

2016 ANNUAL  
REPORT OF HHS  
PROJECTS TO BUILD  
DATA CAPACITY FOR  
PATIENT-CENTERED  
OUTCOMES  
RESEARCH

Office of Health Policy  
Assistant Secretary for Planning and Evaluation  
U.S. Department of Health and Human Services

*Office of the  
Secretary's Patient-  
Centered Outcomes  
Research Trust Fund  
(OS-PCORTF)  
Portfolio Report,  
June 2017*

# Table of Contents

---

I. Executive Summary .....	1
II. Introduction .....	1
Building a Foundational Federal Data Infrastructure for Patient-Centered Care .....	2
Active OS-PCORTF Funded Projects .....	2
III. Agency for Healthcare Research and Quality .....	3
Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries ...	3
IV. Centers for Disease Control and Prevention .....	4
Improving the Mortality Data Infrastructure for Patient-Centered Outcomes .....	4
V. Centers for Medicare and Medicaid Services .....	6
Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service .....	6
VI. Food and Drug Administration .....	8
Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research .....	8
Cross-Network Directory Service .....	9
Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data .....	10
Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data .....	11
Utilizing Data from Various Data Partners in a Distributed Manner .....	12
VII. National Institutes of Health .....	13
Creation of LOINC Equivalence Classes .....	13
Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-reported Information with Other Data Sets and Assess Its Validity .....	14
VIII. Office of the National Coordinator for Health Information Technology .....	15
Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data .....	15
Creating the Foundational Blocks for the Learning Health Care System: Data Access Framework .....	17
Creating the Foundational Blocks for the Learning Health Care System: Structured Data Capture .....	18
Security and Privacy Standards for Patient Matching, Linking and Aggregation .....	19

IX. Cross-Agency Funded Projects .....	21
Development of a Natural Language Processing Web Service for Public Health Use .....	21
Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy .....	22
X. Conclusion .....	24
Appendix A. OS-PCORTF Project Portfolio .....	26
Appendix B. Abbreviations .....	28
Appendix C. Glossary .....	29
References.....	31

## I. Executive Summary

This 2016 Annual Report provides a synopsis of active projects in the Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio. The report describes project goals, objectives, accomplishments, and interagency collaborations for developing data capacity for conducting patient-centered outcomes research (PCOR).

The Office of Health Policy (HP) in the Office of the Assistant Secretary for Planning and Evaluation (ASPE) is responsible for administering the OS-PCORTF on behalf of the Secretary of Health and Human Services (HHS). HP meets this responsibility in partnership with HHS leaders, researchers, and data experts across the nation. HHS agencies and their partners have common and converging needs for high quality research data that enable robust analyses on patient health outcomes. Since 2011, the OS-PCORTF has funded 26 projects, eight of which have concluded (Appendix A). This report reflects the work of 18 projects that were active as of December 2016.

Projects described within the report are organized by the federal agency primarily responsible for administering the project. The report underscores the breadth of work to build data capacity in HHS and the expanding capacity of data for PCOR and other studies in HHS.

## II. Introduction

HHS agencies routinely collect, link, and analyze data that can be used to generate new scientific knowledge about federal programs and the patient populations these programs serve. These data are foundational to research that expands knowledge about the outcomes and effectiveness of health care treatments and interventions. As a consumer, producer, and regulator of key national health data, HHS is uniquely positioned to coordinate its programs to build national data capacity in support of the mission, statutory authorities, and annual priorities of each HHS agency and the Department as a whole. Coordination of efforts to build data capacity across HHS strengthens its research, analyses, and public reporting programs while simultaneously reducing unnecessary duplication, inefficiencies, and reporting burdens on patients and health care providers. HHS agencies have common and converging needs for high quality, timely data that enable independent researchers and government analysts to collaborate and conduct studies quicker, more accurately, and at a lower cost.

The advantages of building the national data capacity for research are not limited to current programs in HHS. An expanded data capacity also supports independent research programs in academia, industry, and private nonprofit organizations. Expanding data capacity aligns with several new legislative provisions for patient-centered research within HHS agencies like the patient-focused drug development provisions in the 21st Century Cures Act and the value-based provisions of the Medicare Access and CHIP Reauthorization Act (MACRA). Moreover, it establishes a pathway for the future of HHS research programs to leverage the power of extremely large datasets (sometimes referred to as “big data”) and new computational tools that can be applied to understand the patterns and determinants of health care interventions on patient outcomes.

The goal of OS-PCORTF is to build the national data capacity and infrastructure needed to collect, link, and analyze data for PCOR, and to enable PCOR findings to be integrated into clinical practice. The ultimate aim of these efforts is to allow patients, providers, and caregivers to make more informed health care decisions. Achievement of this goal will require a data infrastructure that supports the collection of and access to a range of data sources by researchers who can generate and broadly disseminate clinically relevant insights to patients, providers, and other key stakeholders.

Since 2010, ASPE has administered the OS-PCORTF on behalf of the Office of Secretary and convened a council of HHS agency leaders to identify priorities and make recommendations for projects that build data capacity for patient-centered research. The council includes representatives from the Agency for Healthcare Research and Quality (AHRQ), Office of the Assistant Secretary for Preparedness and Response (ASPR), Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid Services (CMS), Food and Drug Administration (FDA), Health Resources and Service Administration (HRSA), Office of the National Coordinator for Health Information Technology (ONC), National Institutes of Health (NIH), and the HHS Chief Technology Officer (CTO). A major goal of building data capacity is to support HHS research programs that generate scientific evidence that informs decisions about patient health outcomes. The common interest in building data capacity for patient-centered research brings together the expertise of HHS agency leaders, informaticians, technologists, and researchers to identify priorities.

## Building a Foundational Federal Data Infrastructure for Patient-Centered Care

Through the OS-PCORTF, ASPE is currently funding projects to *build data capacity for patient-centered research*. Investments in PCOR data infrastructure address foundational components that will support both current and future efforts to advance patient-centered research, as well as other federal initiatives.

The OS-PCORTF's current focus is enhancing and improving data infrastructure related to five core functionalities.<sup>1</sup> These are:

- **Use of Clinical Data for Research.** There are multiple sources of clinical data available for research (e.g., electronic health records (EHRs), patient portals, registries); work in this area is focused on improving access and interoperability of clinical data for query and analysis.
- **Standardized Collection of Standardized Clinical Data.** Supports the use of common data elements to enable more effective and efficient linking and aggregation across data sources.
- **Linking of Clinical and Other Data for Research.** Allows researchers to collect longitudinal patient information and to link data sets with other relevant information for research.
- **Collection of Participant-Provided Information.** New data collection technologies, facilitate the collection of patient-generated information critical to PCOR.
- **Use of Enhanced Publicly Funded Data Systems for Research.** Focuses on efforts to leverage current investments in infrastructure to inform future infrastructure development.

Collectively, these investments and initiatives have begun to lay the foundation for a data capacity and infrastructure that can be leveraged for patient-centered research by users across the federal government and researchers in the field.

This first Annual Report provides a synopsis of active OS-PCORTF projects. The goal of this report is to inform stakeholders, including policymakers, PCOR researchers, federal entities, providers, consumers, and the health informatics community about the objectives, activities, and outcomes of currently funded OS-PCORTF projects and to highlight the progress and contributions these projects have made to develop PCOR data infrastructure.

### Active OS-PCORTF Funded Projects

Since 2011, the OS-PCORTF has funded a diverse set of projects in PCOR data infrastructure. This portfolio report reflects projects that are currently active (see Appendix A for a reference list of past OS-PCORTF-funded projects). The summaries that follow were constructed based on a review of the project statements of work, quarterly progress reports submitted by the awardees to ASPE, calls between project officer and awardees, submitted deliverables, and information publicly posted to HHS agency websites.

The projects are organized by the federal agency responsible for administering the project. This structure highlights the breadth of federal partners involved and the applicability of building PCOR data capacity to each agency's core mission, as well as to the shared goals of improving research, quality, and patient outcomes through PCOR and the learning health system.

### III. Agency for Healthcare Research and Quality

AHRQ is administering one active OS-PCORTF-funded project, Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries.

#### Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries

EHRs and registries contain a variety of rich clinical information including demographics, diagnoses, medications, allergies, and laboratory values. These data have the potential to support hypothesis generation and large scale clinical research studies. To this end, vocabulary and standardization of electronic formats and methods to electronically exchange information has been a priority for EHRs and registries. However, less focus has been placed on differences in clinical definitions across institutions or on efforts to harmonize these definitions. It is essential to address the variability in clinical definitions in order to meaningfully interpret results of studies and use the results to improve patient outcomes. One example is variability within quality measures and public reporting requirements related to care processes, use of medications, mortality, and readmission rates after myocardial infarction. While existing measures provide standard definitions for the process and outcomes, they do not address the variability in the definition of myocardial infarction used by individual health care institutions. For example, one study investigating the impact of differing clinical definitions on study findings showed that in one registry, the rate of reported myocardial infarction was 7.2% using the biomarker CKMB and 24.3% using the biomarker troponin as the criteria, representing a three-fold difference in ascertainment.<sup>ii</sup>

*AHRQ is actively working with registry owners to set up a common group of outcome measures, in a standardized vocabulary, and get the registries' outcome measures harmonized so the data can be combined and analyzed together.*

To address this need, AHRQ is developing an Outcomes Measures Framework—a conceptual model for developing standard outcome measures—to design and pilot test a tool for collecting and reporting about outcome measures in a system such as the AHRQ Registry of Patient Registries.<sup>iii</sup> This OS-PCORTF project supports the creation of outcome measures for the Outcome Measures Framework. The Outcomes Measures Framework will support standardized collection of standardized data, which is key for facilitating linkage and comparing or aggregating of data across sources.

**Project Purpose and Goals:** This project will convene a series of clinical topic-specific working groups to discuss the various definitions currently in use and how these definitions can be harmonized to promote common definitions for outcome measures across data collection and reporting systems. The goal of the project is to select a range of topics that represent different clinical conditions and a wide variety of patient populations. The working groups will solicit input from a broad stakeholder community, including registry holders, EHR developers, policymakers developing quality measures and other types of mandatory reporting measures, as well as clinicians, health systems, industry representatives, federal health and human service agencies, and patients.

The deliverable from these working groups will be a consensus set of clinical data elements that can be consistently used to represent specific outcome measures and best practices for harmonization and governance. The definitions and best practices will be made publicly available for use by registry owners and PCOR researchers.

The project objectives are to:

- Develop a consensus set of clinical data element definitions, for outcome measures for each of five clinical topic areas, to be used in registries;
- Develop best practices for governance of data element definition libraries; and
- Develop best practices for harmonization of outcome measures between different registries, EHRs, and reporting requirements.

**Accomplishments and Deliverables:** AHRQ is in the process of selecting the five clinical topic focus areas for this project. The first topic – atrial fibrillation – has been identified. Additional topics currently under consideration for inclusion in this project are asthma, depression, and low back pain. AHRQ has held two public webinars to date during which they sought input from the stakeholder community and is currently synthesizing the feedback from these webinars to help inform topic selection.

**Coordination with other Federal Agencies:** AHRQ is working with the NIH to add the AHRQ-developed clinical data elements, when complete, to the NIH's Common Data Element (CDE) Repository. The NIH CDE Repository is a platform that contains data element definitions to use in research and other purposes. The NIH CDE Repository definitions include “machine-readable” definitions (i.e. those that can be interpreted by an EHR). Adapting the AHRQ clinical data elements in machine-readable format will facilitate standard data collection at the point of care.

Period of Performance: 4/29/16 – 4/28/18

Federal Point of Contact: Elise Berliner

## IV. Centers for Disease Control and Prevention

The CDC is administering three active OS-PCORTF-funded projects: 1) Development of a Natural Language Processing (NLP) Web Service for Public Health Use, 2) Improving the Mortality Data Infrastructure for Patient-Centered Outcomes, and 3) Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy. The Development of a NLP Web Service is a joint project with the FDA and the Privacy and Security Blueprint is a project in collaboration with ONC. Projects that are cooperatively funded —the NLP and Privacy and Security Blueprint—are summarized jointly (see Cross-Agency Funded Projects).

### Improving the Mortality Data Infrastructure for Patient-Centered Outcomes

The timely reporting of death information is critical for public health surveillance and comparative effectiveness research (CER). Comprised of all U.S. mortality events since 1979, the National Death Index (NDI) database allows researchers to match entries in the NDI to those participating in longitudinal clinical and epidemiologic studies to determine both fact and cause of death. A significant challenge with

the NDI has been the lag between the date of death and the availability of the record for matching purposes. This has limited the NDI for timely patient follow-up and survival outcomes determination.

**Project Purpose and Goals:** The CDC’s National Center for Health Statistics (NCHS) is working to improve the infrastructure to support more timely and complete mortality data collection through more timely delivery of state death records (e.g., cause of death) to the NDI database and by linking NDI records with nationally collected hospital datasets to obtain a more complete picture of patient care.<sup>iv</sup>

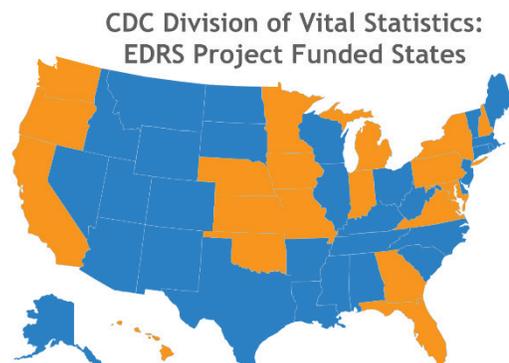
The CDC is engaged in three inter-related tasks to enhance the national mortality system:

- *Strengthen existing state mortality data collection infrastructure.* Currently, 38 states have Electronic Death Registration Systems (EDRS); however, not all are statewide and they do not cover all death events in a given region. This project uses the existing EDRS to forge a comprehensive network that support electronic data transfer to the NDI for the more timely delivery of state records to the NDI database. The intended outcome is for states to report 80% of deaths electronically through existing EDRS to the NDI, thereby enabling more timely linkage of clinical data to the NDI for research.
- *Pilot draft national standards for the exchange of relevant electronic death data from EHRs to EDRS.* Integrating EHR systems with vital records systems provides another opportunity to improve quality and timeliness of mortality data collection and distribution through standardization and reduction of duplicate data entry. The CDC will expand the Health Level Seven International (HL7) standards pilot from Utah to California, leveraging original pilot findings to expand the use case to support bi-directional exchange from the NDI to state EDRS and subsequently to hospitals in order to support quality of care assessments and CER. The pilot will also address previously identified workflow and infrastructure challenges.
- *Pilot the linkage of the National Hospital Care Survey in-patient and emergency department data with the NDI.* The pilot will assess the feasibility, validity, and reliability of measuring in-patient and post-discharge mortality. This linkage will provide additional information on patient survival after hospitalization and emergency care and eventually allow identification of patterns of care, and individual and provider characteristics associated with in-hospital and post-discharge mortality.

*Leveraging existing state EDRS capacity has the potential to decrease the lag between the data of death and the availability of the record for matching from 2-plus years to 10 days.*

**Accomplishments and Deliverables:** The CDC has made significant progress as it enters into the final year of the project.

- CDC funded 19 states (depicted in orange below), all currently in an execution phase, to address EDRS barriers to improve the timeliness of reporting. For example, some states are using the funding to update their EDRS to remove manual steps, while others are promoting physicians and medical provider community trainings to increase participation in local EDRS.



- HL7 standards have been developed and a new implementation guide drafted, the “HL7 v2.6 Vital Records Death Reporting Implementation Guide.”<sup>v</sup> NCHS and California have successfully exchanged “live” mortality records using the HL7 draft standards. The implementation guide is currently undergoing review through the HL7 consensus and approval process.
- The pilot linkage has been completed. CDC has prepared a report describing the methods and preliminary results of the data linkage project.

**Period of Performance:** 4/3/15 – 09/22/17

**Federal Point of Contact:** Delton Atkinson

## V. Centers for Medicare and Medicaid Services

CMS is administering the OS-PCORTF-funded project, Improving Beneficiary Access to their Health Information through an Enhance Blue Button Services.

### Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service

The current CMS Blue Button service was established in 2010 as a joint effort of CMS and the Veterans Administration and allows CMS beneficiaries to download their CMS information via MyMedicare.gov. Blue Button is a standard that makes the patient the custodian of his/her data. Patients can view and download their health records on their computer or mobile device, where they are stored locally. The Blue Button platform, in its many instantiations (e.g., BlueButton, Consumer Information Exchange), engages patients to actively access and share their own data. However, the current download functionality supported in the Blue Button beneficiary portal presents patient data in formats (e.g., plain text and PDF) that make it difficult to reliably extract data that is then useable by applications, researchers, and other service providers with whom beneficiaries wish to share their information.

CMS has undertaken a conceptual re-design of Blue Button to be a “data-as-a-service” platform, allowing data to be accessed on demand and presented in a “human-readable” form to users. To accomplish this re-design, CMS will utilize the HL7 Fast Healthcare Interoperability Resources (FHIR) standard.<sup>vi</sup> FHIR is a new HL7 standard for exchanging health information electronically designed to simplify implementation without sacrificing information integrity.

**Project Purpose and Goals:** The enhanced Blue Button service, BBonFHIR, will create an application programming interface (API) that enables CMS beneficiaries to connect their MyMedicare.gov data to applications and services they trust, including research platforms related to research studies in which the beneficiary may be interested in participating. Utilizing the FHIR framework to ensure data is in a structured format that can be accepted by a wide range of applications, the CMS team is designing an interface (BBonFHIR) that puts Medicare beneficiaries in control of connecting their data to third-party applications.<sup>vii</sup> In this model, beneficiaries are empowered to select

*Beneficiaries will be able to donate their information effortlessly, the same way that you might connect an application in social media platforms. This will allow patients the option to donate their data to research, 3<sup>rd</sup> party commercial health applications, or doctors who are outside of their core record system*

participation in research based on their preferences (e.g., enrolling in a genetic cancer research study because of a familial history of cancer).

Research study investigators will be able to direct interested beneficiaries to a web page or app that guides the beneficiary through the process of granting permission to investigators to access their Medicare part A, B, and D data, held at CMS. This dramatically simplifies acquisition of beneficiary claims information to support beneficiary participation in clinical research studies. Once complete, the BBonFHIR Service API will allow researchers to selectively pull beneficiaries' data for specific research needs. For example, queries could be constructed to access all Part D medication claims and all claims occurring since the last data request.

The project objectives are to:

- Develop the BBonFHIR service and publish the code on open source software;
- Pilot the BBonFHIR service;
- Publish a technical paper on CMS Blue Button Structured Data;
- Launch full production of the BBonFHIR service; and
- Promote the availability of BBonFHIR to external sources.

This project will contribute to building data capacity for PCOR by working towards enabling the use of enhanced publicly-funded data systems for research, which are key sources of data for PCOR.

**Accomplishments and Deliverables:** During this first phase of the CMS project, the first two objectives were addressed. The pre-pilot BBonFHIR service code was published to GitHub, a repository of open source software code, and a working BBonFHIR prototype that included sample data for 11,000 synthetic patients and 1 million claims was implemented in a secure cloud environment. The second objective and its related task are to launch the pilot BBonFHIR service. CMS has completed pilot testing and produced a report describing pilot test results for public posting on the ASPE website. In order to promote the BBonFHIR service to developers from the private sector and research communities, the project team also held a Code-a-Thon during the 2016 Datapalooza conference where 11 teams created application prototypes using the synthetic data provided through the BBonFHIR prototype.

**Project Collaborations:** BBonFHIR will deliver a service that can be more easily integrated with other applications and services. OS-PCORTF funding will support CMS integration of BBonFHIR into the Sync for Science registration process and leverage project and standards across initiatives. Integrating the BBonFHIR application will facilitate Medicare beneficiaries donating their CMS claims data for precision medicine research.

*When patients can pair their Medicare data with data from other sources the value of the data is magnified.*

**Period of Performance:** 4/30/16 – 4/29/18

**Federal Point of Contact:** Christine Cox

## VI. Food and Drug Administration

The FDA has six active OS-PCORTF projects: 1) Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research; 2) Cross-Network Directory Service; 3) Development of a NLP Web Service for Public Health Use; 4) Source Data Capture from EHRs: Using Standardized Clinical Research Data; 5) Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data; and 6) Utilizing Data from Various Data Partners in a Distributed Manner. Cooperatively funded projects (NLP) and are summarized jointly (see Cross-Agency Funded Projects).

### Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research

A critical component of the FDA's mission is to monitor the safety of its regulated products. The FDA has developed a robust national electronic surveillance system, Sentinel, for monitoring FDA-regulated medical products and devices. Sentinel, is a distributed research network, using existing electronic health care data from multiple sources to support FDA surveillance activities. In a distributed network, data is held locally at the data partner site, and queried to request information on a specific research question. Currently, Sentinel is comprised largely of administrative and claims data from health insurers (i.e., data partners) and there is no way to incorporate information from patients.

**Project Purpose and Goals:** This project will create the infrastructure for collecting data from patients through a mobile device application, allowing patient-generated data to be linked with a single data partner that participates in the Sentinel distributed network. The project will develop and pilot a mobile application to capture data from pregnant women. As a result, researchers will be able to query Sentinel for both the new patient-generated health data (PGHD) and the data routinely captured by the Sentinel data partner.

#### **Evidence Generation in Practice**

- *Since pregnant women are often excluded from clinical trials, studying outcomes after the use of medical products by pregnant women is important to raise the level of evidence associated with the use of medical products in this population..*

The project objectives are to:

- Identify and recruit pregnant women to participate in the data partner pilot;
- Develop a mobile device application; and
- Link data provided from patients through the mobile application with existing data from the pilot data partner collaborator.

This effort will serve as a pilot for future applications. For example, the web-based interface has a customizable design; therefore, the mobile application could be re-configured to collect information from another patient population or a disease cohort so that other researchers or HHS agencies could answer other research questions of interest.

**Accomplishments and Deliverables:** All contributing project partners have been identified, including the data partner pilot entity. The data partner is currently working on the Institutional Review Board (IRB) application required to contact patients to facilitate testing of the mobile application and the official data collection period.

The FDA has drafted the requirements and visual design for the mobile application, web portal, and data storage system.

**Period of Performance:** 7/15/15 – 6/30/18

#### ***Active Patient Engagement***

- *Patient representatives are on the panel developing the medical/epidemiological data elements and interface to capture pregnancy exposures, outcomes, confounders, and risk factors.*
- *Pregnant women will conduct user acceptance testing of the application.*

**Federal Point of Contact:** David Martin

### **Cross-Network Directory Service**

The FDA is tasked with developing and implementing a Cross-Network Directory Service (CNDS) that addresses the stand-alone nature of existing distributed research networks and barriers to working across these networks. Distributed research networks facilitate large scale comparative and effectiveness studies by allowing researchers to send data queries to multiple organizations, while allowing those organizations to maintain possession and protection

over the data they house. A CNDS will enable individual networks to become a community of interoperable networks, allowing end-users to participate in and move around from network to network as research needs dictate. This product will enhance network scalability and enable data partners to engage with multiple networks by allowing them to decide their level of participation and their governance rules.

*CNDS can be described as an “online community for data organizations...” starting as a private community open to small groups and organizations within PCORnet, with the potential to be open to other users in the future as governance and adoption are expanded.*

**Project Purpose and Goals:** This project will create an open source interoperable service that allows data partners to participate in multiple data research networks, query across the networks, and share analytic capabilities and knowledge across networks. The project will be piloted across two existing networks: FDA’s Sentinel System and the Patient-Centered Outcomes Research Institute’s (PCORI) National Patient-Centered Clinical Research Network (PCORnet).

The project objectives are to:

- Identify the key functionalities and an overarching technical design for a CNDS;
- Develop and test a detailed design for the CNDS with at least two existing distributed research networks; and
- Conduct additional analyses of the robustness of the CNDS and produce user materials.

This project will produce the following deliverables: a CNDS Software and Source Code; a report on the Failure Mode and Effects Analysis; technical documentation; and user documentation. Additionally, the

information gained from the CNDS will help inform additional FDA projects, including the Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data.

**Accomplishments and Deliverables:** The design phase of the project has been completed; which included identifying and conveying business requirements, determining key functions and usability needs, software design, and producing a Use Case Summary report. CNDS software development has begun, with the first of four components and prototype user interfaces already completed. Additionally, PCORnet and Sentinel partners for pilot testing have been identified.

**Period of Performance:** 7/15/15 – 6/30/18

**Federal Point of Contact:** Carlos Bell and Aaron Niman

## Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data

The current information systems and data models that define both clinical care and clinical research are disparate in many ways. Thus, using data from EHRs and electronic data capture systems (EDCs) necessitates a detailed understanding of each data source. Moreover, even data sources that are from the same database can have extensive differences in coding, schema, format, and usage.<sup>viii</sup> As clinical research and clinical trial data are now mostly collected through EHRs and EDCs, a guideline or protocol for the use of health care data use in clinical research and trials could be helpful to enable PCOR studies. In 2013, FDA published guidance on Electronic Source Data in Clinical Investigations,<sup>ix</sup> which provides recommendations on the capture, review, and retention of electronic data in FDA-regulated clinical trials. To provide a working example of the FDA's recommended approach, the FDA has initiated this project, which will provide a real world demonstration of their guidelines and commitment to EHR-to- EDC in clinical research environments.

**Project Purpose and Goals:** This project will create a single point data capture approach from the EHR to EDC systems using the Retrieve Form for Data Capture (RFD) standard. PCOR stakeholders will be provided with a cloud-based Health Insurance Portability and Accountability Act compliant and 21C.F.R. Part 11-compliant tool to seamlessly integrate EHR and EDC systems.

The project objectives are to:

- Demonstrate an end-to-end, EHR-EDC, standards-based technology solution for the capture and transmission of regulated clinical research data;
- Assess the value of the standards-based technology solution processes for FDA inspection and reconstruction of clinical investigations;
- Assess the impact of the standards-based technology solution on current processes at the sites, and on remote monitoring of the study;
- Develop guidelines for future implementations in both health care and clinical research;
- Provide recommendations for the improvement of existing standards and implementations including standards-development organizations and structured data capture; and
- Develop a general framework (technologies, processes, policies, governance, and standards) for the use of electronic structured data capture in regulated clinical trials.

This project will produce the following deliverables: a demonstration project plan for end-to-end standards; a publicly available report of the gap analysis between CDEs collected in a health care setting by EHRs versus CDEs required for regulated clinical research; publicly available guidelines document for use by PCOR researchers; a report detailing recommendations to the Clinical Data Interchange Standards Consortium and ONC's Structured Data Capture initiative; and a publicly available paper laying out a framework for structured data capture in regulated clinical trials.

**Accomplishments and Deliverables:** This project was awarded at the end of the fourth quarter of Federal Fiscal Year 2016. The first deliverable, a demonstration project plan for end-to-end standards and metrics to be collected during the demonstration, is expected within Quarter 2 of Federal Fiscal Year 2017.

**Period of Performance:** 8/26/16 – 9/30/17

**Federal Point of Contact:** Mitra Rocca

## **Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data**

Currently, no standards exist for describing and characterizing the quality and completeness of electronic health data. In order to best utilize data, it is imperative that researchers and investigators know the fitness-for-use and reliability of data. In order to address this gap, the FDA will build upon the connections and knowledge gained from the FDA CNDS to discover and design data quality metrics for metadata standards; standards that would further regulate the “data about data” helpful in tagging, structuring, formatting, and coding data according to its categorization and fitness.<sup>x</sup> Researchers affiliated with research networks will be the primary audience for this system. However, researchers (outside distributed research networks) with access to clinical data partners will also be able to utilize the CNDS to describe their data sources in a standardized format. The standardized format will include a way for researchers to describe data sources, domains, attributes and other critical data characteristics, enabling cross-data partner comparisons that will help researchers identify and use fitting data sources, matching data to the needs of the research team.

**Project Purpose and Goals:** This project will create and implement a metadata standards data capture and querying system for: 1) data quality and characteristics; 2) data source and institutional characteristics; and 3) “fitness-for-use”.

This project will produce the following deliverables: metadata standards and technical specifications for implementation; development and testing of the metadata standards data capture and querying approach; implementation of the new metadata standards in at least two distributed networks; and incorporation of the standards in an open-source software release.

**Accomplishments and Deliverables:** This project was awarded at the end of the fourth quarter of Federal Fiscal Year 2016 and is in the beginning of its first phase. The first deliverable, a document of metadata standards and technical specifications for implementation and a dictionary describing each metadata standard, is expected Quarter 3 of Federal Fiscal Year 2017.

**Period of Performance:** 8/20/16 – 9/30/19

**Federal Point of Contact:** Carlos Bell

## Utilizing Data from Various Data Partners in a Distributed Manner

Currently information on a patients' health care is captured across various data sources. Health care information about diagnosis, treatment, and management of disease and illness are available through claims data submitted by health care providers to private and public payers. However, more nuanced clinical data on patients including laboratory values, health outcomes and details of care may only be present in their medical records or registries designed to collect information on diseases, procedures, or devices. The ability to link data across health care databases would provide more robust cross-sectional or longitudinal patient profiles, enhancing secondary uses of electronic health care information for research purposes, and improving access to information that would not be present in claims or registry data or EHRs alone. In order to address this gap, the FDA seeks to build upon previous distributed linear regression analysis efforts by developing enhanced analytic capabilities and fully automating distributed linear regression analysis of patient data across organizations. This functionality will be enabled through a new open source software application that will allow stakeholders within a distributed data network to access and analyze data within a single dataset.

**Project Purpose and Goals:** This project will develop and test the capability to conduct timely and secure distributed regression analysis in distributed data networks. Additionally, it will explore the feasibility of creating virtual linkage capabilities to: 1) utilize data from multiple data sources with unique populations (*horizontally partitioned data*); and 2) utilize data for one specific patient with information at different institutions (*vertically partitioned data*) through a unique key used to identify the patient. This would allow research networks to maintain control of patient-level data while generating valid regression estimates within and across networks without the need to transfer protected health information, providing a balance between analytic requirements, patient privacy and confidentiality, and proprietary considerations. The identified use case for this pilot test is the comparison of bariatric surgery survival outcomes, (e.g., body mass index change and hospitalization one year post-surgery).

*A prototype tool that can be converted to allow multivariable-adjusted outcome regression analysis without transferring of patient-level data, providing balance between research needs, patient security, and proprietary concerns.*

The project objectives are to:

- Develop a new open-source software application that will use PopMedNet™ (PMN), an open source software application that enables the creation, operation, and governance of distributed health networks, to automate multi-step interactive processes and allow stakeholders to conduct distributed regression analyses with data from different people held at different institutions without sharing potentially identifiable information across sites;
- Develop this software application so that it can be supported by PMN and can be modified and adopted for non-PMN applications;
- Test the new, distributed regression application in an actual distributed research network;
- Provide technical and user documentation to accompany the new software and allow for its widespread adoption; and
- Explore the feasibility of conducting distributed regression analyses in which data from the same people are held at different institutions.

This project will produce the following deliverables for horizontally partitioned data: develop and test code for multiple regression analyses; software prototype and source code; and a technical report. After completing those deliverables, additional deliverables will be produced for vertically partitioned data: develop and test code for distributed linear regression and a technical report.

**Accomplishments and Deliverables:** The FDA has achieved significant milestone progress towards each of the project deliverables. The FDA has used a flexible and responsive software development methodology called Agile. Application software development is ahead of schedule, including work on the back-end core infrastructure as well as the user interface. Software testing has started and will be an ongoing activity through the agile development cycle. The project team has developed and tested the statistical framework and code necessary to perform the distributed regression analysis. It is anticipated that full automation in a test environment will be available by the end of the second quarter of Federal Fiscal Year 2017.

**Period of Performance:** 7/15/15 – 6/30/18      **Federal Point of Contact:** Carlos Bell and Aaron Niman

## VII. National Institutes of Health

NIH has two active projects: 1) Creation of LOINC Equivalence Classes; and 2) Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-reported Information with Other Data Sets and Assess Its Validity.

### Creation of LOINC Equivalence Classes

Interoperable health information exchange relies in part on common vocabulary standards that allow disparate systems to interpret and exchange information with the same clinical meanings. Logical Observation Identifiers Names and Codes (LOINC<sup>®</sup>), developed by the Regenstrief Institute, Inc., is a universal coding system for laboratory tests and other clinical observations. LOINC is a national standard with widespread adoption and recognition of its utility; for example, LOINC is required under the CMS Meaningful Use program, state and national public health reporting mandates, and in research networks like PCORnet Common Data Model.

However, mapping local terms to LOINC codes can be complex and resource intensive. This complexity can lead to differences in mapping granularity and specificity that limit interoperability. Furthermore, not all important data uses require the full granularity of detailed LOINC names; for some purposes, “equivalencing” LOINC codes may be sufficient. “Equivalencing” or rolling-up a hierarchy groups similar LOINC codes together in clinically relevant and similar groups. This process addresses the important interoperability barriers created when clinical and laboratory terms are mapped to LOINC terms with different levels of granularity.

**Project Purpose and Goals:** The goal of this project is to create a flexible, extensible, and computable mechanism for rolling LOINC codes into clinically relevant equivalence groups that enable more efficient processing aggregation of laboratory data and other data from diverse health information technology (IT) systems. The primary focus of this work will be on laboratory tests.

The project objectives are to:

- Identify high priority and clinically relevant content subsets for representing in the new LOINC hierarchy. To support this work, the team will develop an enhanced software tool that searches the LOINC database and stores the subset for later retrieval. Input on the high priority subsets will be gathered from an expert advisory group and the global LOINC user community;
- Develop a clinically relevant roll-up hierarchy for LOINC terms. The Regenstrief LOINC team will review the candidate subsets to identify “equivalence groups” and define the term attributes necessary to create the group. The goal is to create rule-based group definitions and save them so that they can be updated and re-executed over time and in subsequent LOINC releases; and
- Disseminate the aggregation hierarchy within the main LOINC release distribution. The new hierarchy will be published with each bi-annual LOINC release and available publicly at no cost.

**Accomplishments and Deliverables:** This project was awarded at the end of the fourth quarter of Federal Fiscal Year 2016. The team has identified all members of the subject matter expert advisory committee whose first meeting is scheduled for the next federal fiscal quarter. The team also presented at a public LOINC meeting beginning the process of soliciting input from the global LOINC community. The first LOINC hierarchy release is scheduled for June 2017.

**Period of Performance:** 08/31/16 – 9/30/17

**Federal Point of Contact:** Todd Wilson

## Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-reported Information with Other Data Sets and Assess Its Validity

The ADAPTABLE (Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-Term Effectiveness) trial is the first major randomized comparative effectiveness trial to be conducted by PCORnet. This pragmatic clinical trial will compare the efficacy and safety of two different daily doses of aspirin widely used for patients with chronic cardiovascular disease.<sup>xi</sup> As part of this new type of comparative effectiveness trial, the ADAPTABLE trial encompasses several key features, including enrollment of 20,000 patients across six large health care systems; an internet portal to consent patients and collect

*Integration of patient-reported information and EHR-derived data has the potential to both enhance evaluation of outcomes that are meaningful to patients and to improve data quality and validity for PCOR.*

patient-reported information regarding risk factors, medications, and experiences; and reliance on existing EHR data sources for baseline characteristics and outcomes follow-up.

**Project Purpose and Goals:** Because ADAPTABLE will rely on patients to report key information at baseline and throughout follow-up, it represents a unique opportunity to develop, pilot, and evaluate methods to validate and integrate patient reported information with data obtained from the EHR. The project will generate tools and data standards that could be deployed in other PCOR studies beyond the ADAPTABLE trial.

The project objectives are to:

- Develop, test, and validate metadata standards for patient-reported information to describe the completeness, consistency, and fitness-for-use of patient-reported data in EHR research;
- Evaluate the validity of patient reported data through systemic comparison with EHR data. The project will develop a Patient-Reported Data Assessment Tool to quickly and efficiently evaluate concordance of patient-reported data and EHR data; and
- Develop approaches to resolve inconsistencies between patient-reported data and EHR-derived data.

This project will inform future efforts to synthesize potentially inconsistent data from patient-reported and EHR sources, to identify opportunities to streamline data capture, and to facilitate enrollment in study-specific target populations within larger health systems.

**Accomplishments and Deliverables:** This project was also awarded at the end of the fourth quarter of Federal Fiscal Year 2016. Funding of the ADAPTABLE trial, while causing initial project delays, has resulted in favorable alignment of the ADAPTABLE trial research and NIH monitoring. The ADAPTABLE project team held a kick-off meeting in December 2016. The date for the first NIH deliverable, a peer-reviewed manuscript focused on the first and second objective, is set for April 2018.

**Coordination with Other Federal Agencies:** The patient-reported data elements submitted for inclusion into LOINC will also be published on the PCORnet website and submitted for inclusion in the National Library of Medicine (NLM) CDE Resource Portal and the REDCap shared library to support implementation of patient-reported data elements in future studies.

**Period of Performance:** 8/31/16 – 9/30/17

**Federal Point of Contact:** Wendy Weber

## VIII. Office of the National Coordinator for Health Information Technology

ONC has five active PCORTF projects: 1) Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data; Creating the Foundational Blocks for the Learning Health Care System: 2) Data Access Standards for Electronic Health Records; 3) Creating the Foundational Blocks for the Learning Health Care System: Structured Data Capture; 4) Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy; and 5) Security and Privacy Standards for Patient Matching, Linking and Aggregation. The Privacy and Security Blueprint is a joint project with the CDC. Projects that are cooperatively funded (the Privacy and Security Blueprint) are summarized jointly (see Cross-Agency Funded Projects).

### Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data

PGHD are “health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern.”<sup>xii</sup> PGHD are different from data gathered in clinical settings and through provider encounters in two ways: 1) patients are responsible for capturing the data;

and 2) patients decide how to share or distribute the data with clinicians in other health care settings and others. Examples of PGHD include blood glucose readings using a home blood glucose monitor, mobile fitness tracking apps, and survey data collected from questionnaires in both paper and electronic format. The use of PGHD offers a unique opportunity to fill in gaps in information and provide a more comprehensive picture of ongoing patient health for use during care, resulting in potential cost savings and improvements in health care quality and outcomes, care coordination, and patient safety. Understanding the potential benefits and concerns surrounding PGHD, ONC started a series of activities in 2012 to increase the capture and use of PGHD to meet the needs of patients, researchers, and providers. Lessons learned from those activities informed this project, which aims to identify best practices, gaps, and opportunities for progress in these seven PGHD topic areas:

1. Collection and validation of data and tools, including mobile health applications and wearable technologies;
2. Data donation- consent management/patient preferences and patient expectations;
3. Legislative and regulatory gaps;
4. Ability to combine PGHD data with medical record data in multiple ways;
5. Patient recruitment;
6. Data interoperability; and
7. Big data analysis.

**Project Purpose and Goals:** This project will develop a policy framework for the use of PGHD in research and care delivery that addresses data collection tools, data donation policies, regulatory gaps, the combination of PGHD with medical record data, and interoperability of PGHD across multiple health information systems and devices. Additionally, the project will conduct pilots that test the concepts and implementation of the policy framework.

The main project objective is to develop a framework that improves the current state of PGHD being collected in a variety of formats, both paper and electronic, through the following goals:

- Patients, providers (including safety-net providers), and researchers can better trust, share, and use the data and tools for their own purposes; and
- Delineation of a path forward for digitizing more PGHD so that it can be used and shared more easily, particularly for research.

This project will produce the following deliverables: a Policy Framework White Paper and a report on piloting capture and use of PGHD.

**Accomplishments and Deliverables:** ONC has made considerable progress in developing their PGHD Policy Framework White Paper. The project team has performed research and drafted environmental scans and literature reviews across the chosen topic areas for capture and use of PGHD. Additionally, interviews with subject matter experts were performed to provide further information about activities at organizations using PGHD. The results of these efforts were used to develop a draft white paper<sup>xiii</sup> of the final framework deliverable, available now for public review and comment. Public feedback and the pilot demonstration results will inform the final white paper outlining ONC's proposed PGHD Policy Framework. In August 2016, pilot testing started at two unique sites that will bring diverse information and perspectives to the landscape of PGHD. Working with these teams, ONC will test the framework in

remotely collected PGHD across patients in a diabetes-focused research study and gather PGHD across multiple medical conditions, respectively.

**Period of Performance:**

6/16/15 – 6/30/18

**Federal Point of Contact:**

Michelle Murray, Caroline Coy, and Jan Hice-Smith

## **Creating the Foundational Blocks for the Learning Health Care System: Data Access Framework**

EHRs have the potential to provide useful information for PCOR by aggregating patient data across disparate systems. Many current PCOR projects rely on patients' clinical data collected from site- or system-specific EHRs. While other ONC activities are focused on standardizing the way the EHR data is collected within EHRs, building on near widespread adoption of EHRs utilizing standard data and vocabulary requirements as specified by Meaningful Use and MACRA, researchers also need standardized ways to access that data. This project seeks to make it easier to extract data from an EHR in a consistent and reproducible way and is a critical next step to enabling and simplifying data aggregation across widely distributed EHR systems (i.e., distributed population queries).<sup>xiv</sup> To accomplish this goal, ONC will develop an API that will "connect" to a provider's EHR to extract data in a standard way. An API is a technology that allows one software program to access the services provided by another software program.<sup>xv</sup>

**Project Purpose and Goals:** The goal of this project is to develop technical standards for how health care providers, researchers, and the public health community access and extract data from EHRs in order to conduct PCOR. The project has three phases each building on the capacity developed in a prior phase.

The project objectives are to:

- *Local Access API Initiative:* Develop an API that will allow providers to access data in their own EHR in a standardized way. The API will focus on a standard way to extract patient, practice, and outcomes data from the EHR. ONC will develop two use cases to develop a standard way to query an EHR and a standard format for how that data is returned.
- *Secure Enterprise Access Initiative:* Using the local access data standards, add a standardized interface to allow researchers outside of a particular organization with remote access to another organization's EHR data. This phase will focus on robust, secure authentication and authorization of the researcher for access to EHR data.
- *Distributed Access Initiative:* Leveraging both the Local Access and Secure Enterprise Access standards, ONC's work will focus on the development of a governance structure for standardizing distributed research queries. This phase will build upon ONC's former Standards and Interoperability (S&I) Framework Query Health Initiative, aimed at identifying standards and services to support distributed population queries.

In each of the phases, ONC will establish draft standards through an open and consensus-based standards organization, test and pilot the standards, and develop testing and certification tools to support consistent and interoperable implementations.

**Accomplishments and Deliverables:** Recent work has focused on Phase Three, the Distributed Access Initiative, which leverages the standards from the Local Access API and the Secure Enterprise Access phases that were completed in Federal Fiscal Year 2014. The Data Access Framework standards have been defined within two implementation guides, the IHE DAF Profile<sup>xvi</sup> and the HL7 FHIR DAF Profile.<sup>xvii</sup> ONC completed pilot testing for four Distributed Access use cases.

**Period of Performance:** 11/1/13 – 9/30/17

**Federal Point of Contact:** Farrah Darbouze

## Creating the Foundational Blocks for the Learning Health Care System: Structured Data Capture

EHRs have the potential to provide useful information for PCOR. Many PCOR studies rely on patients' clinical information collected and stored in EHRs. The use of this information for PCOR would be increased if relevant patient data (e.g., lab-test results) were defined and collected in a common way across research studies using structured data definitions (i.e., common data elements) that comply with the consensus-derived health data standards (e.g., LOINC, Systematized Nomenclature of Medicine (SNOMED)) which were included under the CMS EHR Incentive Program. Development and adoption of CDE standards will enable interoperability in real-world settings to support PCOR.

**Project Purpose and Goals:** This project will identify and develop the functional and technical specifications necessary to enable an EHR system to retrieve, display, and fill a structured form or template, and store and submit the completed form to an external repository. The goal of this project is to develop, pilot, and ballot technical data standards for CDE as well as an electronic template for use in case reporting. Electronic case reporting refers to the ability of an EHR to automatically identify and report specific cases and submit this information in a particular format or template to an end-point (e.g., clinical research, public health registry surveillance system).

*The standard balloting process improves the proposed standards and is necessary to ensure buy-in from the vendor and user communities, and hence, widespread use of the standards.*

The project objectives are to:

- Conduct an environmental scan of existing technical standards and current use of data elements in research settings;
- Create a detailed use case document to guide the standards development process;
- Prepare a reference implementation guide for each of the technical standards; and
- Develop, select, validate, and ballot standards for CDEs and the electronic case-reporting template. This includes CDE standards for the data elements for the form, the structure and design of the form, a standard for how the EHR interacts with the form, and a standard to enable these forms to auto-populate with data extracted from the EHR.

**Accomplishments and Deliverables:** ONC has made significant progress towards balloting data standards for structured data capture in EHRs. ONC completed its environmental scan and resulting summary of the current landscape of available technical standards available to develop the electronic

case reporting form. A detailed use case document was developed to provide the basis for further analysis of the technical standards needed to support practical interoperability for clinical data generated in the health care settings with research data.

ONC has developed and balloted four standards for the CDE and electronic case report form (eCRF) through two standards organizations—HL7 and Integrating the Healthcare Enterprise (IHE) frameworks. The four standards are: (1) standards for common data elements, (2) template (form) for data entry, (3) methods for pre-population or auto-population, and (4) standards for transmission or movement of gathered information. Pilot testing of the standards and supporting reference implementation guides were performed with two test sites. During the final project funding quarter, ONC will update the standards based on lessons learned from the pilot implementations and then focus efforts on dissemination.

**Project Collaborations:** Coordinate with the NLM's efforts to develop a repository for CDEs and eCRFs. ONC has completed the mapping of the AHRQ Data Elements to the Structured Data Capture CDE standards format and has uploaded the format to the NLM repository.

**Period of Performance:** 7/15/13 – 9/30/17

**Federal Point of Contact:** Farrah Darbouze

## Security and Privacy Standards for Patient Matching, Linking and Aggregation

Linking and aggregating patient data from disparate sources cannot be done without “matching” these data in a secure manner that protects patient privacy. Effective patient matching allows users of the data (whether patients, providers, or researchers) to draw correct inferences using the data that have been linked. Without accurate patient matching, stakeholders may inadvertently draw inaccurate conclusions that could significantly impair patient safety and privacy.<sup>xviii, xix</sup> Challenges arise when policies on data use and patient matching are in conflict.

Aggregating records requires strong and standardized methods (i.e., patient matching algorithms) to ensure that data on the same patient are kept together, and that data on different patients are kept separate. These patient matching algorithms rely on analyzing both unique identifiers and personal characteristics of the patient record (i.e., patient attributes) to match records. The accuracy and reliability of patient matching results vary widely due to inconsistent data quality and the variability of algorithms used in matching. For instance, one may believe “John Smith” and the likely misspelled “Jon Smtih” to be the same person. However, even small differences introduce the chance that two patient records do not correspond to the same patient. Measuring this risk and developing risk thresholds is challenging and dependent on the end use of the linked datasets. These issues may be further exacerbated when information collected in the EHR is incomplete making it more difficult to uniquely identify patients across data sources. While there are data exchange standards in place to electronically exchange information across health IT systems, there are no widely adopted standards that consistently address patient identity, authorization, and consent.

*It is critical to have the ability to uniquely match patients for any infrastructure building around data sharing.*

**Project Purpose and Goals:** This project will identify the best patient attributes to address the challenge of linking patients' data across research, clinical, and claims data sets in order to support the PCOR data infrastructure that enables standardization and sharing of patient data across organizations. This project includes work along five distinct tracks.

The project objectives are to:

- Improve data quality by standardizing patient attributes and algorithms that can be used to reliably perform patient matching across clinical and claims data sets to improve algorithm match rates. A use case, environmental scan, standards list, and a specifications/implementation guide will be developed for this portion of the project. All the patient matching standards will also be balloted through a recognized standards development organization and piloted with a health IT vendor;
- Create an open source visual tool for patient matching and aggregation. ONC will repurpose an existing set of open source tools, developed for creating synthetic patient records for testing clinical quality measures, to create a patient matching toolkit. The toolkit will allow researchers to: 1) inspect patient match results; 2) quickly create test data for sharing; 3) incorporate results from clinical and claims feeds and PCORnet; and 4) model new patient attributes;
- Create a privacy and security API or PCOR infrastructure security "layer." The privacy and security "layer" (i.e., authentication, authorization, consent, and data provenance) will ensure that the data are being matched: 1) with the appropriate consents; 2) by a user who is authorized to do the matching; and 3) with authentication (i.e., validation that people are who they say they are) of those trying to access the data. The API work will provide the specifications for this layer along with corresponding implementation guides to manage the authentication, authorization, and consent policies necessary for all participants (patients, researchers, providers) to safely retrieve and contribute patient data to the PCOR infrastructure;
- Include clinical data research networks in the piloting and testing of the proposed standards and services; and
- Integrate the National Plan and Provider Enumeration System (NPPES) provider identification as an additional attribute to improve patient matching across sources.

**Accomplishments and Deliverables:** A number of open source tools and resources have been developed under this project. ONC secured a commitment of participation from a Community Health Applied Research Network partner to pilot all aspects of the algorithm, data management, and the privacy and security layer at six sites. The pilot testing is nearly complete; lessons learned will be documented in a key findings report planned for release in July 2017. ONC is working with a subcontractor to develop patient matching testing services. Work continues on the development of a series of patient matching use cases for the privacy and security API that demonstrate the use of these security protocols for both clinical and research use. ONC is utilizing the Health Relationship Trust (HEART) Profile, an authentication and authorization protocol to facilitate more consistent user access management. The API development work is being furthered by open vendor challenges (e.g., the Move Health Data Forward Challenge). ONC has completed its deliverables for the NPPES work to support the provider directory integration of the NPPES. As a result, ONC and the Federal Health Architecture have formed a collaborative, the Healthare Directory Technology Learning Community, to explore health care directory governance and sustainability issues.

**Period of Performance:** 6/16/15 – 9/30/18

**Federal Point of Contact:** Debbie Bucci

## **IX. Cross-Agency Funded Projects**

Two projects are jointly funded collaborations between federal agencies – the Development of a NLP Web Service of Public Health Use and the PCOR Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy. While working collaboratively to accomplish shared goals, the federal agency awardees have distinct project activities.

### **Development of a Natural Language Processing Web Service for Public Health Use**

The Development of a NLP Web Service for Public Health Use is a joint project between the CDC and the FDA. NLP identifies and codes isolated clinical concepts in order to extract meaningful information and represent events from free-text.<sup>xx</sup> The CDC and FDA are leveraging a recently developed NLP system to “translate” unstructured, free-text data submitted to existing agency surveillance systems into a structured form.

In the U.S., central cancer registries collect, manage, and analyze longitudinal data about cancer cases and cancer deaths. Cancer data are collected from multiple sources such as hospitals, laboratories, physician offices, and independent diagnostic and treatment centers. Hospital reporting of cancer cases has been standardized for over a decade; however, as the provision of cancer care has shifted away from the hospital, registries have had to expand their data collection efforts to include data from non-standard systems that contain large amounts of unstructured data. The CMS Meaningful Use EHR Incentive Program, established under the Health Information Technology for Economic and Clinical Health (HITECH) Act, established support for widespread adoption of EHRs and data and transmission standards to facilitate interoperable data exchange. The public health measures under Meaningful Use included specific cancer registry reporting requirements that facilitated the collection of structured cancer data from provider EHR systems. Additionally, over 90% of cancer cases are diagnosed using methods that generate pathology reports that are often documented in the form of unstructured text. The process of abstracting these crucial cancer data is very labor intensive and expensive.

This unstructured data limits the ability of researchers to analyze the information without manual review. Similarly, a considerable amount of clinical information submitted to the FDA Spontaneous Reporting Systems is unstructured. One of the FDA’s major responsibilities is the post-marketing safety surveillance through the review of spontaneous reports submitted to the Vaccine Adverse Event Reporting Systems (VAERS) and the FDA Adverse Event Report System (FAERS). The adverse events description in VAERS and FAERS data are coded to the Medical Dictionary for Regulatory Activities (MedDRA) terms—a library of standardized medical terminology to facilitate sharing of regulatory information (e.g., registration, documentation, and safety monitoring) for medical products used by humans both before and after the product has been approved for use.<sup>xxi</sup> However, a considerable amount of clinical information in both systems is either not coded (e.g., medical and family history) or is not linked to the MedDRA codes (e.g., exact time information for each symptom). Additionally, there may be duplicate entries for the same event, a phenomenon that impacts the surveillance process, requiring manual review of submitted reports to trace the adverse event.

**Project Purpose and Goals:** This project proposes to develop an NLP service that will be accessible and publicly available to researchers on the Public Health Community Platform (PHCP)—a cooperative platform for sharing interoperable technologies to address public health priority areas aimed at improving population health outcomes and health equity (e.g., tobacco use).<sup>xxii</sup> The NLP service will enable functionality that processes spontaneous report narratives, extracts clinical and temporal information from the text, formats the data for presentation, and maps unstructured medical concepts (e.g., cancer data and safety surveillance data) into structured and standardized data (i.e., International Classification of Diseases 10<sup>th</sup> Edition Clinical Modification (ICD-10-CM), LOINC, SNOMED, and MedDRA).

As a joint project, some project deliverables will be coordinated by a lead organization; these responsibilities are noted below.

The project objectives are to:

- Conduct an “as-is” environmental scan and literature review of existing NLP algorithms, methods, and tools to inform the development of the NLP Web Service (FDA lead; CDC contributor);
- Design the NLP Web Service technical requirements (CDC lead; FDA contributor);
- Build structured datasets using CDC and FDA resources to capture data;
- Pilot the NLP Web Service on the PCHP;
- Evaluate the pilot; and
- Release the final NLP Web Service.

**Accomplishments and Deliverables:** The first deliverable, the Environmental Scan Report has been completed. The report includes a complete inventory of the existing NLP algorithms. The findings of the report will support the design of the system architecture. Additionally, a manuscript detailing the literature review is being drafted with a planned submission for publication in early 2017.

**Coordination with other Federal Agencies:** The CDC and FDA met with the National Cancer Institute for a full-day session to discuss opportunities for collaboration.

**CDC Period of Performance:**  
6/1/16 – 5/31/18

**FDA Period of Performance:**  
6/1/16 – 5/31/21

**CDC Federal Point of Contact:**  
Sandy Jones

**FDA Federal Point of Contact:**  
Taxiarchis Botsis (FDA)

## **Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy**

The Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use and Use of Technology for Privacy project is a joint project between the CDC and ONC. Patient-level data are essential to understanding and improving health outcomes. These data must be made available to researchers in a way that ensures the protection of patient privacy while providing sufficient granularity to allow meaningful research questions to be assessed. However, current laws and policies around the use

of patient level data are nuanced and sometimes conflicting, creating confusion for researchers, providers, and patients.

Making PCOR data-sharing a reality requires addressing many privacy and security-related policy issues. This project is addressing these issues by: 1) conceptualizing and developing a privacy and security data infrastructure blueprint; 2) conceptualizing and developing the legal analysis and ethical frameworks needed to balance individual privacy rights with data use, sharing, and disclosure for PCOR; and 3) identifying, refining, harmonizing, validating, recommending, and piloting standards that support individual consent and preferences for research.

**Project Purpose and Goals:** ONC will develop a privacy and security data infrastructure blueprint, legal analysis, and ethical framework to address the many legal and privacy and security-related policy issues that affect the use of data for various types of PCOR. The CDC is leading the analysis of the public health legal and ethical analysis for use broadly and will supplement the overall privacy and security data infrastructure blueprint developed by ONC. In particular, the analysis will explore CDC's traditional data collection activities in the context of supporting individually identifiable data for PCOR.

The CDC project objectives are to:

- Conduct an “as-is” analysis of public health laws and ethical considerations that relate to the release and use of CDC public health and health surveillance data. These data include public health data generated or collected by the CDC and generated or collected by other agencies or organizations using CDC funds. The “as-is” analysis includes an environmental scan, analysis of findings, and gap analysis, each of which will result in a report;
- Prepare a final white paper of findings and recommendations for best practice practices for data release beyond CDC, and provide guidance to researchers interested in using CDC data for PCOR; and
- Convene a work group of public health legal and ethics experts to inform the “as-is” analysis and provide a “view” of agency wide data use and report of findings.

ONC project objectives are to:

- *Privacy and Security Data Infrastructure Blueprint:* This portion of the project will identify common PCOR use cases diagrams and user scenarios/stories. These use cases will facilitate privacy and security analysis of the privacy and security requirements for PCOR infrastructure based on the identified user stories. ONC will engage with a broad spectrum of PCOR stakeholders to identify and diagram these use cases.
- *Legal Analysis and Ethical Framework:* The purpose of this framework is to enable PCOR researchers to analyze and evaluate their data needs against federal privacy, public health, and security laws. This will rely on a consistent understanding of the legal obligations related to privacy, confidentiality, and security of data to determine if and how PCOR can support those obligations. This framework will also include a legal and regulatory gap analysis of privacy and security protections, models of data flows mapping to privacy and security legal requirements, identification and definition of updates to policies and minimum requirements for de-identification and re-identification, and ethical implications (CDC lead) for PCOR data use.

- *Individual Consent and Preferences for Research:* Currently, an individual's/participant's consent/authorization/preference is often obtained through a paper form. As electronic health information exchange continues to evolve and PCOR develops, technology will play an increasing role in electronically capturing, maintaining, and communicating consent for use and sharing health information for research. A number of different models for electronically capturing and managing patient consent exist. No one operating model has emerged as the best practice. Depending on the results of an analysis of these existing models, standards will either be developed or updated and then piloted for electronically capturing, tracking, storing, managing, and transmitting consent for research both within and outside the traditional health care system.

**CDC Accomplishments and Deliverables:** The CDC is nearing completion of the “as-is” analysis. The environmental legal and ethical scan and analysis of findings reports are complete. The environmental scan and analysis focused on data de-identification as a useful tool to further PCOR and the need for clarity surrounding the use of a single IRB versus multiple IRBs. Further, these two reports showed that there is scant overarching guidance on de-identification, which will be part of the focus of the gap analysis. The work group is actively working on the gap analysis of applicable laws, policies, and ethical implications of sharing public health data for PCOR. As part of the “as-is” assessment, the CDC developed three CDC public health data use scenarios to evaluate the applicability of law and policy to those scenarios. Two of the use cases related to existing CDC data collection activities (e.g. nationwide cancer registry report). The third use case explores new data access and use scenarios that “liberate” data for PCOR. The white paper deliverable is on track for completion in Quarter 3 of Federal Fiscal Year 2017.

**ONC Accomplishments and Deliverables:** ONC is making steady progress towards achieving all deliverables with no identified risks or challenges. To date, completed deliverables include the research data use scenarios and accompanying Use Case Report Technical Volume. For the Framework, the team has prepared a legal scoping document as well as the Federal, Legal, Regulatory and Policy Gap Analysis, which identified and defined legal, regulatory, and policy gaps in privacy and security protections.

**CDC Period of Performance:**  
7/10/15 – 9/30/17

**CDC Federal Point of Contact:**  
Carissa Holmes

**ONC Period of Performance:**  
6/16/15 – 9/30/18

**ONC Federal Point of Contact:**  
Devi Mehta

## X. Conclusion

The OS-PCORTF portfolio has several ongoing projects dedicated to building PCOR data capacity. The projects funded through the OS-PCORTF represent a broad range of topics that target multiple components and functionalities needed for expanding data capacity for outcomes research. Critical to building data capacity, these projects address knowledge and technical gap areas in standards development, data governance, and evolving technologies that support novel approaches to collecting, linking, and analyzing data and help advance the goals of the learning health system.

Additionally, the projects are poised to make significant contributions to several recent initiatives including the Precision Medicine Initiative and the 21<sup>st</sup> Century Cures Act. The Precision Medicine Initiative (recently rebranded as All of Us) aims to invite at least a million individuals to contribute their health information for research purposes. The implementation of this program as well as several efforts arising from the 21<sup>st</sup> Century Cures Act will benefit from the knowledge gleaned from the projects described in this portfolio that are addressing issues around patient consent, collection and integration of patient-provided data and privacy, and security and ethical concerns around these data.

There are multiple synergies (e.g., improving the utility of EHR data, collection and use of PGHD), across the project portfolio. These project synergies extend beyond the cross-agency collaborations already underway that are highlighted in this report, but demonstrate the many applications and context in which this work can be expanded within the research community. The tools and products resulting from these funded projects have utility to federal and non-federal stakeholders will be available to the larger research community to support implementation among users. Furthermore, projects that involve developing a service or standard, have plans to pilot test service in real-world situations, producing best practices and implementation resources aimed at improving readily useable solutions.

## Appendix A. OS-PCORTF Project Portfolio

**Table A1.** Active OS-PCORTF Projects

Funded Agency	Project Title
<b>Agency for Healthcare Research and Quality</b>	Harmonization of Clinical Data Element Definitions for Outcome Measures in Registries
<b>Centers for Disease Control and Prevention</b>	Development of a Natural Language Processing Web Service for Public Health Use*
	Improving the Mortality Data Infrastructure for Patient-Centered Outcomes
	PCOR: Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy*
<b>Centers for Medicare and Medicaid Services</b>	Improving Beneficiary Access to their Health Information through an Enhanced Blue Button Service
<b>Food and Drug Administration</b>	Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research
	Cross-Network Directory Service
	Development of a Natural Language Processing Web Service for Public Health Use*
	Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data
	Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data
	Utilizing Data from Various Data Partners in a Distributed Manner
<b>National Institutes of Health</b>	Creation of LOINC Equivalence Classes
	Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-reported Information with Other Data Sets and Assess Its Validity
<b>Office of the National Coordinator for Health Information Technology</b>	Conceptualization of a Data Infrastructure for the Capture and Use of Patient-Generated Health Data
	Creating the Foundational Blocks for the Learning Health Care System: Data Access Framework
	Creating the Foundational Blocks for the Learning Health Care System: Structured Data Capture
	PCOR: Privacy and Security Blueprint, Legal Analysis and Ethics Framework for Data Use, & Use of Technology for Privacy*
	Security and Privacy Standards for Patient Matching, Linking and Aggregation

\* OS-PCORTF project funding awarded to multiple agencies.

**Table A2.** Completed OS-PCORTF Projects

Funded Agency	Project Title
<b>Assistant Secretary for Planning and Evaluation</b>	Beta Testing of the Multi-Payer Claims Data <sup>^</sup>
	Comparative Effectiveness Research Inventory
<b>Centers for Disease Control and Prevention</b>	Expanding Data Collection Infrastructure of the National Program of Cancer Registries for Comparative Effectiveness Research
<b>Centers for Medicare and Medicaid Services</b>	Maintenance and Support of the Chronic Conditions Warehouse for Comparative Effectiveness Research
<b>Health Resources and Services Administration</b>	Strengthening and Expanding the Community Health Applied Research Network (CHARN) Registry to Conduct Patient-Centered Outcomes Research
<b>National Libraries of Medicine</b>	Development of Data Infrastructure for Use of EHRs in Comparative Effectiveness Research*
<b>Office of the National Coordinator for Health Information Technology</b>	Strategic Opportunities for Building Data Infrastructure for Patient-Centered Outcomes Research
	Development of Data Infrastructure for Use of EHRs in Comparative Effectiveness Research*

<sup>^</sup> The Multi-Payer Claims Data (MPCD) project was a \$16 million CMS project with a contract period of performance of 09/14/2010 to 09/15/2013. On 09/24/2012, the contract was modified with ASPE-provided OS PCORTF funding to conduct a Beta Test. ASPE was responsible for leadership oversight of the Beta Testing of MPCD.

\* OS-PCORTF project funding awarded to multiple agencies.

## Appendix B. Abbreviations

AHRQ	Agency for Healthcare Research and Quality
API	Application Programming Interfaces
ASPE	Office of the Assistant Secretary for Planning and Evaluation
ASPR	Office of the Assistant Secretary for Preparedness and Response
CDC	Centers for Disease Control and Prevention
CDE	Clinical Data Element
CER	Comparative Effectiveness Research
CMS	Centers for Medicare and Medicaid Services
CNDS	Cross-Network Directory Service
CTO	Chief Technology Officer
EDC	Electronic Data Capture
EDRS	Electronic Death Registration Systems
EHR	Electronic Health Record
FAERS	Food and Drug Administration Adverse Event Report System
FDA	U.S. Food and Drug Administration
FHIR	Fast Healthcare Interoperability Resources
HEART	Health Relationship Trust
HHS	U.S. Department of Health and Human Services
HITECH	Health Information Technology for Economic and Clinical Health Act
HL7	Health Level Seven International
HP	Office of Health Policy
ICD-10	International Classification of Diseases 10 <sup>th</sup> Edition
IHE	Integrating the Healthcare Enterprise
IRB	Institutional Review Board
IT	Information Technology
LOINC	Logical Observation Identifiers Names and Codes
MACRA	Medicare Access and CHIP Reauthorization Act
NCHS	National Center for Health Statistics
NDI	National Death Index
NIH	National Institutes of Health
NLM	National Library of Medicine
NLP	Natural Language Processing
NORC	NORC at the University of Chicago
ONC	Office of the National Coordinator for Health Information Technology
OS	Office of the Secretary
PCOR	Patient-Centered Outcomes Research
PCORI	Patient-Centered Outcomes Research Institute
PCORnet	PCORI's National Patient-Centered Clinical Research Network
PCORTF	Patient-Centered Outcomes Research Trust Fund
PGHD	Patient-generated health data
PHCP	Public Health Community Platform
PMN	PopMedNet™
SNOMED	Systematized Nomenclature of Medicine
VAERS	Vaccine Adverse Event Reporting Systems

## Appendix C. Glossary

<b>Clinical Element Models (CEM)</b>	An approach to representing detailed clinical data models and the instances of data which conform to these models.
<b>Clinical Data Research Networks (CDRN)</b>	System-based networks (such as hospital systems) that have the potential to become an ideal electronic network, without structural impediments.
<b>Common Data Models (CDM)</b>	An aggregated or centralized data model copies data from original sources and brings and standardizes these data in a centralized place. The copied data can then be queried and analyzed.
<b>Comparative Effectiveness Research (CER)</b>	CER is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat, and monitor health conditions in “real-world” settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.
<b>Data Governance</b>	The process by which stewardship responsibilities are conceptualized and carried out, that is, the policies and approaches that enable stewardship.
<b>Distributed Research Network (DRN)</b>	A DRN is an approach in which data holders maintain control over their protected data and its uses. A DRN features a central portal that performs network functions, such as operations (e.g., workflow, policy rules, auditing, query formation, distribution) and security (e.g., authentication, authorization) and distributed data marts that remain under the control of the data holders.
<b>Electronic Clinical Data (ECD)</b>	Data in electronic form that may include clinical data from sources such as long term care facilities, pharmacies, laboratories, genetic tests, emergency medical services, and other clinical settings.
<b>Electronic Health Record (EHR)</b>	An electronic record of health-related information for a patient that contains information captured in clinical visits, lab and imaging studies, and other information important to the patient’s medical past.
<b>Health Information Technology for Economic and Clinical Health (HITECH) Act</b>	A law passed by Congress in 2009 that authorizes expenditures of approximately \$20 billion over five years to promote the adoption and use of electronic health record technologies that will be connected through a national health information network.
<b>Interoperability</b>	The ability of health information technology (health IT) systems from different vendors to communicate and share information.
<b>Institutional Review Board (IRB)</b>	A group that follows federal regulations, state laws, and institutional policy to review, monitor, and approve research in order to protect the ethical rights and privacy of the subjects involved.
<b>Learning Health Care System</b>	A health care system in which knowledge generation for research, science, quality assessment, outcomes, and safety standards are aligned for improvement and innovation.
<b>Medicare Access and CHIP Reauthorization Act (MACRA)</b>	The Medicare Access and CHIP Reauthorization Act of 2015 ended the Sustainable Growth Rate formula and consolidated three existing quality programs into two new tracks under the Quality Payment Program.
<b>Metadata</b>	The term metadata refers to “data about data”. The term is ambiguous, as it is used to describe two fundamentally different concepts. Structural metadata concerns the design and specification of data structures and is more properly called “data about the containers of data”; descriptive metadata, on the other hand, concerns individual instances of application data, that is, the data content.
<b>Natural Language Processing (NLP)</b>	A computational model that analyzes texts using several linguistic approaches, such as syntax, semantics, and pragmatics, for the purpose of achieving human-like language results.

<p><b>Patient-Centered Outcomes Research (PCOR)</b></p>	<p>Patient-Centered Outcomes Research helps people make informed health care decisions and allows their voices to be heard in assessing the value of health care options. It answers four patient-focused questions:</p> <ul style="list-style-type: none"> <li>■ “Given my personal characteristics, conditions and preferences, what should I expect will happen to me?”</li> <li>■ “What are my options and what are the benefits and harms of those options?”</li> <li>■ “What can I do to improve the outcomes that are most important to me?”</li> </ul> <p>“How can the health care system improve my chances of achieving the outcomes I prefer?”</p>
<p><b>Patient-Generated Health Data (PGHD)</b></p>	<p>Patient-generated health data is health-related data created, recorded, gathered, or inferred by or from patients or their designees to help address a health concerns.</p>
<p><b>PopMedNet™</b></p>	<p>PopMedNet™ is an open source software application that enables the creation, operation, and governance of distributed health data networks through a no cost license.</p>
<p><b>Semantic Interoperability</b></p>	<p>The ability of computer systems to transmit data with unambiguous, shared meaning.</p>

## References

---

- <sup>i</sup> OS PCORTF Measurement and Evaluation Plan Final Report. Prepared by RTI International for the Division of Healthcare Quality and Outcomes Office of Health Policy Office of the Assistant Secretary for Planning and Evaluation (2016).
- <sup>ii</sup> Novak V. et al. Troponin Criteria for Myocardial Infarction After Percutaneous Coronary Intervention. *Arch Intern Med* 2012; 172(6):502-208.
- <sup>iii</sup> Gliklich RE, Leavy MB, Karl J, Campion DM, Levy D. Outcome Measure Framework Design Document. Effective Health Care Program Research Report No. 43. (Prepared by the Outcome DEcIDE Center under Contract No. 290-2005-0035-1). AHRQ Publication No. 14-EHC019-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2014. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).
- <sup>iv</sup> Centers for Disease Control and Prevention National Center for Health Statistics National Hospital Care Survey [website] <https://www.cdc.gov/nchs/nhcs/>.
- <sup>v</sup> HL7 International [website]. [http://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=209](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=209)
- <sup>vi</sup> Centers for Medicare and Medicaid Services Blue Button on FHIR: Improving Medicare Beneficiary Access to Their Health Information [website]. <https://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/Blue-Button/index.html>.
- <sup>vii</sup> Centers for Medicare and Medicaid Services Blue Button on FHIR: Improving Medicare Beneficiary Access to Their Health Information [website]. <https://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/Blue-Button/index.html>
- <sup>viii</sup> Voss, Erica A., Rupa Makadia, Amy Matcho, Qianli Ma, Chris Knoll, Martijn Schuemie, Frank J. DeFalco, Ajit Londhe, Vivienne Zhu, and Patrick B. Ryan. "Feasibility and utility of applications of the common data model to multiple, disparate observational health databases." *Journal of the American Medical Informatics Association* 22, no. 3 (2015): 553-564.
- <sup>ix</sup> Food and Drug Administration Guidance of Industry Electronic Source Data in Clinical Investigations. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf>.
- <sup>x</sup> Baca, Murtha, ed. *Introduction to metadata*. Getty Publications, 2016.
- <sup>xi</sup> Adaptable, the Aspirin Study: A Patient-Centered Trial [website]. <http://theaspirinstudy.org/>.
- <sup>xii</sup> HealthIT.gov Consumer eHealth Patient-Generated Health Data [website]. <https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data>.
- <sup>xiii</sup> "Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024." Accenture Federal Services. [https://www.healthit.gov/sites/default/files/Draft\\_White\\_Paper\\_PGHD\\_Policy\\_Framework.pdf](https://www.healthit.gov/sites/default/files/Draft_White_Paper_PGHD_Policy_Framework.pdf)
- <sup>xiv</sup> ONC Standards and Interoperability (S&I) Framework QueryHealth Initiative Project Charter [website] <http://wiki.siframework.org/Query+Health+-+Project+Charter>.
- <sup>xv</sup> ONC Health IT Joint Committee Collaboration API Task Force. April 19, 2016 [website]. [https://www.healthit.gov/FACAS/sites/faca/files/HITJC\\_Final\\_API\\_Update\\_Presentation\\_2016-04-19.pdf](https://www.healthit.gov/FACAS/sites/faca/files/HITJC_Final_API_Update_Presentation_2016-04-19.pdf)
- <sup>xvi</sup> The Data Access Framework (DAF) Document Metadata Based Access Implementation Guide. IHE Patient Care Coordination US National Extension Implementation Guide. September 2016. [http://ihe.net/uploadedFiles/Documents/PCC/IHE\\_PCC\\_IG\\_DAF\\_National-Extension.pdf](http://ihe.net/uploadedFiles/Documents/PCC/IHE_PCC_IG_DAF_National-Extension.pdf)
- <sup>xvii</sup> U.S. Data Access Framework (DAF) FHIR Implementation Guide (IG) [website]: <http://hl7.org/fhir/2015May/daf.html>.

<sup>xviii</sup> Shah, G.H., Lertwachara, K., and Ayanso, A. Record linkage in healthcare: applications, opportunities, and challenges for public health. *International Journal of Healthcare Delivery Reform Initiatives*, 2010. 2(3): 29-47.

<sup>xix</sup> Dimitropoulos, L. L. Privacy and Security Solutions for Interoperable Health Information Exchange; Perspectives on Patient Matching: Approaches, Findings, and Challenges. Report written for Jodi Daniel and Steven Posnack at the Office of the National Coordinator for Health IT and P. Jon White at the Agency for Healthcare Research and Quality. June 2009.

<sup>xx</sup> Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform.* 2009 Oct;42(5):760-72.

<sup>xxi</sup> International Conference on Harmonization of Technical requirements for Registration of Pharmaceuticals for Human Use (ICH). <http://www.meddra.org/how-to-use/support-documentation/english>.

<sup>xxii</sup> The Public Health Community Platform [website]. <http://www.thephcp.org/home>