National Health IT Priorities to Advance Research: Workshop Synthesis

Drafted by:

RTI International

Stephanie Rizk, MS

Jacqueline Bagwell, MS, MMCi

Alexa Ortiz, MSN, RN

Jonathan Wald, MD, MPH, FACMI

Linda Dimitropoulos, PhD

Draft Date: 2/27/19

Table of Contents

Executive Summary

In July 2018, the Office of the National Coordinator for Health Information Technology (ONC) Chief Scientist Division hosted a 2-day in-person workshop to identify challenges related to the health IT infrastructure that ONC should address over the next 3-5 years to improve scientific discovery and application.

The workshop featured a mix of presentations and breakout group discussions based, in part, on the findings of a background report. A background report identified gaps in the health data infrastructure in the following six areas: 1) adaptability of the health IT infrastructure; 2) ability to produce data for research; 3) functionality needed for research; 4) data aggregation across multiple platforms; 5) advancing patient engagement in research; and 6) realizing a transparent and scalable architecture. Six breakout groups were organized around these initial gaps to facilitate discussions about underlying challenges.

A high-level analysis of the inputs from the workshop participants identified crosscutting challenges to achieving a future state in which research happens faster, better, easier, and improves outcomes. This document describes the key discussion points related to the challenges identified during the workshop including:

- A need for solutions that will lead to **improvements in health care data quality and availability for research**, such as:
    - increasing support for functionalities that ensure transparent and interoperable health-related data;
    - development and use of tools that allow researchers to learn how to use, interact with, and share standardized data from electronic health record (EHR) systems; and,
    - solutions to enable aggregation across multiple, non-EHR–based data sources s (e.g., socio-economic data, patient-generated data, patient-reported outcomes).
- Improvements in education and coordination **activities that increase the ability for patients, providers, and other stakeholders to participate** in the research ecosystem, such as:
    - improved coordination and sharing of functional solutions for patient matching and identity management;
    - improved education, coordination, and technical solutions to manage consent necessary for data sharing in research;
    - providing research opportunities to areas and organizations that have traditionally been underserved or underutilized in research participation; and,
    - promotion of opportunities to encourage dialogue and education on the use of health IT infrastructure for research.

1. Introduction

1.1. Workshop Objectives

The Office of the National Coordinator for Health Information Technology (ONC) Chief Scientist Division hosted an in-person workshop to identify key health IT infrastructure challenges that ONC should address over the next 3-5 years to improve scientific discovery and application. The activities planned and executed during the workshop were intended to identify specific challenges and potential solutions necessary to inform a policy and development agenda.

1.2. Workshop Details

1.2.1. Organization of the Workshop

Dates, Venue, Participants

ONC convened the "National Health IT Priorities to Advance Research" workshop on July 24th and 25th, 2018. In total, 26 experts (see ***Appendix A***) from both the public and private sectors attended either one or both days of the workshop. On the first day of the workshop, Dr. Kenneth Mandl, director of the Computational Health Informatics Program at Boston Children's Hospital, provided a keynote presentation on successful approaches to pursuing innovative research by using data in a rapidly changing health IT infrastructure and improving access to data.

Materials and Resources

Approximately 3 weeks before the workshop, RTI distributed a background report to participants which identified six gap areas between the current health IT infrastructure and the needs biomedical and health services researchers. Planning guides were developed in advance of the workshop to organize breakout and facilitator activities. A training session was held for facilitators and notetakers 3 days prior to the workshop to review planning documents and discuss any concerns or questions raised by the project staff.

1.2.2. Workshop Discussion Topics

The background report identified six gap areas relating to health IT infrastructure that must be addressed to support the advanced research needs of health services and biomedical researchers. These gap areas included:

- **Limited adaptability for research within the changing health IT infrastructure:** Because health IT is incrementally adapted to accommodate new scientific discoveries, operational and workflow needs of clinical care, and evolving business and clinical priorities, research priorities and the impact on research should also be considered.
- **Limited health IT infrastructure support to produce research data:** Health IT produces tremendous amounts of data that researchers are increasingly interested in accessing to inform their work. However, systems typically do not capture data in ways that would better support research, such as consistently identifying captured data, metadata about timing and the context and other details necessary for research-quality data.
- **Limited health IT infrastructure support for research functions:** Certain functions of the health IT infrastructure that are important for conducting research—such as locating data elements relevant to the specific research question, searching for data across multiple sources, indexing data of interest, querying for matching records, and identifying consenting status—are also relevant to other stakeholders including providers and patients. Specific functions that are important when disseminating research findings—such as implementing guideline-driven decision support triggers

and rules, application programming interfaces (APIs) with third-party functions, and other IT-driven changes in workflow—would enhance the value that research brings to the routine delivery of care.

- **Limited health IT infrastructure support for aggregation across research platforms:** Health IT that is used as a data source for a research platform should better support the platform's major functions by, for example,: receiving and processing multiple data streams, matching and linking the data, honoring data use agreements, identifying redundant data, managing updates to data and metadata, and supplying varying data formats.
- **Limited health IT infrastructure support for patient and family engagement in research:** Health IT could potentially be used to directly engage patients and their families in research. Such involvement may include data contributed by an individual through a survey (patient-reported outcomes [PROs]) or a personal/wearable device, encouraging patients to view their data and review it for accuracy, providing mechanisms to manage permission to use data, or enabling patient and family research leadership.
- **Lack of a robust health IT architecture to support research:** A consistent software architecture across different health IT components and technologies at many different organizations would advance research through better access to data, improved understanding of data context; more powerful tools that support collection, aggregation, and analysis of large and disparate data sources, and more transparent coordination across various systems.

Participants were assigned to one of six breakout groups focused on these six gap areas. Participants engaged in a facilitated discussion intended to identify specific challenges and potential solutions relative to each gap area. ***Appendix B*** contains the final participant agenda.

2. Synthesis of Cross-cutting Challenges

Most of the challenges that emerged from individual breakout sessions were relevant to more than one gap area, therefore the synthesis provided here is organized around the major challenges rather than the original gap areas. The information from the breakout sessions was synthesized using a grounded theory approach in which the information captured during the workshop by notetakers was reviewed carefully and grouped into recurring themes until patterns emerged. ***Table 1*** displays critical topics that were identified during review of each breakout group discussion.

These cross-cutting topics, which underly multiple gap areas, created the foundation of each identified challenge described in the synthesis below. Each challenge identified includes a summary of the descriptive information about key discussion points, potential solutions, and recommendations for overcoming challenges across all breakout groups.

Table 1. Critical Crosscutting Workshop Discussion Topics

	Workshop Breakout Sessions					
Cross-cutting Topic Areas	***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***
Patient identification and matching	•	•	•	•	•	•
Patient engagement, patient-centered consent	•	•	•	•	•	•
Inclusion of nonclinical data sources		•	•	•	•	•
Common data elements, common data model development	•	•	•	•		•
Development of standards for transparent exchange	•	•	•	•		
Research for underserved organizations/areas			•		•	
Creation of national dialogue/education	•	•				

2.1. Improvements in health care data quality and availability for research

2.1.1. Challenge 1: The research community requires transparent and interoperable health-related data

Researchers often require access to specific details regarding clinical treatment that may not be included within the data element as currently captured in an EHR system. These additional details can be captured in associated data fields but require additional inputs by the clinician at the point of entry and create a more complexities with the transfer and analysis of the data. Ideally, metadata fields would be used to collect additional details about a data element and tag them within the data element itself. Metadata examples include the taxonomies and ontologies in use during data collection, the purpose for capturing the data, capture methods (such as time, devices used, or mode of administration), provenance, and permissions. Standardizing certain metadata fields as part of the overall effort to promote the use of a single standard for a specific data element would greatly benefit the research field. In addition, the ability to identify what information is available and where it resides within an EHR system is imperative to improving the ability for researchers to perform better and more efficient analyses using EHR data.

Key Discussion Points

Stakeholders across the research and health IT infrastructure ecosystem should be incentivized or required to prioritize FAIR Principles—ensuring data are findable, accessible, interoperable, and reusable—to feasibly create an interoperable, universal system for collecting and utilizing patient health data. A critical technical component for achieving transparency is ensuring that clinicians and researchers can obtain important metadata such as provenance and how certain variables are calculated from source systems. Requirements related to open APIs allow for data linkage, but do not provide information about the functionality needed to understand how the data was captured or calculated; therefore, EHR developers need to allow their customers to share information about the vendor- and site-specific algorithms used to compile data across the instance of their product.

Key Recommendations

- ONC should support and/or engage in the following types of activities:
    - Create standards and offer guidance and clarity on how metadata collected both within and outside of the care setting can be shared or utilized to support researchers.
    - Build on the 2015 Edition of the Health IT Certification Criteria related to APIs to require developers to make public their underlying architectures and schemas to provide a platform to develop tools that can query data across multiple systems.

2.1.2. Challenge 2: There is a need for tools that allow researchers to learn how to use, interact with, and share standardized data from EHR systems

The use of shared common data models (CDMs) facilitates the ability to standardize and share data across institutions and research teams. There are currently several initiatives aimed at developing and harmonizing different CDMs. The ONC-led Common Data Model Harmonization project is currently working to harmonize data models from Sentinel, PCORnet, OHDSI, and i2b2 to further advance the utility and interoperability of the data within these networks for use in patient-centered outcomes research. While standardizing across various data models is a beneficial activity for many researchers, many types of research questions cannot be answered within the boundaries of a single CDM and must create unique data models or collect data elements in addition to those available in an established CDM. Therefore, development of a single, harmonized CDM did not seem desirable, especially given that standards around devices and taxonomies

often do exist but are evolving rapidly, making it difficult for a single entity to rely on a CDM to capture all essential data for a project. Standards developed on devices built at different times produce different data. Therefore, there must be a frequency with which the standards need to be applied with regard to emerging interoperability standards for medical devices.

Key Discussion Points

At the workshop, several conversations emerged regarding both the usefulness and the difficulties inherent to relying on one CDM for research. Although there is consensus in the industry that developing a unique data model for every research question is inefficient, opinions differ on what a CDM should include or how efficient it would be at streamlining the labor-intensive process of collecting research data.

Key Recommendations

- ONC should coordinate with leadership at the National Library of Medicine (NLM), the Agency for Healthcare Research and Quality (AHRQ), and FDA to investigate the feasibility of a "research workbench" tool in which researchers could
    - Specify the data elements they are interested in collecting, including metadata standards describing the elements.
    - Register their data elements, with corresponding metadata.
    - Provide specifics on collection methods.

2.1.3. Challenge 3: Rapid support for solutions to enable aggregation across multiple, non-EHR–based data sources is critical

A strong and growing body of evidence indicates that a wide variety of environmental, social, and personal factors play a critical role in health outcomes. Data related to these factors are found in a variety of sources that include, but are not limited to, social determinants of health (SDOH) data and other non-electronic health record (EHR)–based data sources including PROs and patient-generated health data (PGHD). These data may be available through services like Apple's HealthKit or other third-party developers that are entering the market. However, they remain separate from the clinical record (which is typically stored in an EHR system) and have not yet gained the same level of standardization across devices that many EHR data elements have. Digital access to these data provides an opportunity for patients, providers, and researchers to have actionable insight into the factors that impact health outcomes. To facilitate research, health data must be accessible and sharable through a health IT infrastructure that supports interoperability among data types and organizations.

Key Discussion Points

The current health IT infrastructure does not include requirements for metadata standards needed to facilitate population health research at a large scale. The ability to standardize the capture of research-relevant metadata such as time, devices used, or mode of administration would support the potential of integrating data sources like PGHD. Work needs to be done to identify additional metadata standards that could be critical to population health research – for example, location and geography could potentially support the inclusion of social and environmental data points that are increasingly identified as key factors in identifying risk and improving outcomes. Not only should the collection of this metadata be coordinated with the workflow needs of those providing routine clinical care; specific focus needs to be given to improving the methods of feeding research findings back into the healthcare system. Improving metadata standards and patient identification methods should not only improve researchers' access to data and tools to perform needed analyses but should greatly increase the ability to communicate back to patients and providers.

Key Recommendations

- ONC should support and/or engage in the following types of activities:
    - Continue to work with the Food and Drug Administration (FDA) and the Centers for Medicare & Medicaid Services (CMS) while evaluating industry standards, to enforce the use of these standards for data sources other than the EHR to interact more seamlessly with the current infrastructure.
    - Fund more demonstration projects that focus on identifying small sets of data elements about individuals that can be used to support and show the effectiveness of analyzing data from multiple sources.
    - Fund the use and development of Fast Healthcare Interoperability Resources (FHIR) resources and third-party applications.
    - Work with the National Institutes of Health (NIH) to identify opportunities for co-funding or supporting improvements in the health IT infrastructure itself.

2.2. Improvements in education and coordination activities that increase the ability for patients, providers, and other stakeholders to participate in the research ecosystem

2.2.1. Challenge 4: Lack of coordination and sharing of functional solutions for patient matching and identity management

The healthcare and research industries realize that there is a critical need to find solutions that will allow for better, more reliable matching of patients' data, which are stored across many different systems and devices. When data cannot be reliably or effectively associated with an individual, it reduces the ability to coordinate care, and to link data to conduct accurate biomedical and health services research.

Key Discussion Points

Although HHS is not authorized to issue a unique, single patient identifier (ID) for health care use, some in the private sector have attempted to address this problem (e.g., College of Healthcare Information Management Executives [CHIME] National Patient ID Challenge) with mixed results, although many continue to work on the problem by improving matching algorithms. One potential solution discussed during the workshop was to pursue the development of an individualized patient profile for research, which would include the assignment of a unique tracking number that could be used across systems to track a single patient's data for research purposes.

Key Recommendations

- ONC should support and/or engage in the following types of activities:
    - Fund challenges and other demonstration mechanisms that investigate the potential utility of creating a universal patient ID as organizations in private industry continue discussions on this and other patient matching solutions.
    - Strengthen support to increase the likelihood of developing a common identifier for research as a separate (but related) activity to develop patient matching solutions.

2.2.2. Challenge 5: Lack of education, coordination, and technical solutions to manage consent necessary for data sharing in research

Consent for research is typically collected on a project-specific level, requiring signing of a form (typically paper) at every individual location. These forms rarely include consent for data sharing, which reduces the flow of data within the research community. A more transparent and dynamic consent management system

would not only improve the ability for researchers to share data to increase reliability, validity, and to study additional research questions, but it would also likely support patient engagement by providing them with technical solutions to controlling the flow of their data. This would benefit researchers and patients alike, as well as clinical staff who are often called on to mediate the consent process.

Key Discussion Points

A centrally managed and standardized e-consent product would allow patients to control and update the data they consent to share with the research community and would remove many of the barriers often imposed by researchers and/or their organizations around sharing data for research. Researchers also require additional education from a central authority around the processes needed to appropriately de-identify data to improve their willingness to share the data.

Key Recommendations

- ONC should support and/or engage in the following types of activities:
    - Use patient-facing e-consent options that are designed to work within the current (and future) health IT infrastructure.
    - Offer educational activities related to consent for both patients and researchers.
    - Develop activities that seek to improve the general population's perception of data collection, storage, and use for research
    - Develop activities that are aimed at increasing participants' trust in research activities and provide assurances that personal data will not be used for financial gain, for example.
    - As a neutral convener, engage stakeholders to identify solutions to managing patient permissions for use of their data for research at an infrastructure level instead of limiting data to individual organizations.

2.2.3. Challenge 6: A need to provide research opportunities to areas and organizations that have traditionally been underserved or underutilized in research participation

A vast number of patients are not able to participate in research because the health care organizations that serve them do not have the resources to participate or lead research projects. Research advances are often conducted by academic medical centers and other large integrated delivery systems that provide strong, supportive expert staff and operational services; however, this is not where most patients receive their routine day-to-day care.

Key Discussion Points

The full spectrum of the research process will be strengthened if the health IT infrastructure is used to extend meaningful participation in research to patient populations that are largely cared for outside of the boundaries of a large academic medical center. Large health systems could help to develop best practices for sharing their research expertise with smaller, less resourced institutions. Health IT infrastructure should make research-focused resources accessible to a variety of institutions.

Key Recommendations

- ONC should support the following types of activities to encourage its stated objective to broaden the reach of research:
    - Increase focus on testing and translating improvements in the research-related informatics infrastructure to underserved areas and populations by

- Extending demonstrations and challenges specific to these largely untouched groups.
- Supporting research funded by other agencies (e.g., AHRQ, NIH).

2.2.4. Challenge 7: A need to better promote opportunities to encourage dialogue and education on the use of the health IT infrastructure for research

While communications and outreach campaigns to notify the public of research opportunities are not a function housed within the health IT infrastructure, it is an essential component to ensure broad awareness of opportunities and drive engagement in research participation. To ensure the broad use of the functionalities developed within the health IT infrastructure to support research, participants suggested that a process to identify knowledge and resource gaps across institutions would be important. Optimal, widespread use of these functions will require coordination among larger academic medical centers and integrated delivery networks to disseminate best practices for sharing research expertise and resources, as well as coordination with vendors to develop the infrastructure. Engaging stakeholders from across the spectrum of the research process—from patients to providers to hospital staff to researchers themselves—will support an assessment of current resources and identify how they should be expanded.

Key Discussion Points

Participants noted that in many cases, such as with de-identification protocols and use of standards and CDMs, various tools and resources exist; however, many researchers are not aware of them. A broader dissemination effort must be made, not only with existing tools, but also with newly created tools, to try to reach both researchers and patients, and those that are interested but do not have the resources to participate fully.

Key Recommendations

- ONC should support the following types of activities:
  - Create a national dialogue on healthcare data including how and why the industry will use the data (e.g., appropriate use/release of data, positive patient outcomes).
  - With the Department of Education, promote research as a civic engagement and offer information about risks and benefits.
  - Provide support for activities that improve education for scientists around consent and the importance of collecting health data.

2.3. Specific Recommendations by Breakout Group

The section above provided a synthesis of issues and recommendations that emerged across the six original gap areas. To provide easy reference to the specific recommendations submitted by each individual breakout group which made up the synthesis of cross-cutting challenges in section 2.1 and 2.2, **Table 2** provides a summary of the individual suggestions provided by each breakout group during the workshop.

Table 2. Summary of Individual Breakout Topic Group Suggestions and Recommendations

Breakout Group	Group Suggestions and Recommendations
Adaptable Health IT Infrastructure	• ***Create API for all data elements*** of an EHR; not only for the continuity of care document (CCD). • ***Renew efforts on data standards***, investigate gaps, and support a continuous lifecycle for standards (i.e., build, test, update, innovate).

	• ***Renew focus on patient identification and matching***. Designate an entity (commercial or public) to establish common patient IDs/profiles for research. • Support research to understand what patients want and how to provide it to them. Patients need solutions embedded and/or supported by the health IT infrastructure to ***control sharing of their data for research***, including access to their own data and dynamic consent options. • ***Support more work and advancement in clinical decision support (CDS)***, including solutions that are scalable to the infrastructure and the national system for public health emergencies.
Producing Data for Research	• Provide support to ***develop and disseminate sharable tools*** to empower data access and ensure that data are consumable. • Improve the business case and ***decrease provider burden*** by creating cost-efficient tools for capturing data for research at the point of care. • ***Provide better education around de-identification*** and the importance of efficient data accessibility for research. • ***Create a national dialogue*** on how healthcare data can be used for the common good (along with appropriate safeguards). • ***Establish a working relationship with EHR developers*** to achieve consensus about how to increase data accessibility.
Functionality Needed for Research	• ***Define and establish connections with data*** that influences health (outside of the traditional health system). • Require EHR systems to ***share publication of the back-end data models***/schemas so that researchers can understand how data points are calculated, especially given the customization of EHR systems per institution. • ONC and FDA should support standards development organizations and other appropriate entities to ***converge on metadata standards for medical devices***, which are currently determined solely by vendors.
Aggregating Data for Research	• ***Create a common patient profile for research***, including a unique ID. This would allow patients a way to access and control their data and make it easier to access patient data for research. • ***Develop a research "workbench" tool*** to support the issue of continually emerging standards and problems with developing a single CDM. Researchers could then document data elements, methods, and links to resources like FHIR; suggest improvements; and agree to standardized data elements and reuse of elements. • ***Develop metadata standards*** for sensor-based products, wearables, and medical devices for basic data points like time and place. • ***Avoid the replication of the research ideas from one setting to another***. Locate opportunities to co-fund research across settings that will further improve processes and quality, reduce burden, and more. • Work with NIH to ***inform study sections about the functionalities and limitations*** in the current health IT infrastructure to inform better research funding decisions.
Patient Engagement	• ***Provide better tools to patients to manage their data***, as consent for research and inclusivity are paramount. The current standard of signing a paper form at every location is a disservice to our current needs and environment. • ***Provide patients with an increased level of trust*** by using effective security measures as data starts flowing more freely.

	• Create better tools to ***allow patients to share information they generate with their physicians***, through CDS or another similar mechanism and integrate information into the EHR. Patients want to be collaborators; this will create value for patients to participate in research. • ***Support small practices and hospitals that would like to participate in research but are under-resourced*** because of research costs; however, these organizations have the highest need.
Transparent and Scalable Architecture	• ***Develop and share tools that help researchers centralize data*** that has been stored in different systems in different ways (vertically vs. horizontally) to break down barriers that currently exist within the architecture. • ***Recognize that research is not a priority*** for many organizations and that they will only participate when the barriers to entry are reduced (level of effort to participate becomes tenable) and they see return on investment by receiving usable results of research findings at the point of care. • ***Acknowledge that CDMs are important***, but they do not provide complete data for researchers. • ***Educate developers on the positive impact*** of a more useful, open-source EHR infrastructure for researchers rather than maintaining a strict interpretation on requirements related to confidentiality which are embedded into system functionalities. • ***Expand our approach to incorporating new technologies*** into the architecture (i.e., Blockchain).

3. Detailed Summary of Breakout Sessions

This section provides a short summary of the detailed notes captured during each facilitated breakout session. Each summary includes areas of strength; areas of weakness/limitation; areas of opportunity; barriers and gaps, top priorities and action items for ONC, and areas for further exploration.

3.1 Summary 1: Adapting Health IT Infrastructure

3.1.1. Areas of Strength

- FHIR is a big advancement; however, there is still a need to increase support for the development of similar solutions.
- There is excitement around machine learning and the understanding of the value of research is growing.
- Institution-based operations have a solid base of documentation built around health IT, which is important for long-term sustainability.

3.1.2. Areas of Weakness/Limitation

- EHR systems are built for clinical purposes rather than for research. There are currently two separate infrastructures that both lack adaptability.
- Vendor contracts heavily influence the capabilities of healthcare organizations to exchange health information, and these contracts are interpreted as explicitly excluding considerations that might better support research.
- Open flow of research data and methods needs to be improved and supported on a national level. Universal plug-and-play applications to support data flows are needed more than data collection and storage.
- Extreme care must be taken to ensure that data are properly controlled, and that better interoperability does not lead to reclassified data. We need to learn how to better use the data we already have before expanding our efforts.
- Researchers rely on data that comes from information standardized across EHRs through certification, which does not provide a full picture of a patient's health-related activities; therefore, research should be able to include activities that happen outside of healthcare institutions as well.

3.1.3. Areas of Opportunity

- Increase dialogue with patients about how their data are being used. Give patients (and clinicians) something in return for sharing data.
- Develop an encrypted persistent identifier for individuals who participate in research and/or a nationwide strategy for patient matching across systems.
- Achieve widespread agreement on the acceptable risks related to data sharing.
- Develop the ability to launch application processes at the appropriate point within the physician workflow.
- Improve health IT infrastructure to speed up research. If data could be accessed earlier and faster for research, it may lead to improvements in the vendor/EHR space.
- Support a shift in the market power dynamics in favor of providers and systems to better negotiate vendors to support desired system functionality.
- Establish functionality within the health IT infrastructure to help researchers gain a full and complete picture of patient care.

- Increase exposure to CDS among healthcare workers.
- Fully capture patient outcomes (e.g., long-term measures of well-being).
- Alter the focus of clinical information systems from interfacing between systems to a focus on better design on the back-end which, if built with interoperability in mind, will provide better access and exchange of data.
- Identify areas that highlight early/easy wins for machine learning rather than looking to artificial intelligence (AI) to solve the most complex problems first. Apply machine learning to data points such as laboratory tests that are already in a coded format to demonstrate a straightforward use case.
- Identify FHIR standards and resources in ways that will encourage the productive use of live data.
- Cross map records that are optimized both for care and for research.

3.1.4. Barriers and Gaps

- There is push back from vendors to expand API requirements to all elements of the EHR.
- There is no standardized directory or clearinghouse for API specifications that allows users to track value/utilization of the data received through the API.

Use Case: Machine Learning 1

Machine learning can (1) be applied to simple problems first to demonstrate the concept, using complex but coded data (e.g., labs); (2) be applied to more complex coding systems (ontologies); and (3) then help integrate terms, models, and standards across settings.

- A widespread onsensus has not been reached on appropriate standards and there is not enough instruction provided on how various coding systems/ontologies/taxonomies relate and how they interact with data models. ***This may be a good use case for AI***. There are obvious challenges to approaching standards from the top down. A crowdsourcing approach might be much more effective to discourage building new standards that already exist and a way to evaluate and improve current standards. The health IT infrastructure should be able to support a better understanding of the standards environment through machine-readable standards, for example, which should be able to help integrate terms, models, standards, and metastructures, and help prioritize standards across multiple settings.
- There is a lack of common understanding of the evolution of standards—the lifecycle that requires testing and implementation before widespread adoption.
- Patient matching is not developed sufficiently enough that there is enough data available to identify improvements or measure them. Having no universal identifier makes it difficult to match patients involved in a clinical trial across systems, which introduces data integrity problems for researchers and potential safety, privacy, and fraud issues for patients. The lack of available tools and lack of time to develop them is also a big challenge. Patients should be educated on how they may the benefit from solutions for matching and identification issues.
- There is also a demand for mechanisms that allow patients to make their own choices and control decisions about how their data are shared for research. A more transparent and dynamic consent system may reduce barriers to engaging patients.

"The healthcare system disempowers patients—how can we get patients to take control?"

"By not offering choices to patients, we lose them."

- Research data are not easily returned to patient records to inform their care. The technical infrastructure should be capable of supporting providers and patients to make better decisions based on research outcomes.
- The system for triggering a CDS system based on research requires a redesigned health IT infrastructure—one that is faster and more open (including open EHRs). There is no consensus on which agency or organization should carry this responsibility (e.g., private sector, public-private partnership, government).

Use Case: National Public Health CDS

Working across agencies, ONC and others could work to develop a national public health clinical decision support system to demonstrate the importance using the health IT infrastructure to notify providers about an identified public health threat.

- The concept of a ***national system for public health CDS*** is perhaps a way for the public sector to get involved and provide a model infrastructure and a case study that would resonate as important and effective across stakeholder groups.
- The incentives for various groups to share data vary; researchers, payers, and system developers all have different objectives.

3.1.5. Top Priorities and Proposed Action Items for ONC

- Investigate the benefits of ***requiring APIs to be applied to all EHR elements***, not just CCDs, which would require information on useful research hooks and necessary standards.
- ***Separate the adoption of standards*** from billing codes.
- Identify ***existing standards and identify gaps***. Require stronger incentives for use.
- Provide leadership and direction to ***develop and enforce universal patient matching and identification*** methods.
- Provide leadership to ***develop a tiered consent system with consideration of control by patients*** to share data for research. Support advocacy for patient engagement and control of data.
- Investigate proven methods to ***speed up research analysis*** and feed ***information back into the system*** at the provider level to improve patient care.
- Support development of ***tools to analyze free text/clinical notes***.
- Partner with the Department of Education to formally ***educate youth on the value and importance of research and data***.
- Work with the Office for Civil Rights (OCR) to ***consider developing a data protection bureau.***

3.1.6. Areas for Further Exploration

- Development of a national strategy for enforcing centralized policy and reporting of death certificate information, including access and use of that information.

3.2. Summary 2: Producing Data for Research

3.2.1. Areas of Strength

- The National Center for Health Statistics (NCHS) administers provider surveys about visit-level information to help make decisions about health services use in hospitals and long-term settings. Organizations are increasingly utilizing EHR systems to support survey data collection (e.g., National Health and Nutrition Examination Study [NHANES]).
- The Centers for Disease Control and Prevention (CDC) Office of Public Health Scientific Services (OPHSS) focuses on generating knowledge quickly rather than increasing and improving research and data quality. OPHSS monitors 120 diseases across 130 systems to build triggers, but do not consider care flow as part of their work.
- The Patient Centered Outcomes Research Institute (PCORI) has been able to develop partial solutions to facilitate clinical research. They have reduced privacy concerns by distributed data networks. PCORI's PCORNet project and Sentinel use a similar CDM (OMOP). The PCORNet network provides an intense level of data quality assessment and has 130 health systems organized into 13 research areas. Some data types such as lab data are straightforward; others, such as prescriptions, are more difficult.

3.2.2. Areas of Weakness/Limitation

- Challenges in survey data collection, such as for NHANES, include a lack of willingness for providers to participate, and a lack of available infrastructure. EHR vendors must work to implement interfaces. Some necessary data are collected from CCDs and other interfaces and implementation guides (IGs) specify the data elements and data requirements, leveraging HL7 standards. However, important data are still missing. Up to 60%–70% of the records do not include diagnosis codes and processing this data will require a huge effort to prevent data quality issues.
- Consent for research should be collected at the provider level. This is difficult to incorporate into the computerized physician order entry, even in large systems. One large system currently uses six different EHRs and is trying to merge data into one EHR; however, that alone will not solve the problem of standardizing consent processes.

3.2.3. Areas of Opportunity

- NHANES data are available for public use but require a fair amount of editing/data cleaning. The algorithm developed for each IG is transparent. IGs for specific surveys and long-term care and acute settings are distinct and are propagated in the settings with system updates and upgrades; therefore, provider input is important but EHR vendor participation is essential. Vendors are the "gate keepers" for accessing this information.
- PCORI's PCORNet has provided funding to hospitals with little research capacity to build portals to access aggregated NHANES data.
- OPHSS is more interested in non-health data to inform their future work (e.g., geography (census), socioeconomic, social determinants, pharmacy data).
- Researchers assume that patient data are collected over time. In PCORNet, data are distributed across multiple providers. Linkages to claims data are important to provide longitudinal information over time, including more complete dispensing information. The capacity to link claims and EHR is paramount. Randomized controlled trials have the advantage of participant communication—a distributed research network does not. Claims

data currently live with the payers; therefore, health plans would need to be brought into the network for these linkages to work.

3.2.4. Barriers and Gaps

- NCHS data are siloed and difficult to share and use to create national databases. Not enough is known about how IGs are used. If they are not used correctly, there is no current method of knowing, nor is there a way to evaluate the data quality.
- OPHSS challenges are due to the complex relationship between providers and vendors. The organization processes a billion records and focuses on triangulation—not curation of the data. Data may need to be curated at the point of care.
- Maintaining confidentiality is also a constant struggle. Data linkages are always temporary due to the risk of identification, and new ways to protect and share data should be developed.
- We must incentivize data sharing from participants—and change the culture of resistance—to reduce the burden of collecting data.
- PCORI has difficulty dealing with nonstandard data elements in the EHR with the building of PCORNet. Individual institutions have rich data, and the process of standardizing and accessing is expensive and slow.
- Because of the increasing discomfort with participant data linkages, the concept of consent should be examined in terms of who is able to access the data, when, and how; and who owns the data. Genetic data will become an identifier.
- Most of PCORNet's work is project-based and duplicative. For instance, local organizations can customize Epic scripts for their needs of the EHR system, but data in these fields are not easily shared.
- In order to move forward, an evaluation needs to be done to determine how current data sharing challenges lead to information blocking addressed in the Cures Act. The health IT infrastructure should make it easier for independent third parties to help. Institutions that want to share information do not necessarily know how to do so.

> *"ONC should make sure that the source data are available for others to put their hooks into."*
>
> ONC should explore IT solutions to access data from the EHR through a third party rather than depending on the vendor for data extraction.

- EHR data tables and taxonomies are considered a trade secret and must be more universally accessible. In PCORNet, the data model is constantly expanding. A data model must be created for each CDC-identified outbreak; researchers need better methods for submitting new data to a distributed data center.
- Recruitment under the NIH *All of Us* Research Program has been difficult. People are more apt to participate in local research due to the culture of mistrust and lack of education surrounding data and consent.
- Incredibly rich private and proprietary databases exist across the research field; however, the rate of biobank research consent is decreasing, caused by the recent cultural shift around sharing data. People seem more willing to share data in social media than in medical settings or with the government. At the same time, they often consent to data sharing that are incented.

- The Privacy Rule applies to government data. The Department of Veterans Affairs (VA) does not believe in de-identification of data (Million Veteran program); however, people are wary of sharing identifiable data. This leads to the question: Is there a need for two systems, one identifiable and one de-identified?

3.2.5. Top Priorities and Proposed Action Items for ONC

- ***Require each EHR vendor to have a webpage describing the schema and design of EHR tools*** 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.
- ***Increased clarity around the process and rules for appropriate de-identification*** should be developed in addition to guidance on data ownership terms that are more favorable to the appropriate sharing of information to increase efficient access to data. Sharing data should require fewer steps and systems should establish a single location to set up permissions. ***Create a national dialogue with patients*** on the benefits of data sharing, create more patient-/consumer-centered systems, and increase specificity around rules (as in the Health Insurance Portability and Accountability Act [HIPAA]).
- ***Create a clear platform for data sharing.*** At the institution level, there are transparency risks, privacy and consent issues, and budget concerns. Data quality is poor and data extraction and cleaning is expensive; the burden of ensuring better quality data should not be on the provider but should be covered through improvements to the system. ***Develop basic tools and services*** to help with these frequent, widespread problems. ***Use a machine learning use case*** to examine well-structured data. ONC should ensure that the source data are available for others to develop tests on extracting and cleaning data.
- ***Create inexpensive and efficient tools for mining and cleaning data.*** Some tools currently exist but need improvements to show the real value of increasing the integrity of data. The tools would ideally exist within the EHR to enable use at the point of care.

3.2.6. Areas for Further Exploration

- Provide clinicians better information about the importance of quality data input and how data entry impacts data quality and how to re-think the role that data plays as part of their clinical care and not view it solely as an output of care. Make research data available to clinicians for decision making.
- Create a business model for sharing data, which may help overcome issues of confidence and trust that lead to data blocking and the fear of misuse of APIs.
- Increase speed and cycle time of testing to potentially offer near-real-time access to data. Much research is time-sensitive and currently takes far too long.

3.3. Summary 3: Defining the Health IT Functionality That Supports Research

3.3.1. Areas of Strength

- Increased adoption of EHRs allows patient data to be captured across different organizations and patient records to be extracted electronically from their systems.
- Other forms of technology (e.g., wearables) have increased the electronic capture of patient data outside traditional health IT data. This may enable insight into SDOH (beyond what is collected in the EHR).
- ONC's role in determining certification criteria for EHRs is an important factor in moving toward an infrastructure that ensures clinical data captured by healthcare organizations is adequately structured for research purposes.

3.3.2. Areas of Weakness/Limitations

- Current EHRs provide a "snapshot" of a patient's health within the boundaries of the healthcare system, and rarely capture health-related activity that occurs outside of the system.
- The current healthcare system is blocked by regulation from creating a single patient ID, which makes aggregating information across data sources complex and time consuming and captures data that is likely inaccurate.
- The type of data captured routinely by clinicians often differs from that valued by researchers. The separation of the infrastructures between healthcare systems and research facilities impedes our ability to provide better continuity of care and improved care to patients.
- Data are recorded and extracted with high variability, which decreases their value for researchers and prevents efficient analysis. The benchmark/gold standard for data collection differs based on the different research questions being asked and the unique contextual and environmental variables, which leads to unstandardized and unstructured data.
- Sync for Science (S4S) provides access to longitudinal records via APIs; however, there is little to no incentive for organizations with EHRs to start using the protocols outlined in S4S.
- Patient involvement in clinical trials is infrequent due to limited value proposition. Patients do not want to participate in research if it means ending their established relationship with their physician. In addition, there are barriers (e.g., monetary costs, travel time) associated with clinical trial participation that limit involvement from large subgroups of the population; therefore, the infrastructure should be structured to reach a larger, more diverse group of providers/organizations.
- The methodology used for many clinical trials is not as straightforward and stable as we might expect; the data collection process is often complicated and may result in missing data points. Many "gold standard" clinical trial practices restrict who can participate, excluding some of the sickest patients, and therefore also the external validity of the results. The fragmentation in the current infrastructure prevents us from being able to assess the overall quality of the data and the process we are using for research.
- A high level of burden is put on the physician as an unintended consequence of how the current health IT infrastructure has been developed. Physicians are already required to capture data for billing purposes that are not clinically relevant; asking them to collect data for research purposes will likely create additional burden.

- Current policies and regulations are still often interpreted in ways that prevent access and sharing of data. Healthcare organizations hide behind HIPAA as a reason not to share their patient's data and block the transfer of data across institutions.
- Our understanding of what makes people healthier is rapidly changing. The infrastructure does not currently support methods to access or aggregate the data necessary to easily discover or validate new information.
- Research is not a high priority among healthcare institutions, providers, or EHR developers, which creates a gap in the health IT infrastructure that prevents researchers from accessing usable data.

3.3.3. Areas of Opportunity

- Create standardized, infrastructure-supported methods for researchers to recruit, retain, and return value to clinical trial participants.
- Increase access to clinical trials participation by including functionalities within the health IT infrastructure that make it easy for people to contribute their data.
- Include underserved populations in research in addition to the socioeconomically advantaged participants and populations who have access to large academic medical centers.
- Develop a repository of research-ready data sets to encourage the reuse of data for various types of analyses.
- Support innovation through cohesive, open APIs and scalable software design to merge relevant data sets in new ways.
- Moving forward, create more explicit requirements in policies and regulation to ensure that healthcare organizations do not block appropriate access to patient data.
- Support the development of services and tools that use automated methods of cleaning and reconciling data.

> *"You don't need to regulate an API or say how an API should be implemented. What we need is commonality across what's available with the APIs. [We] want the ONC to specify certain aspects about APIs. We need open APIs and minimal interoperability standards, so vendors don't capture and trap you."*

- Encourage innovative solutions and standards to easily capture data at the point of care within the provider's established workflow.
- Collect data for clinical research during routine care for easy translation from practice to research, and vice versa. Develop an infrastructure to better support the components of a learning health system, including capabilities to incorporate research findings, disseminate information, and guide providers' decisions.
- Create a more organic process for data sharing and harmonization. Improve methods to obtain, bridge, and map data around SDOH and population health. Include data available both within and outside of the EHR and determine how best use it.

3.3.4. Barriers and Gaps

- Current incentives to improve the health IT infrastructure are not aligned with innovation. Establishing incentives for health IT developers to create higher levels of interoperability and innovation is less of a priority than following the many business operations and quality reporting requirements that often need tailored solutions to meet the needs of the end users.
- Current EHR functionality does not support analysis across varying data sources.
- Attempts to comply with current regulations end up diluting EHR platforms.
- Few research registries exist that contain enough people needed for most analyses.
- There is a lack of appropriately tagged data provenance information, limiting health data from being routinely shared and reused for research.
- Once data are developed and curated for research purposes, they become locked into specific user interfaces. In part, this is due to a lack of standards and infrastructure for data sharing, which prevents the translation of research findings back into the point of care.

> *"We should push clinical research to the point of routine care. Once we do that, clinicians will have new incentives that are brought to the table and can support and finance extraction of variables for research."*
>
> *"Sync4Science allows access to longitudinal records, but there's not the incentive to do it on the scale we need."*

3.3.5. Top Priorities and Proposed Action Items for ONC

- Specify certain aspects for APIs to ***ensure commonality and prevent barriers*** in the implementation of interoperability standards.
- Review current ONC regulatory authority to identify potential activities ONC could support to ***make patient involvement in clinical trials more feasible*** and assess revisions to ONC-led regulations that could have a measurable impact on research capabilities.
- Expand the support for and share findings on ***bulk FHIR and population health data***.
- Support activities focused on the ***creation of and mapping to a CDM*** that includes standard data elements commonly used in research and pursue incentives to collect and reuse structured data.
- ***Contribute a vision*** that will help stakeholders understand the flow of information between clinical practice and clinical research, the type of data collected, the scope of data collected, and where the data originates.
- The industry must ***converge on standards, specifically around metadata elements*** and recommend best practices for metadata use cases. ONC should work with CMS (with support from the FDA) to require medical devices to record and share metadata to receive reimbursement.
- ONC could develop use cases to ***facilitate data bridging and mapping of data*** to and from the EHR. Healthcare organizations are stuck with the responsibility and cost of trying to incorporate functionalities like incorporating PROs into the point of care without a clear return on that investment.

3.3.6. Areas for Further Exploration

- Creating a feedback loop to integrate research findings into the EHR at the point of care. To examine this further, ONC could investigate smaller, successful, scalable examples.

- Determining how research data impact outcomes and create knowledge.
- Taking a more modular approach to EHR development may resolve issues of information overload and provider burden by allowing providers to access and contribute data through a more application-based software environment; for this to be possible, however, application developers must be able to access information about the endpoints available in the enterprise system (EHR) to build these modules.
- Discovering how to make research a higher priority among providers and healthcare organizations. ONC could use rulemaking to drive this initiative.

3.4. Summary 4: Aggregating Data Across Multiple Platforms

3.4.1. Areas of Strength

- There are some examples of how using CDMs allows multiple health organizations to pool multiple types of data and creates new opportunities for collaboration (e.g., possible for one hospital with Epic to send records to another Epic hospital).
- Sentinel has answered some important questions but applies more to medical surveillance. PCORNet can be used to help answer questions about risks associated with certain conditions but relates more to population health monitoring than health services or biomedical research.
- The National Cancer Institute (NCI) is working to identify fields that are unique to cancer and other disease areas with no adequate billing code equivalents. NCI is considering drafting recommendations for structured data elements to track specific disease progression.
- There are hundreds of researchers who are using digital health data in their studies. A group at Mount Sinai tried to create a clinicaltrials.gov for digital health studies; this evolved and allowed researchers to register what they were studying and supported collaboration between researchers.

3.4.2. Areas of Weakness/Limitation

- The development of standard data models will not necessarily address the need for individual researchers to understand how to access the data they need in a meaningful way; researchers may perhaps be more interested in standardizing elements of a data model than being given a CDM to use.
- Researchers need a mechanism to register data elements. Researchers need tools to find similar data elements or to improve data elements.

"You're asking researchers to switch from developing their own data model to giving them a data model to use. I'm not convinced researchers are ready for this. Data models and standard elements are two different pieces."

"One of the challenges of [a CDM] and encouraging collaboration is different investigators might have different aims. There might be overlap in the data they want to collect but there are nuances that don't match up. It's a great concept, but every researcher has their own grants and own aims that are unique."

3.4.3. Areas of Opportunity

- Data generated during healthcare activities could be better used by researchers if they were made available in a common data model such as OMOP.
- Additional instructions on how to work with FQHCs that are not as familiar with research should be developed; however, they are now sending their data to the *All of Us* Research Program. All participant data collected at research sites goes through an initial round of curation and can then be dumped into the data center run by Vanderbilt/Verily.
- Device manufacturers may mutually benefit if researchers explored ways to improve data from medical devices (even hospital grade). Is it possible to start with the data elements and just pick 12 that are consistent low-hanging fruit?
- How can we empower and encourage researchers and innovators to collaborate? Could we create a common space to register and share data elements? For example, Fitbit devices can identify if a wearer is running, cycling, or swimming, because the data elements were standardized based on a group's consensus on the parameters for each activity.

Use Case: Registering Data Elements via Research Workbench

"Data elements are the way to go… If there was a way we could share what we have then we could start to see re-use at the data level."

- Having a ***research data "workbench"*** with utilities that verify the permissions for using data and calibrates data to a common standard is a possibility. Data could be input for short cycle modeling, creating a combination of test bed and workbench. This tool could be used across multiple parts of the cycle of research and could test created constructs and how well they map to a research model, and eventually map to FHIR resources using pre-existing elements.
- Establishing a ***national research profile***, perhaps with a common identifier for research, would be a huge step toward building trust and providing the elements needed to match patients across data sources. *All of Us* is one example of how a profile could be structured from the top-down. A different, more grassroots method could work in parallel, but would not be managed at the top level. This would serve as a demonstration project that could evolve into something more structured. A current example is the public's willingness to share their location data.

Use Case: National Research Profile

Although All of Us *is a top-down structure for collecting data, creating a national research profile system would work as a parallel, grassroots track. If it started as a demonstration project, it could build evidence of the value of patient matching in a system that is crowdsourced around trust.*

- The EHR is a data source but not necessarily a research platform. Although clinicians utilize it for clinical care, it could support evidence-based medicine, and the research should loop back to the point of care. The EHR should be used as both a data source and repository. To demonstrate this in the short term, we should start inputting research data into the EHR.

3.4.4. Barriers and Gaps

- One of the biggest barriers to data aggregation is the accuracy of data (quality) and the quality of metadata. FDA allows a 20% range in home glucose monitoring machines. Even when the researcher is collecting their own data, many elements are recorded. Researchers can make individual adjustments on a case-by-case basis, and although they seek the same data, the data may vary. The provenance data should be available to enable another researcher to decide whether the data are comparable.
- A panel of experts/researchers could request they types of metadata that would be valuable for clinicians and researchers. ONC could then work with a standards panel to develop data parameters/requirements for devices (e.g.., must report time at the minute level or sync to the NIST atomic clock).

3.4.5. Top Priorities and Proposed Action Items for ONC

- ***Develop a wish list of the metadata*** that would be most valuable for clinicians and researchers that ONC could take to the standards panel—data items and specifications for devices. Investigate and then set parameters needed for data aggregation.
- Researchers can ***drive the industry through demonstration projects***. A project or set of projects focused on demonstrating the ***value of having standardized data models and***

common data elements could have long-term positive effects on the research and health IT infrastructures.

- Investigate the appropriate entity to ***develop a research workbench tool***, with ONC specifying the initial elements that should be populated by researchers. This concept could be developed in several ways, depending on user needs. It could start as a tool to register data elements or find similar ones. The community could then verify elements and drive toward standardization. Eventually, users could begin creating constructs that map to their data models and even work toward a defined set of CDMs. Over time, FHIR resources or any other components that researchers found helpful could be added, and it could become a test bed and source of information for emerging standard's needs (e.g., exchange, metadata, etc.).
- ONC can use challenges, cooperative agreements, or pilots to create use cases for the long-term goal of ***developing a common centralized patient profile for research***. In addition, if a unique ID for research and a centralized e-consent management platform were introduced through this profile (sharable in machine-readable format and directed by the patient), it could be a valuable resource for engaging patients by ensuring transparency and control of their data. This would also support cross-system communication with researchers and perhaps even include data for systems that allow patients to consent using their PGHD from various devices.

3.4.6. Areas for Further Exploration

- Those in the health IT field are increasingly interested in finding better ways to collect real-world data rather than data that is generated within a research study. There is a shift in focus away from simply trying to overcome the lack of current data for research, for example, tracking disease progression or outcomes.
- ONC should consider how they might collaborate to shape co-funded research initiatives. These co-funded initiatives may not be research projects themselves, by definition, but are envisioned as projects that better enable research. ONC could request a research agenda from NLM around the ethics of advanced data discovery. A majority of these co-funded projects focus on problem solving, but because there is no informatics advancement, they cannot be funded under the current parameters of most research funding streams.
- There is a new movement to index computable knowledge, which will mobilize computable biomedical knowledge.

3.5. Summary 5: Advancing Patient Engagement

3.5.1. ONC Areas of Strength

- CMS's Meaningful Use Stage 3 requirements opened the door for participants/patients to engage with their healthcare data electronically. Across institutions, EHR patient portals have served as a one-way communication mechanism (from provider organization to patient) which can also be leveraged to draw information from patients when needed, for example, to review for inclusion criteria. The question that remains is how can it be leveraged for patient engagement in research?

3.5.2. ONC Areas of Weakness/Limitation

- There is still a lack of transparency and variation in how consent for research is handled, especially for research purposes (e.g., opt-in/opt-out options). Researchers fear that if consent were more transparent, the current low research participation rate would decrease even further. We need to understand the reasons behind attrition.
- The General Data Protection Regulation will likely have an impact on patient participation, as it requires businesses to specify purpose of use and ownership of data.
- There are high attrition rates. Opt-in and opt-out options enable patients to control their own data.
- Each area of research area uses a different consent form. There is no standard language, but there is a basic framework that can be used and shared. A broad-based consent such as Partners Biobank's, should be included in all research consent forms.

3.5.3. ONC Areas of Opportunity

- Quality improvement (QI) initiatives do not require Institutional Review Board (IRB) approval for clinical data use. If potentially publishable findings were a byproduct of QI work, providers/researchers could retroactively obtain IRB approval and conduct research analyses on the data sets, thereby integrating QI and research data.
- Consider supporting a culture shift to allow patients to easily retract participation. ONC's View, Download, and Transmit requirements have defined EHR/provider-based specifications; OCR is responsible for regulating consent (broadly). ONC's challenges on Model Privacy Notice and awareness of data currently require coordinating with OCR to establish how data will be used.
- Support creation of tech platforms built around privacy and security management and e-consent (PCRO).
- Use AHRQ-funded Patient-Centered Outcomes Research Clinical Decision Support Learning Network to understand how to access and collect relevant and accurate data *from and for* patients. When the patient is engaged and involved in conversations, it impacts endpoints.
- Continue to leverage the ability to maintain robust/mature information within FHIR resources, such as identification information.
- Broaden the conversation around research participation. Clinical trial participation is not diverse and does not consider food or housing insecurity. There is low participation with community health centers.
- If patients are determined, they often find the information they seek. APIs could increase patient access to information.

3.5.4. Barriers and Gaps

- OCR is still working to clarify guidance around HIPAA and the right of access of data. Providers are still unsure about the approved methods for distributing patient data requested by the patient.
- If participants/patients were to control their data, the institution or provider would need to relinquish control. Regulations should specify which data points should be controlled by each party.
- Current research consent models are designed to protect the institution—not inform the patient and should be modified to create and reinforce trust.
- The infrastructure should be improved to support "thing identity" (i.e., having computable biomedical knowledge to identify the person, device, etc.), should have metadata and provenance information to identify the source, and offer a method to consent to use the information from the "thing." To identify patients and their device data, a linkage between the two is needed and proof (audit quality) of what was collected/managed/shared.
- There is still much to learn about the best way to deliver information to patients and the most effective way to transport knowledge and data. Multiple portals ("portalosis") have become an unintended consequence of the certification criteria. There is low value for the patient in portals, and they are not widely accessible to certain populations.
- Many under-represented populations use mobile devices and therefore may be more reachable through SMS/text instead of through a poorly designed patient portal, but text messaging is considered less secure. The challenge of inclusiveness is one of the embedded issues in digital research.
- Concerns about inappropriate release of sensitive information and medical identity theft are still prevalent, as are conversations around how to balance privacy and security. Many systems across the healthcare landscape are vulnerable to data compromise.
- Although direct control over PGHD is not within ONC's authority, there would be potential benefits to making this data available to providers. Would it be possible for APIs to "write" data to EHRs in addition to "reading"/extracting it, and give providers access to PGHD from within the EHR? Would ignoring the importance of this capability (due to lack of clarity on clinical relevance and potential liability issues) further impede healthcare providers from engaging with their patients?

3.5.5. Top Priorities and Proposed Action Items for ONC

- ***Include education on the importance of functionality and patient portals in the EHR certification program***. Provide modules on consent, specifically research consent, engagement, and research participation in the patient portal. Provide patients with more than just access to their EHR records; build on the paradigm shift and emphasize the positive impacts of engaging patients in research.
- ***Standards should be created for how CDS driven by patient-generated data (e.g., PGHD, PROs) should be entered into an EHR system***. The health IT infrastructure should include mechanisms that allow providers and researchers to meet patients where they are in their own healthcare engagement level.
- Find ways to ***support small practices and hospitals that are under-resourced*** (and their patients) to participate in research.

3.5.6. Areas for Further Exploration

- The current state of the consent document is a disservice to researcher needs, both now and looking toward the future. Transparency in consent for research is needed if the goal is to scale up data aggregation capabilities across the sites that store and curate research data. Clinical researchers successfully receive participant consent, but more research should be done regarding how well patients are being informed in that process. A mechanism to allow researchers to make collected data more widely available should be created.
- Overall, key themes for patient engagement include consent for research, standards, inclusivity, data security, and the value proposition for the patient.

3.6. Summary 6: Realizing a Transparent and Scalable Architecture

3.6.1. Areas of Strength

- None discussed or identified in the notes.

3.6.2. Areas of Weakness/Limitation

- Imposing an architecture after the components have been deployed is difficult and not efficient. Updates must then be made through the products (e.g., APIs) as they are not embedded in the infrastructure.
- It is safer for institutions to keep the data at the original source.
- Data extracted from the EHR does not include data that is not yet mapped into the CDM; it is not observational data. Researchers should be trained on how data is stored in the network. Only then can APIs be used as a tool to support transparency and streaming data access.
- It is difficult to track how and what information is being shared across the distributed architecture.
- Some architectures include horizontal partitions of patient data in which calculations are run on each stack and the information is aggregated up and down the stacks. Others include vertical partitions of patient data, which first require linkages and then algorithms to operate across the vertical partitions of data for analysis (e.g., NHLBI's Strong Heart Study).

3.6.3. Areas of Opportunity

- The VA uses VA Informatics and Computing Infrastructure (VINCI) to clean, aggregate, and provide access to data for their patients, and has begun collaborating with private providers; 65% of veterans already see Medicare providers. Exchanging data between these two systems presents a challenge, as their policies and systems are very different. Once the software is in place, however, the analysts have no need to understand the complexity of the underlying systems (vertical vs. horizontal) because they only receive a set of vertically partitioned data to analyze.
- The VA has partnered with the Department of Energy to perform secure analysis of large digital health and genomic data (i.e., big data). Through this project, the agencies will develop a centralized comprehensive data coordination center and create algorithms to pull together disparate data and make deidentified data available to researchers.
- The VA is trying to lead the effort to make all their data (de-identified) a national resource; this is an ongoing effort.
- A Memorandum of Understanding may be used to coordinate collaboration between entities (but its details can be complicated), and establish the organization responsible for maintaining the data, running algorithms, and establishing methods to request information.
- ONC has used the EHR certification program to encourage health information exchange functionality within the EHR vendor community, which could also be used to improve research-specific functionality.
- Blockchain is an important architectural principle to consider, as it supports decentralization, transparency in transactions, and reproducibility.
- Purchasers of EHR systems could have considerable leverage to demand better quality products that are scalable and customizable from the start (e.g., accommodate clinical workflow variations, research needs, etc.) but need help mobilizing the effort.

3.6.4. Barriers and Gaps

- We should not focus completely on creating centralized data hub solutions. Hybrid models may be dynamic and flexible depending on the study and can use data hubs on an "as needed" basis.
- Loss of data control and re-identification are major concerns. Organizations increasingly desire to exchange information, but also recognize the associated risk.
- The clinical environment views health services research as a separate realm with no overlap. We should shift funding to promote more analyses to more quickly show clinicians how research impacts their work. Current mechanisms can return information to the EHR but need to be more robust and automated.
- Data should be linkable across systems, with the ability to include social determinants data.

3.6.5. Top Priorities and Proposed Action Items for ONC

- ***Study, understand, and describe the architecture*** that already exists (e.g., vertical vs. horizontal) to support common goal-oriented architecture changes with a well-coordinated policy component to collectively evaluate any major decisions.
- Ensure that each ***vendor exposes enough metadata*** to enable a third party to access all data in the system. Researchers need data accompanied by relevant metadata and provenance information.
- Help develop a third-party application embedded into the clinical workflow to ***support both clinicians and researchers in achieving their goals***. More research, demonstrations, and understanding are needed around the enhancement of CDS Hooks.
- Focus on activities that ***reduce provider burden*** across the infrastructure, such as promoting quality metrics with better alignment, supporting incentives for provider participation in research, and developing plug-and-play tools that support research.

3.6.6. Areas for Further Exploration

- Define the properties of a health IT system that is supportive of research.
- Consider ways to engage the patient around consent. This may include e-consent (SageBio is a leader in the private sector) and/or making the patient part of the trust bundle (e.g., notified each time someone accesses the EHR).
- Focus opportunities toward small and midsize vendors that may be more receptive to working with ONC in informing their product development and not placing restrictive provisions on the release of their internal schemas.

Appendix A: Workshop Participant List

ONC Technical Expert Workshop
National Health IT Priorities to Advance Research

PARTICIPANT LIST

Kenneth 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
US 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
US Department of Veterans Affairs

Paul Varghese, MD
Head, Health Informatics
Verily

Appendix B: Final Workshop Agenda

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

Location: RTI International, Washington D.C.
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 and COFFEE	All
9:00 am – 9:20 am	WELCOME, INTRODUCTION and OPENING REMARKS	**Jon 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	**Ken Mandl**, *Director, Computational Health Informatics Program, Boston Children's Hospital*
10:00 am – 10:20 am	FACILITATION PROCESSES and GROUP ASSIGNMENTS + 10 minute break	**Stephanie Rizk**, *Workshop Lead, RTI*
10:30 am – 12:00 pm	BREAKOUTS (breaks as needed) - *Adaptability of the Health IT Infrastructure* (Wald) - *Producing Data for Research (Banger)* - *Health IT Functionality Needed for Research (Rizk)*	*All* Groups facilitated by **Jon Wald, Stephanie Rizk, Alison Banger** *Focus on **Future Vision** and **Current Challenges***
12:00 pm – 12:30pm	LUNCH - Boxed lunches delivered	All
12:30 pm – 2:00 pm	BREAKOUTS, cont.	*All* Groups facilitated by **Jon 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	**Jon Wald**
5:30 pm – 7:30 pm	GROUP DINNER	Old Ebbitt Grill 675 15th St NW (202)347-4800

Day 2: Wednesday July 25, 2018		
8:15 am – 9:00 am	ARRIVAL and COFFEE	All
9:00 am – 9:30 am	REVIEW DAY 2 TOPICS	**Jon Wald**, *Project Director, RTI*
9:30 am – 10:45 am	BREAKOUTS - *Data Aggregation across Multiple Research Platforms (Rizk)* - *Advancing Patient Engagement in Research (Banger)* - *Realizing a Transparent and Scalable Architecture (Wald)*	*All* Groups facilitated by **Jon Wald, Stephanie Rizk, Alison Banger** *Focus on **Future Vision** and **Current Challenges***
10:45 am – 11:00 am	BREAK	*All*
11:00 am – 12:00pm	BREAKOUTS, cont.	*All* Groups facilitated by **Jon Wald, Stephanie Rizk, Alison Banger** *Focus on **Overcoming Gaps** and **Activities/Responsibilities***
12:00 pm – 12:30 pm	LUNCH - Boxed lunches delivered	*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**