provide input on

participant retention, privacy, security, and how best to return information to participants in a meaningful way.

5.8.2 Pursue infrastructure improvements that enable participation from a diverse patient population

Action Category	Description	Partners
Education and Communication	Co-sponsor and/or directly launch communications and outreach campaigns that utilize health IT-enabled applications to notify the public of research opportunities to promote diversity and drive engagement.	Health IT developers, IT sector, researchers

As a key part of the *All of Us* Research Program, NIH has specifically requested that partner organizations, or healthcare provider organizations, support recruitment of a diverse population of participants. This includes diversity in age, genetic background, and lifestyle, and an intentional focus on those who have traditionally been underrepresented in biomedical research.[43] To conduct outreach, *All of Us* uses multiple methods of targeted advertising, personal interest groups, healthcare providers, and volunteers to notify potential participants and achieve diversity.

5.8.3 Expand research opportunities beyond large health systems

Action Category	Description	Partners
Education and Communication	Support workforce development and capacity-building programs to increase the availability of in-demand capabilities (e.g., data scientists); support for certification programs in health data analysis across the healthcare sector.	Educational institutions
Access to Tools and Services	Lower barriers to entry for organizations without a robust research infrastructure by making the tools and processes for using the health IT infrastructure for research more accessible, including availability of shared technical resources (e.g., plug and play apps) and organization-to-organization mentoring.	Educational institutions and healthcare provider organizations

Various federal agencies fund research that is specifically targeted to specific populations or patient groups. The Health Resources and Services Administration (HRSA) provides significant funding through the Federal Office of Rural Health Policy to support improved access to care in populations that often do not have easy access to large academic medical centers and hospitals, and can serve as a partner in engaging patients served by HRSA-funded institutions in research.

AHRQ and PCORI work closely to support various patient engagement and dissemination projects, which aim to improve the process by which researchers can better share their findings with intended end users,[44,45] as well as countless tools and documents intended to reach providers and patients with research findings. While these tools are most often intended to reach a wide audience, there is little tracking or evidence regarding whether these materials are reaching outside the larger health system

community. Funding opportunities for research across most federal agencies outside of HRSA do not often target smaller or less-resourced institutions. Without access to the expertise or experience often seen in large health systems, opportunities to participate in research are often limited. The ONC Workforce Training to Educate Health Care Professionals in Health Information Technology program, funded under the HITECH Act in 2009, may prove to be a useful model.

5.9 PRIORITY 9: ACCELERATE INTEGRATION OF KNOWLEDGE AT THE POINT OF CARE

5.9.1 Advance new methods to accelerate the digitization of evidence into computable knowledge

Action Category	Description	Partners
Collaboration	Collaborate across federal agencies to fund solutions that support digital knowledge that is standardized and scalable across the architecture, especially those that have strong visibility or positive impact on the population (e.g., a national alert system for public health emergencies).	Federal partners, researchers, standards development organizations
Collaboration	Federal agencies and private organizations that fund research can collaborate to incent researchers to develop digital knowledge of evidence-based findings.	Federal partners, foundations, research funding entities

In 2018, an HL7 workgroup was approved to investigate the application of FHIR-based resources to support Evidence-Based Medicine (EBM) Knowledge Assets (EBMonFHIR). Their goal is to provide interoperable standards for "producing, analyzing, synthesizing, disseminating and implementing clinical research (evidence) and recommendations for clinical care (clinical practice guidelines, or CPG)."[46] Expanding on this work, in early 2019, a CPG-focused HL7 workgroup gained approval that focused on developing FHIR specifications related to the Centers for Disease Control and Prevention Adapting Clinical Guidelines for the Digital Age initiative. The goal of this effort is to connect research and evidence swiftly and accurately to those who need it most. Additional related use cases should be considered, particularly by those who fund research in the area of clinical guidelines.

In conjunction with NIH, researchers at the University of Michigan have developed a group reviewing legal and technical challenges to Mobilizing Computable Biomedical Knowledge (MCBK). Participants in the MCBK group include representatives from across the research and academic field, various federal agencies and initiatives, and private industry.

5.9.2 Develop tools to support the translation of computable knowledge at the point of care supporting providers and patients

Action Category	Description	Partners
Research and Evaluation	Investigate the functionality and effectiveness of bi-directional FHIR-based tools that leverage previous developments in the area of clinical decision support.	Federal partners, foundations, researchers, health IT developers, research funding entities

The link between evidence generation and clinical practice has been theorized but has suffered from a lack of large-scale implementation and evaluation. Additionally, there have been numerous research studies focused on the development of clinical decision support tools and artifacts. There continues to be a need for health IT to support functions that easily incorporate computable knowledge into practice through decision support triggers and rules and API tools with third-party functions to both read and write to clinical systems. With the rapid growth of FHIR-based tools and API access, health IT developers could be encouraged to build on previous initiatives, such as CDS Hooks, to expand and demonstrate the ability to effectively reduce the length of time between scientific discovery and implementation in the clinical setting.

6. Conclusion

The increased collection of electronic health data and investment in health IT infrastructure over the past decade have created unprecedented opportunities for biomedical and health services research. However, challenges to achieving these advances in research have inhibited progress. Data quality is inconsistent, and there is an unmet need for tools and services to support researchers in finding, interacting with, and managing data across disparate sources. Solutions are needed that speed the pace of discovery and the ability of research to inform care delivery and policy.

This effort undertook an important first step by identifying challenges for health service providers and biomedical researchers in obtaining and using health data to support scientific exploration. The Policy and Development Agenda outlines a set of priorities and actions that can be taken by relevant stakeholders to advance those priorities.

References

1. Parasrampuria S, Henry J. Hospitals' use of electronic health records data, 2015-2017. Washington, DC: Office of the National Coordinator for Health Information Technology; 2019. ONC Data Brief, no. 46.

2. HL7 International. Welcome to the Argonaut Project. 2019. http://argonautwiki.hl7.org/index.php?title=Main_Page. Accessed July 24, 2019.

3. JASON at the MITRE Corporation under Contract No. JSR-13-700. A robust health data infrastructure. Rockville, MD: Agency for Healthcare Research and Quality; 2014.

4. Amoozegar J, Blumenfeld B, Brown S, et al. RTI International under Contract No. HHSP23320150039I. Building data capacity for patient-centered outcomes research in HHS: a formative evaluation of 2012-2016 projects. Washington, DC: Office of Health Policy, U.S. Department of Health and Human Services; 2017. https://aspe.hhs.gov/system/files/pdf/259016/ASPEPCORTFEvaluation.pdf.

5. National Academy of Medicine. The future of health services research: advancing health systems research and practice in the United States. Washington, DC: National Academy of Medicine; 2018. https://nam.edu/wp-content/uploads/2018/11/The-Future-of-Health-Services-Research_web-copy.pdf.

6. National Committee on Vital and Health Statistics. Health information privacy beyond HIPAA: a framework for use and protection. Washington, DC: U.S. Department of Health and Human Services; 2019. https://ncvhs.hhs.gov/wp-content/uploads/2019/07/Report-Framework-for-Health-Information-Privacy.pdf.

7. National Institutes of Health, Office of the Director. Fast Healthcare Interoperability Resources (FHIR®) Standard. NOT-OD-19-122; 2019 July 30. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html.

8. National Institutes of Health. Notice of special interest: Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) applications directed at the adoption of the Fast Healthcare Interoperability Resources (FHIR®) Standard. NOT-OD-19-127; 2019 July 30. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-127.html.

9. National Institutes of Health, Office of Data Science Strategy. NIH Strategic Plan for Data Science. 2018. https://datascience.nih.gov/strategicplan. Accessed September 17, 2019.

10. Zayas Cabán T, Wald J, White J, Brennan P. National health IT priorities to advance research. AMIA 2018 Annual Symposium; 3-7 Nov 2018; San Francisco, CA.

11. Zayas Cabán T, Chaney K, Kho A, Mandel J. National health IT priorities to advance research. AMIA 2019 Informatics Summit; 25-28 March 2019; San Francisco, CA.

12. Schneeweiss S, Brown JS, Bate A, et al. Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products. *Clin Pharmacol Ther* 2019. doi: 10.1002/cpt.1577

13. Ogunyemi OI, Meeker D, Kim HE, et al. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. *Med Care* 2013;51(8 Suppl 3):S45-52. doi: 10.1097/MLR.0b013e31829b1e0b

14. Infectious Diseases Data Observatory (IDDO). About IDDO. 2019. https://www.iddo.org. Accessed July 24, 2019.

15. Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? *Eur J Hum Genet* 2018;26(7):931-936. doi: 10.1038/s41431-018-0160-0

16. Pisani E, Aaby P, Breugelmans JG, et al. Beyond open data: realising the health benefits of sharing data. *BMJ* 2016;355(5295):i5295. doi: 10.1136/bmj.i5295

17. Yale University Open Data Access. The YODA Project. http://yoda.yale.edu. Accessed July 24, 2019.

18. Evans BJ, Krumholz HM. People-powered data collaboratives: fueling data science with the health-related experiences of individuals. *J Am Med Inform Assoc* 2018. doi: 10.1093/jamia/ocy159

19. HealthIT.gov. Advancing Standards for Precision Medicine. 2019. https://www.healthit.gov/topic/advancing-standards-precision-medicine. Accessed September 17, 2019.

20. SIREN. The Gravity Project: a social determinants of health coding collaborative [version 1.3]. 2019. https://sirenetwork.ucsf.edu/sites/sirenetwork.ucsf.edu/files/wysiwyg/Gravity-Project-Charter.pdf. Accessed September 17, 2019.

21. Kasthurirathne SN, Vest JR, Menachemi N, et al. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *J Am Med Inform Assoc* 2018;25(1):47-53.

22. National Research Council. Toward Precision Medicine. Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC: The National Academies Press; 2011.

23. Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. *J Am Med Inform Assoc* 2015;22(3):621-627. doi: 10.1093/jamia/ocu012

24. Rudin RS, Hillestad R, Ridgely MS, et al. Defining and evaluating patient-empowered approaches to improving record matching. *Rand Health Q* 2019;8(3):3.

25. Morris G, Farnum G, Afzal S, et al. Patient identification and matching final report. Baltimore, MD: Audacious Inquiry, LLC; 2014. http://journal.ahima.org/wp-content/uploads/ONC-Patient-Identification-Matching-Final-Report-February-2014.pdf.

26. Dimitropoulos LL, Grannis SJ, Banger AK, et al. Privacy and security solutions for interoperable health information exchange—perspectives on patient matching: approaches, findings, and challenges. Chicago, IL: RTI International; 2009. https://www.healthit.gov/sites/default/files/patient-matching-white-paper-final-2.pdf.

27. U.S. Department of Health and Human Services. Confidentiality of substance use disorder patient records (42 CFR 2 [SAMHSA-4162-20] RIN 0930–AA32). *Federal Register* 2019;84(165):44568-44589.

28. Lopez MH, Holve E, Sarkar IN, Segal C. Building the informatics infrastructure for comparative effectiveness research (CER): a review of the literature. *Med Care* 2012;50(Suppl):S38-S48. doi: 10.1097/MLR.0b013e318259becd

29. Office of the National Coordinator for Health Information Technology. Federal Health IT Strategic Plan 2015–2020. Washington, DC: U.S. Department of Health and Human Services; 2019. https://www.healthit.gov/sites/default/files/9-5-federalhealthitstratplanfinal_0.pdf. Accessed September 10, 2019.

30. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. *Yearbook of Medical Informatics 2000: Patient-Centered Systems* 2000;65-70.

31. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi: 10.1038/sdata.2016.18

32. Maier C, Lang L, Storf H, et al. Towards implementation of OMOP in a German university hospital consortium. *Appl Clin Inform* 2018;9(1):54-61. doi: 10.1055/s-0037-1617452

33. SMART. What is SMART? https://smarthealthit.org/an-app-platform-for-healthcare/about.

34. U.S. Department of Health and Human Services. Privacy impact assessment. 2016. https://www.hhs.gov/sites/default/files/cms-amazon-web-services.pdf. Accessed September 14, 2018.

35. Shirts BH, Salama JS, Aronson SJ, et al. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. *J Am Med Inform Assoc* 2015;22(6):1231-1242. doi: 10.1093/jamia/ocv065

36. National Academies of Sciences, Engineering, and Medicine. Examining the impact of real-world evidence on medical product development: I. Incentives: proceedings of a workshop—in brief. Washington, DC: The National Academies Press; 2018.

37. Office of the National Coordinator for Health Information Technology. Patient matching, aggregating and linking (PMAL). 2018. https://www.healthit.gov/topic/research-evaluation/patient-matching-aggregating-and-linking-pmal. Accessed July 24, 2019.

38. U.S. Department of Health and Human Services. HHS names patient matching algorithm challenge winners. 2017. https://www.hhs.gov/about/news/2017/11/08/hhs-names-patient-matching-algorithm-challenge-winners.html.

39. Sweeney E. Senators ask GAO to expand patient matching study. FierceHealthcare; 2017. https://www.fiercehealthcare.com/regulatory/elizabeth-warren-gao-patient-matching-21st-century-cures.

40. U.S. Government Accountability Office (GAO). Health information technology: approaches and challenges to electronically matching patients' records across providers. Washington, DC: U.S. Government Accountability Office; 2019. https://www.gao.gov/products/GAO-19-197.

41. The Pew Charitable Trusts. Enhanced patient matching is critical to achieving full promise of digital health records. Philadelphia, PA: The Pew Charitable Trusts; 2018. https://www.pewtrusts.org/-/media/assets/2018/09/healthit_enhancedpatientmatching_report_final.pdf. Accessed July 24, 2019.

42. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2018;22(5):1589-1604. doi: 10.1109/JBHI.2017.2767063

43. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214-223. doi: 10.1016/j.jclinepi.2015.09.016

44. Carpenter D, Nieva V, Albaghal T, Sorra J. Development of a planning tool to guide research dissemination. In: Henriksen K, Battles JB, Marks ES, Lewin D, eds. Advances in Patient Safety: From Research to Implementation (Volume 4: Programs, Tools, and Products). Rockville, MD: Agency for Healthcare Research and Quality; 2005.

45. Mathematica Policy Research. PCORI dissemination & implementation toolkit. Washington, DC: Mathematica Policy Research; 2015. https://www.pcori.org/sites/default/files/PCORI-Dissemination-Implementation-Toolkit.pdf.

46. HL7 International. EBMonFHIR. 2019. https://confluence.hl7.org/display/CDS/EBMonFHIR.

Appendix A. Workshop Participants

We acknowledge and thank the following individuals for their participation in the workshop.

Workshop Participants

Kenneth D. Mandl, MD, MPH (Keynote Speaker)
Director, Computational Health Informatics Program, Boston Children's Hospital

Natalie Boutin
Program Director, Partners HealthCare Biobank
Director, Information Technology at Partners HealthCare Personalized Medicine, Partners HealthCare

Patricia Flatley Brennan, RN, PhD, FACMI
Director, National Library of Medicine

Stephen Downs, SM, FACMI
Chief Technology and Strategy Officer, Robert Wood Johnson Foundation

Chris Dymek, EdD
Director, Division of Health Information Technology, Agency for Healthcare Research and Quality

Barbara Evans, PhD, JD, LLM
Director, Center for Biotechnology & Law, University of Houston Law Center

Matthew Harker, MPH, MBA
Director, Research Customer Solutions, Analytics Center of Excellence (ACE), Duke University

Tony Kerlavage, PhD
Director, Center for Biomedical Informatics and Information Technology, National Cancer Institute

Abel Kho, MD, MS, FACMI
Director, Institute for Public Health and Medicine—Center for Health Information Partnerships, Northwestern University

Sean Khozin, MD, MPH
Associate Director, Oncology Center of Excellence, Food and Drug Administration

Denys Lau, PhD
Division Director, National Center for Health Statistics, Centers for Disease Control and Prevention

Eve Maler
Vice President of Innovation and Emerging Technology, ForgeRock, Co-chair of OpenID's Health Relationship Trust (HEART) Workgroup

Josh Mandel, MD
Chief Architect for Microsoft Healthcare, Microsoft Corporation

Wendy Nilsen, PhD
Program Director, The Smart and Connected Health Program, National Science Foundation

Lucila Ohno-Machado, MD, MBA, PhD, FACMI
Chair, Department of Biomedical Informatics, Associate Dean for Informatics and Technology, UC San Diego School of Medicine

Rachel Ramoni, DMD, ScD
Chief Research and Development Officer, U.S. Department of Veterans Affairs

Raj Ratwani, MA, PhD
Director, National Center for Human Factors in Healthcare, MedStar Health

Rachel Richesson, MS, PhD, MPH, FACMI
Associate Professor, Duke University

Chesley Richards, MD, MPH, FACP
Deputy Director, Office of Public Health Scientific Services, Centers for Disease Control and Prevention

Josh Rising, MD, MPH
Director, Health Care Programs, The Pew Charitable Trusts

Joni Rutter, PhD
Director of Scientific Programs, *All of Us* Research Program, National Institutes of Health

Drew Schiller
Chief Executive Officer, Validic

Eric Schneider, MD, MSc
Senior Vice President for Policy and Research, The Commonwealth Fund

Joe Selby, MD, MPH
Executive Director, Patient-Centered Outcomes Research Institute

Naomi Tomoyasu, PhD
Deputy Director, Health Services Research & Development Service, U.S. Department of Veterans Affairs

Paul Varghese, MD
Head, Health Informatics, Verily

Appendix B. In-Person Workshop Agenda

ONC Technical Expert Workshop—National Health IT Priorities to Advance Research

Location: RTI International, Washington, DC

One Metro Center—701 13th Street, NW, Suite 750—Washington, DC 20005-3967

Day 1: Tuesday, July 24, 2018

8:15 am–9:00 am	ARRIVAL	All
9:00 am–9:20 am	WELCOME, INTRODUCTION and OPENING REMARKS	**Jonathan Wald**, *Project Director, RTI* **Teresa Zayas-Cabán**, *Chief Scientist, ONC* **Jon White**, *Deputy National Coordinator for Health Information Technology, ONC*
9:20 am–10:00 am	KEYNOTE	**Kenneth D. Mandl**, *Director, Computational Health Informatics Program, Boston Children's Hospital*
10:00 am–10:30 am	FACILITATION PROCESSES and GROUP ASSIGNMENTS	**Stephanie Rizk**, *Workshop Lead, RTI*
10:30 am–12:00 pm	BREAKOUTS (breaks as needed) Group A: *Adaptability of the Health IT Infrastructure* Group B: *Producing Data for Research* Group C: *Health IT Functionality Needed for Research*	*All* Groups facilitated by **Jonathan Wald, Stephanie Rizk, Alison Banger** *Focus on **Future Vision** and **Current Challenges***
12:00 pm–12:30 pm	LUNCH BREAK	*All*
12:30 pm–2:00 pm	BREAKOUTS, cont.	*All* Groups facilitated by **Jonathan Wald, Stephanie Rizk, Alison Banger** *Focus on **Overcoming Gaps** and **Activities/Responsibilities***
2:00 pm–2:15 pm	BREAK	*All*
2:15 pm–3:45 pm	REPORT OUT and DISCUSSION	*Representatives from Groups A, B, and C*
3:45 pm–4:45 pm	DAY 1 REVIEW	Facilitated by **Stephanie Rizk**
4:45 pm–5:00 pm	PREVIEW DAY 2	**Jonathan Wald**

Day 2: Wednesday, July 25, 2018

8:15 am–9:00 am	ARRIVAL	*All*
9:00 am–9:30 am	REVIEW DAY 2 TOPICS	**Jonathan Wald**, *Project Director, RTI*
9:30 am–10:45 am	BREAKOUTS	*All*
	Group D: *Patient-Centered Infrastructure Solutions*	Groups facilitated by **Jon Wald, Stephanie Rizk, Alison Banger**
	Group E: *Data Aggregation across Multiple Research Platforms*	*Focus on **Future Vision** and **Current Challenges***
	Group F: *Realizing a Transparent and Scalable Architecture*	
10:45 am–11:00 am	BREAK	*All*
11:00 am–12:00 pm	BREAKOUTS, cont.	*All*
		Groups facilitated by **Jonathan Wald, Stephanie Rizk, Alison Banger**
		*Focus on **Overcoming Gaps** and **Activities/Responsibilities***
12:00 pm–12:30 pm	LUNCH BREAK	*All*
12:30 pm–1:30 pm	REPORT OUT and DISCUSSION	*Representatives from Groups D, E, and F*
1:30 pm–2:30 pm	DAY 2 REVIEW	Facilitated by **Stephanie Rizk**
2:30 pm–3:00 pm	WRAP UP and NEXT STEPS	Facilitated by **Alison Banger** with closing by **Teresa Zayas-Cabán**