Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Signature:

_____ _____

Chade Granderson Date

Bridging Electronic Health Records (EHRs) and Clinical Research Data: An Evaluation of an Academic Medical Center

BY

Chade Granderson

Degree to be awarded: MPH
Executive MPH

Sheri Chernetsky Tejedor MD, SFHM — Date
Committee Chair

Jennifer Prozonic MPH, CHRC — Date
Committee Member

Marc Overcash — Date
Committee Member

Laura Gaydos PhD — Date
Associate Chair for Academic Affairs, Executive MPH Program

Bridging Electronic Health Records (EHRs) and Clinical Research Data: An Evaluation of an Academic Medical Center

By

Chade Granderson

B.S., Psychology, Georgia State University, 2008

Thesis Committee Chair: Sheri Chernetsky Tejedor, MD, SFHM

An abstract of a thesis submitted to the Faculty of the Rollins School of Public Health of Emory University in partial fulfillment of the requirements for the degree of Master of Public Health
in the Executive MPH Program 2017

Abstract

Bridging Electronic Health Records (EHR) and Clinical Research Data:
An Evaluation of an Academic Medical Center

BY

Chade Granderson

Background: The conduct of clinical research leads to innovative health solutions that are critical for improving health outcomes and providing scientific evidence needed to support and guide clinical and health policy decisions. As public health has moved towards evidence based scientific standards and interventions, federal rules and regulations regarding the conduct and reporting of clinical research findings also continue to evolve. To assist in meeting these standards and requirements, clinical research studies and the organizational operations surrounding them have grown larger and more complex requiring additional time, staff, and monetary resources.

Purpose: The clinical research infrastructure has become fragmented and perpetuated by the gaps between health records and research data. The primary goal of this project is to propose a systems architecture integrating electronic health records and research data and assess its feasibility within an academic medical center. We also seek to assess whether this integration has the potential to reduce study timelines and errors within research databases. Lastly, we seek to describe barriers to integration.

Methods: A series of data flow analysis and interviews were conducted to understand current research and business workflows and processes within an academic medical center. Additional interviews and a literature review were used to identify and propose an integrated systems architecture and method for pre-populating research databases with data from electronic health records.

Results: A process and systems architecture for auto-populating research databases with relevant data elements from electronic health records was defined allowing for a reduction in data duplication for research staff using the Retrieve Form for Data Capture method. However, several barriers for system integration were identified.

Conclusions: There are several methods for integrating systems, yet identifying a successful architecture for data sharing between health information and research systems remains a challenge due to time, cost, and regulatory constraints. Awareness and advocacy surrounding the requirements and importance of building systems to promote data sharing between electronic health and research records as to contribute to the advancement of public health are needed.

Bridging Electronic Health Records (EHR) and Clinical Research Data: An Evaluation of an Academic Medical Center

By

Chade Granderson

B.S., Psychology, Georgia State University, 2008

Thesis Committee Chair: Sheri Chernetsky Tejedor, MD, SFHM

A thesis submitted to the Faculty of the Rollins School of Public Health of Emory University in partial fulfillment of the requirements for the degree of Master of Public Health in the Executive MPH Program 2017

Acknowledgments

I would like to thank my thesis committee members for their continued support and dedication throughout the development of my thesis project. Their expertise, guidance, and patience has had a tremendous impact on realizing the potential of the knowledge gained through my experience as a public health informatics student.

I would also like to thank my family and friends for their motivational support, prayers, and patience during my time as a student.

Table of Contents

Chapter 1: Introduction

Introduction and Rationale

The conduct of clinical research leads to innovative health solutions that are critical for improving health outcomes and providing scientific evidence needed to support and guide clinical and health policy decisions. Through the thousands of clinical studies that are conducted globally each year, scientists can contribute medical knowledge, develop medicines and devices, vaccinations, and lifestyle recommendations related to the treatment, diagnosis, and prevention of diseases and conditions that affect population health. Clinical research also provides imperative insight to the efficacy, safety, side effects, and long term outcomes from receiving a health treatment or intervention (Health, 2015). This information is crucial to public health by providing insights to the intended and unintended effects of health interventions, defining disease patterns, developing prevention methods, and conducting disease surveillance on the U.S. population.

As public health has moved towards evidence based scientific standards and interventions, a need for better knowledge of disease frequency and distribution, the determinants and consequences of a disease, and the safety and efficacy of treatments has risen (Victora, Habicht, & Bryce, 2004). Federal rules and regulations regarding the conduct and reporting of clinical research findings also continue to evolve. To assist in meeting these standards and requirements, clinical research studies and the organizational operations surrounding them have grown larger and more complex requiring additional time, staff, and monetary resources. However, a fragmented infrastructure and incompatible databases within

research conducting institutions have done little to support the growth of trials (Sung, Crowley, Jr, Genel, & et al., 2003).

Research organizations often rely on several repositories to collect information related to study conduct. Data for patients participating in clinical research studies are traditionally collected from the electronic health record (EHR) and manually transcribed onto paper case report forms (CRFs) as well as entered into electronic data capture (EDC) systems (Murphy, Ferris, & O'Donnell, 2007). Equivalent information can also be collected and tracked in research management systems intended to support organizational operations and report long term health outcomes to sponsors and various public health entities. This duplicative and manual data entry process results in unintentional transcription errors across systems, reduces time spent performing clinical duties, extends study completion timelines, and increases research costs.

The consequences of duplicative data entry collectively hinder the discovery of new and more efficacious health interventions, the influence clinical research has to accurately guide public health, and organizational efficiency. Poor data quality resulting from excessive transcription errors can lead to inaccurate interpretation of study findings that have the potential to impact standard health care services, public health policy and decision making, and the conduct of disease surveillance. Studies that frequently require timeline extensions are at risk for losing funding prior to reaching study endpoints and producing meaningful results. Replicating information into various databases also decreases productivity for clinical staff who often dually serve as principal investigators for multiple studies. Although the issues

surrounding data duplication have been identified, few research conducting organizations have produced system solutions for bridging data gaps.

Review of the Literature

Many resources are invested into conducting clinical research including money, time, and labor of organizations, sponsors, and participants across the country. Therefore, it is imperative that end data is accurate, and resources are spent wisely to manage research activities allowing for a timely conclusion and meaningful analysis. Electronic data capture within clinical research settings has aided in streamlining the data collection process, reduced research costs, and improved quality. Prior to EDC, clinical research associates (CRAs) during site visits located patient charts, transcribed data onto paper case report forms, and lastly mailed these forms to a central location where data required double entry prior to statistical analysis. Queries, or data requiring correction or clarification, were sent via mail between coordinating sites and sponsors (Goodman, Krueger, & Crowley, 2012). With the increase in adoption of EDC for many research studies, it has become easier for study staff to document clinical data for research purposes in a structured manner. The advent of EDC has also allowed for sponsors and clinical research organizations (CROs) which monitor trials to access data as a study progresses. Reviewing these data throughout the duration of a trial may detect anomalies that can be addressed and corrected prior to its conclusion. This has resulted in a reduction in time spent for data cleanup activities post active trial completion (Murphy et al., 2007).

Although EDC has made for advancements in clinical research, there lies an opportunity to make use of existing data. The migration from paper medical records to electronic databases

has been supported by the Health Information Technology for Economic and Clinical Health (HITECH) Act and the concept of meaningful use (Prevention, 2016). Meaningful use incentives aimed at improving efficiency across health care organizations, and ultimately better clinical outcomes for population health, have been a major actor in the movement towards electronic health (EHR) records (HealthIT.gov, 2015). As the use of EHR systems is continually adopted, they often serve as the main data source for clinical research studies. Through data sharing, the vital health information that an EHR system holds presents an opportunity to enhance and improve the conduct of clinical research.

In September 2013, the Food and Drug Administration (FDA) published guidance advocating the need for capturing electronic source data, including data originating in health care systems (U. S. F. a. D. Administration, 2015). The Center for Drug Evaluation and Research (CDER), responsible for ensuring that "safe and effective drugs are available to improve the health of the people in the United States" supports and encourages the use of seamless data exchange (Administration, 2014). This exchange includes the re-use of information from health care systems to clinical research systems allowing for the one time, point of care, entry of data. In effort to achieve this goal, CDER heads collaboration among the regulated industry, EHR and EDC vendors, standards development organizations (SDOs), and several academic medical centers. This collaboration has led to a published paper supporting projects that test the capability and performance of an end-to-end electronic health records to EDC single point data approach in June 2015 (Register, 2015). Federal efforts to support and encourage the utilization of clinical data originating from an EHR system demonstrate the need for integration.

Integrating systems to prepopulate data from EHR to clinical research databases could greatly reduce or eliminate the need for manual transcription. Such integration has the potential to significantly decrease the number of transcription errors between EHR and research databases, thus reducing the time required for data cleanup and reconciliation resulting in shorter research timelines and costs savings. Major EHR vendors Cerner Corporation, Allscripts, and Greenway Health have presented solutions for electronically capturing and prepopulating data elements into research systems during demonstrations hosted by the Clinical Data Interchange Standards Consortium (CDISC) and Integrating the Healthcare Enterprise (IHE). Cerner, in collaboration with Florida Hospital, created a streamlined system to integrate data capture for an investigator initiated research study into the standard workflow using the Millennium EHR and the Cerner Discover*e* Research data capture system (Marsha Laird-Maddox; Susan B. Mitchell, 2014). Similarly, using a solution developed by Siemens and the Frauenklinik of the Technical University of Munich, a study found that between 48 and 69 percent of an electronic case report form (eCRF) could be prepopulated by integrating EHR with a research database (Markus Schmidt, 2009).

As clinical research studies in the U.S. grow in complexity, the costs and time to product development from study initiation continue to rise. Over the past two decades, the cost of developing a drug has increased at a rate 7.4 percent higher than inflation primarily due to clinical research expenses (Collier, 2009). Although there is variance across therapeutic area and phase, it is estimated that introducing a new drug into the market ranges from $161 million to $2 billion (Evaluation, 2014). Contributing to the financial burden are lengthy timelines that accompany establishing the safety and efficacy of a treatment or intervention. A 2003 analysis

conducted by Dimasi, Hansen, and Grabowski indicated that the average time from clinical testing to marketing for a drug is 7.5 years (DiMasi, Hansen, & Grabowski, 2003). The time to marketing is estimated to have increased in more recent years according to the 2015 Prescription Drug User Fee Act (PDUFA) performance report (F. a. D. Administration, 2015). Lengthy timelines create delays in the discovery of innovative treatments for various diseases and conditions that affect population health.

Although technological advances have been developed to demonstrate methods for EHR and research EDC integration, there are major considerations and potential roadblocks that hinder widespread utilization. Data standards vary among health care and the clinical research community. Goodman et al. noted that many organizations have yet to adopt standards to facilitate data sharing, citing it difficult to transition from HL7 Version 2 to Version 3 (Goodman et al., 2012). Systems that share protected health information (PHI) must also remain in compliance with federal regulations which includes ensuring electronic data is confidential, allowing for provisions for audit trails, providing adequate informed consent describing data uses, access, and risks, as well as maintaining proper security measures (U. S. D. o. H. a. H. S. F. a. D. Administration, 2016). Additionally, many organizations operate clinical research studies across several administrative and registration systems which require integration beyond EHR and EDC. To address the concept of automated integration within an organization that allows for a more cohesive and efficient clinical and research process, an evaluation at the individual organization's workflows and business rules is needed.

Problem Statement

Researchers have identified several causes to increasing clinical research timelines that consequently delay the evaluation of new treatments and therapies, hinder the ability to accurately study long term health effects and accurately monitor disease prevalence, and postpone the creation of standards, policies, and programs aimed at managing population health and bettering outcomes. Major contributors include extensive periods of data collection, monitoring, and data cleanup which also create loss of productivity within a research organization. The adoption of EDC by many sponsors and CROs has significantly reduced these efforts. However, to support the complexity and meet the changing needs of clinical research activities, including the management of administrative tasks, many organizations have developed solutions that operate in disparate and siloed systems. These systems often collect redundant data to be used in various capacities across research operations and the organizational business enterprise.

Electronic health records traditionally serve as the primary source of data that are transcribed and aggregated into one or more research databases. However, data discrepancies within EHR and ultimately among clinical research databases frequently occur creating a need labor intensive data cleanup across systems. Previous studies have calculated error rates between EHR and research databases as high as 27%. Many of these errors are believed to result from missing data values and mistakes made during manual transcription (Goldberg, Niemierko, & Turchin, 2008). Although it is known that cross referencing and manual duplication of data greatly increases the chances for higher error rates, best solutions for preventing errors are largely understudied. Many organizations utilize laborious and time

consuming methods such as double-entry to identify and correct mistakes. Other institutions have developed and adopted costly system solutions to bridge information between EHR and research databases using research warehouses and middle tier applications with varying success rates. However, the best strategy for error prevention and maximizing the use of existing data has yet to be defined and accepted by the research informatics community. An approach for preventing and significantly reducing errors within clinical research databases, meanwhile reducing labor constraints for staff, and improving organizational workflow is needed.

Purpose Statement

The primary goal of this project is to assess whether the automatic transfer of relevant data elements from EHR to a clinical research management system and an EDC system is a feasible solution for improving data quality and reducing timelines for clinical research studies within an academic medical center. We also seek to evaluate whether integrating these systems can improve organizational workflow by reducing or eliminating the need for duplicate data entry. Lastly, we seek to describe a method for systems integration, barriers to systems integration, and discuss an infrastructure needed to overcome these barriers.

Significance Statement

Clinical research is a critical component for monitoring and improving public health, yet the lack of integrated systems within research organizations impairs the ability to deliver timely and accurate study information. Without system integration, institutions remain vulnerable to high error rates inherently resulting from the manual transcription of data requiring prolonged

data cleanup periods post study completion. Timely and reliable data is essential to the conduct of clinical research that births novel health treatments, information needed to support public health policy and decision making, and monitoring long term health outcomes of past and recent interventions. Inaccurate data leads to false interpretation of study results, and therefore a miscalculated perception of the efficacy of health treatments and programs. As thousands of clinical research studies are conducted each year, poor quality data has the potential to affect millions of people with health conditions who rely on developments from research to improve their quality of life. System integration has the potential to transform the way clinical research is conducted, and further bridge the relationship between research and public health by delivering faster, more accurate results, impacting the health of our nation.

Chapter II: Methodology and Approach

Introduction

The purpose of this chapter is to describe the methodology used to develop evaluate the research workflows, business processes, and enterprise architecture for the integration of electronic health records and clinical research systems, utilizing Emory University as a model for academic medical centers. This chapter also describes the approaches taken to propose a solution for integration.

Project Design

A mixed methods approach was applied to develop a solution proposing the integration of electronic health records and clinical research data. A business case was chosen as the approach for developing and presenting the proposed solution as it is a frequently used tool for

project justification in the field of public health informatics. For the purposes of this project, it also served a guide for identifying project requirements across the enterprise. The Centers for Disease Control and Prevention (CDC) Unified Process templates were used to develop the business case ((CDC), 2010). Emory University was chosen as an organizational model for this project as it is a leading academic medical center in conducting clinical research studies (Emory, 2017). Accessibility to systems was also taken into consideration in choosing Emory as the model. No protected health information (PHI) was accessed for the purposes of this project. This project was exempt from institutional review board (IRB) review.

Procedures

Access to the electronic health records system for Emory Healthcare and the clinical research management systems of interest was obtained prior to review. All systems were examined to develop an understanding of their purpose within the organization and the types of data collected. Unstructured interviews were conducted with subject matter experts in the Emory University Office for Clinical Research and the Emory Healthcare Information Services Department to describe the current state of integration, research workflow, and research infrastructure in relation to business processes. A data flow analysis was completed on the research process to identify the types of data being captured, its location and movement among systems, and the actors involved in data capture.

Options for integrating systems were then assessed through several methods. An unstructured interview was conducted with a subject matter expert at a competing academic medical center currently using an integrated health records and research systems framework.

The discussion consisted of reasons for systems integration, perceived value, and a description of the organizational workflow and systems architecture. From this discussion, it was concluded that a research management system was integrated with EHR to display on study notifications, consent documents, and contact information within health records for patient safety purposes. An interview was also conducted with a representative from Forte Research Centers, Inc., developer of the OnCore system, during which options for integration with EDC were discussed. Results from both interviews concluded that a method for pre-populating data elements from electronic health records into an EDC and research management system had not been achieved.

A literature review was then conducted to gain knowledge and compare methods for systems integration. The information was compiled and compared to the Emory enterprise, research infrastructure, and workflow to propose a best suited solution. This solution and alternatives were documented as a business case to serve as a concise presentation of major findings widely accepted by the informatics community. The formal business case including the project overview, justification, enterprise architecture, scope, and cost analysis are included in the executive summary. The following chapters outline a solution, major findings, and barriers to integration.

Chapter III: Proposed Solution

Introduction

This chapter describes the findings concluded from the methodology and approaches taken to define a solution for systems integration, as well as identify and evaluate alternatives to the proposed solution.

The Systems

PowerChart is the EHR system used by Emory University Hospital and serves as the source for research documentation for the majority of clinical research studies at Emory University. PowerChart provides access to records for all disciplines across the enterprise including satellite sites, emergency care, ambulatory clinics, and surgery. Clinicians are able to document and share visit records, orders, current and past medications, vital signs, patient demographics, and allergies. The EHR also houses laboratory results, radiology scans and reports, and outside scanned documents. In addition to the traditional EHR capabilities, PowerChart displays signed research consents in a patient's records by integrating with the PowerTrials of Management (POM) application.

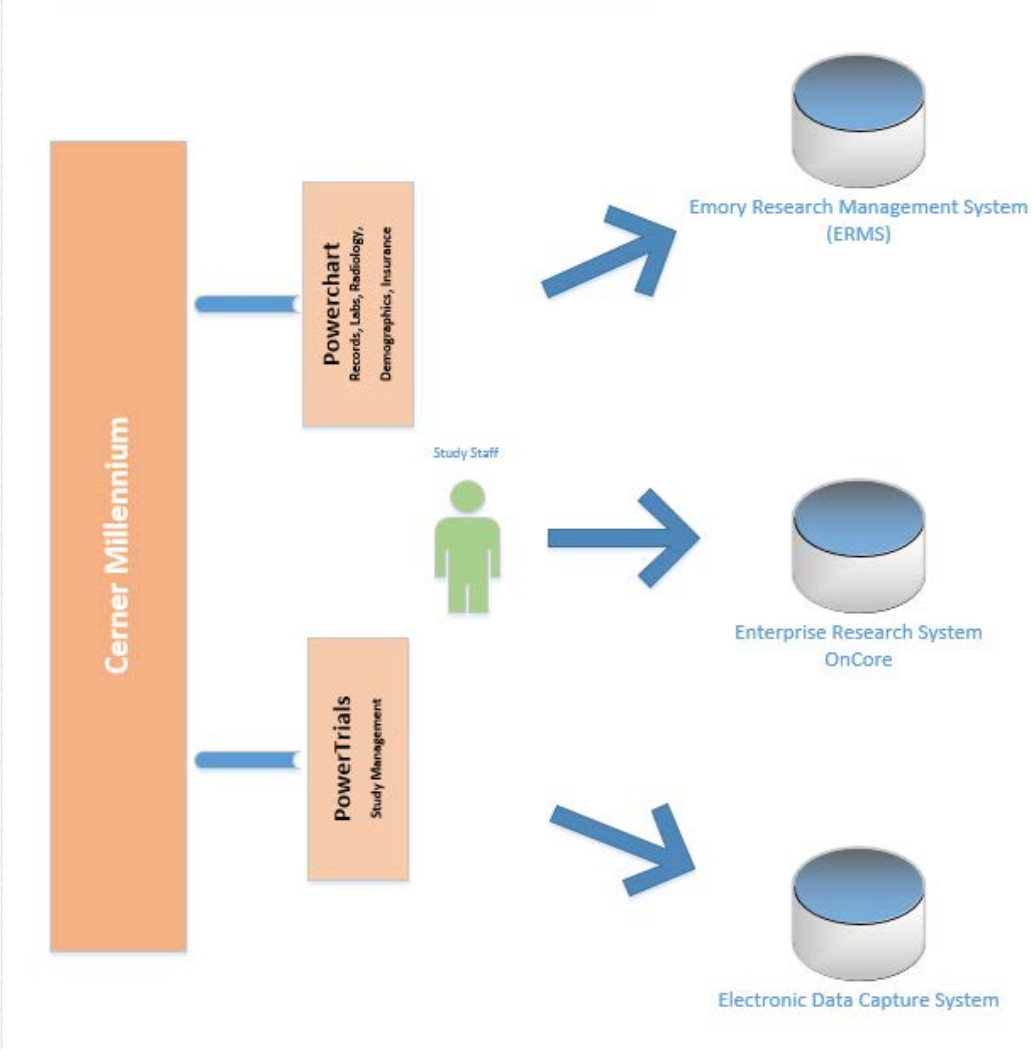
The Emory Research Management System (ERMS) is an internally developed clinical research management system used to register and track clinical research participants for studies that have billable items and services. Participants are registered in this system at the time of study consent. Research staff are required to track study visits per protocol, and update participant status as they progress through the study. ERMS is primarily used for research administration purposes such as aiding in billing compliance and metrics reporting.

OnCore is an enterprise research management system that is currently used solely in the Winship Cancer institute at Emory. OnCore was included in this project due to the large number of clinical studies that are conducted within the cancer institute. As of December 2016, 472 clinical studies were led by Winship investigators making the cancer institute a great contributor to the Emory clinical research sphere (Research, 2016). In its' current state, OnCore is primary used as a tool for tracking patients through study treatment and long term follow up according to calendars generated per protocol schedules. Information regarding investigational products and study orders, as well as study team contact information is stored here. OnCore also serves a reporting tool for study accruals, monitoring and auditing, data safety, protocol deviations, financial management, and quality assurance. The research management system is highly customizable and has the capacity to interface with an EDC module, yet currently these capabilities are not utilized.

Electronic Data Capture systems are a repository for research data collected throughout the study in adherence with the protocol. Specific data elements collected vary between protocols, however much of the data reflects health information related to study treatment. Vital signs, medications and doses, adverse events, and lab results are common elements collected within an EDC. The source for these data are primarily located in the EHR. All data related to a study participant entered into EDC must be de-identified and kept separate from health records. Multiple EDCs are used throughout the organization to store protocol specific data. This variability is due to differences in study sponsors, the size of the trial, funding, and physician discretion.

Figure 1 depicts the current flow of information from electronic health records to multiple research systems. It was determined that the flow of information is a manual, duplicative, and time consuming process requiring data transcription from EHR to multiple systems by research staff.

Figure 1:
Current organizational disconnection between EHR and research systems



The Current State of Integration

The extent of data automation from PowerChart EHR is limited to a clinical data warehouse (CDW) that receives a nightly load of medical, billing, scheduling, lab, and medication, structured and unstructured data from the Emory hospital and clinics. This data is systematically pulled into tables that at request, can be extracted by a data analyst using structured query language (SQL) commands. Although both structured and unstructured data are housed in the CDW, extractable data is restricted to structured information or discreet variables only. Major uses for this data include identifying the number of patients that fit study criteria prior to engaging in research (feasibility studies) and cohort identification. The CDW does not currently integrate with ERMS, OnCore, or a research EDC system.

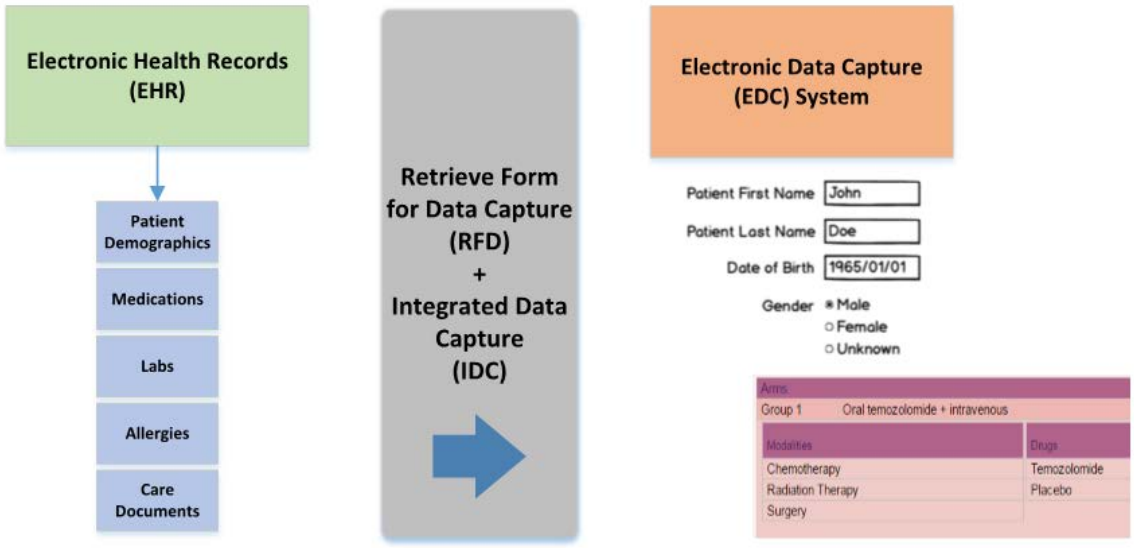
Proposed Solution

Retrieve Form for Data Capture (RFD) is an Integrating the Healthcare Enterprise (IHE) and Clinical Data Interchange Standards Consortium (CDSIC) developed method for gathering data within an application environment to support the pre-population of forms received from an external source. Integrated Data Capture, an extension of RFD, is the process by which data is electronically transmitted from one application to another. Using this method ensures that the process for the pre-population of relevant data meets federal guidelines and regulations regarding the use of electronic records (Marsha Laird-Maddox; Susan B. Mitchell, 2014).

The Retrieve Form for Data Capture process utilizes the continuity of care document (CCD), an XML-based HL7 V3 document generated in the EHR used to exchange clinical information between healthcare providers. The CCD should contain pertinent data elements

needed for electronic data capture. A script code is used to convert the document into a format that is usable by the research system. A new window then displays the eCRF pre-populated with EHR data elements in relevant fields. Within the displayed window, a user can then modify and enter additional research data as needed and save the form within the research system. This process eliminates the need for manual entry of pre-defined data elements needed for research capture (Marsha Laird-Maddox; Susan B. Mitchell, 2014). Figure 2 depicts the Retrieve Form for Data Capture and Integrated Data Capture workflow.

Figure 2: RFD and IDC Workflow

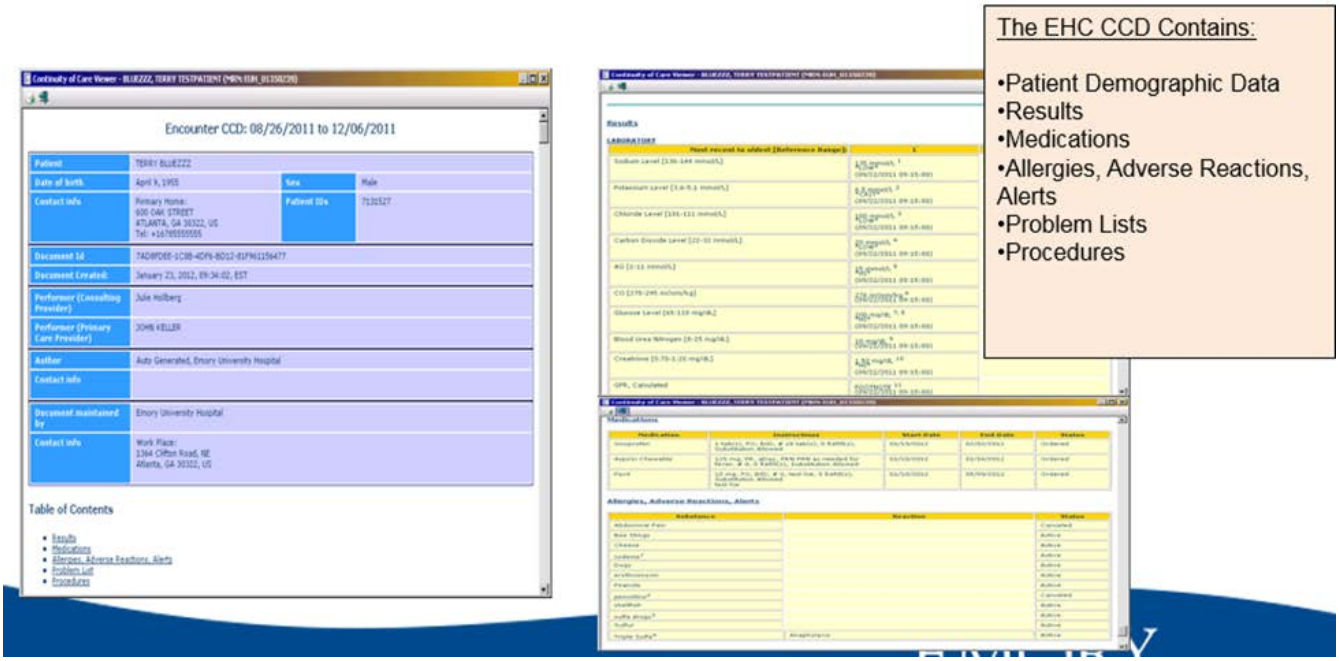


The Emory Healthcare CCD contains the following basic elements:

- patient demographic data
- conditions/problems
- procedures
- medications
- allergies/adverse reactions
- immunizations
- lab results
- microbiology
- vital signs
- visits
- problems list
- procedures

Additional elements include: Diagnosis, Radiology, History and Physicals, Discharge Summary, Provider Reports, and Anatomic Pathology (Dee Cantrell RN, n.d.).

Figure 3: Snapshot of the EUH CCD using a test patient record

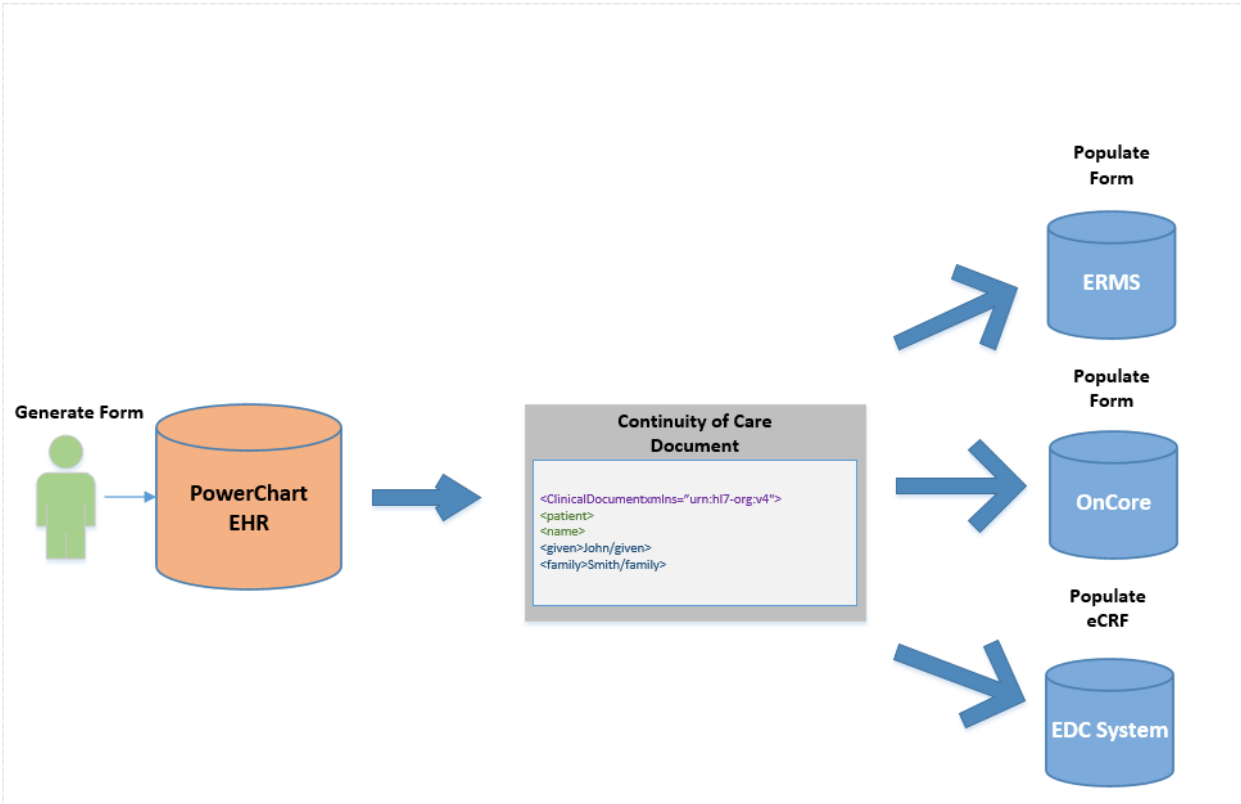


There are several types of notes found within the CCD that pertain to clinical care as patients frequently see multiple providers, often for the same condition or related malignancies. Consequently, there may be inconsistencies within the document related to care for the same patient. For example, patients diagnosed with cancer often undergo several tests for disease staging; such as tumor biopsies and radiology scans. Therefore, there are multiple sources within the CCD that may provide may provide relevant research data elements (i.e. conditions/problems, problems list, radiology, physician notes, and anatomic pathology). Prior to initiating the RFD process, the source of truth related to diagnosis will need to be identified to ensure the most accurate data is captured in the EDC.

The Retrieve Form for Data process can also be applied to the Emory Research Management and OnCore systems. However, the number of relevant data elements pre-populated into these systems are expected to be significantly less than those for EDC as there is less need to house clinical information. Visit dates and participant status correlated with clinical visits can be pre-populated if documented and specified in the continuity of care document. Using the Retrieve Form for Data Capture method to integrate EHR and research systems ensures that there is one source of truth and identical data across systems while reducing the need for manual data transcription for study teams.

The following figure illustrates the data flow with the implementation of the RFD process beginning with the generation of a document by a healthcare provider or staff.

Figure 4: Proposed Data Flow Using RFD



Additional Requirements

Although the Retrieve Form for Data Capture method meets federal compliance when sharing information among electronic health record systems, there are considerable adjustments needed to initiate integration between EHR and clinical research systems. Utilizing closed systems, such as electronic data capture and research management systems, specific procedures and controls must be employed to protect the authenticity, integrity, and confidentiality of the records including the use of digital signatures. Actions must also be taken to ensure that signer cannot deny the validity of the signed record. Federal regulation 21 CFR part 11 outlines the system requirements needed to safeguard electronic records. The following procedures and controls should be established prior to initiating the proposed solution using the Retrieve Form for Data Capture method (U. S. F. a. D. Administration, 2016).

- validation system workflow to ensure accuracy, reliability, consistent intended performance, and the ability to discern invalid or altered records
- the ability to generate accurate and complete copies of records in both human readable and electronic format
- protection of records to enable their accurate and ready retrieval throughout the records retention period
- ability to limit system access to authorized individuals
- the use of secure, computer-generated, time-stamped audit trails to independently record the date and time of entries and actions that create, modify, or delete electronic records
- changes made to the records should not hide previously recorded information and audit trail documentation should be kept as long as that required for the subject electronic records
- the use of operational system checks to enforce permitted sequencing of steps and events
- the use of authority checks to ensure that only authorized individuals can use the system and perform activities assigned for their access level
- the use of device checks to determine the validity of the source of data input or operational instruction

- determination that persons who develop, maintain, or use the electronic record have the education, training, and experience to perform their assigned tasks
- the establishment of, and adherence to, written policies that hold individuals accountable and responsible for actions initiated under their electronic signatures
- the use of appropriate controls over systems documentation including:
    - adequate controls over the distribution of, access to, and use of documentation for system operation and maintenance
    - revision and change control procedures to maintain an audit trail that documents time-sequenced development and modification of systems documentation
- signed electronic records should also contain information that indicates
    - the printed name of the signer
    - the date and time the signature was executed
    - the action associated with the signature such as review, approval, responsibility, or authorship
    - electronic signatures executed in electronic records should be linked to their respective electronic records to ensure that signatures cannot be falsified
- electronic signatures should have the following controls and components
    - be unique to one individual and their identity verified prior to establishing, assigning, and certifying an electronic signature
    - certify electronic signatures used in the systems are legally binding equivalent to traditional handwritten signatures
    - two distinct identification components should be employed for each electronic signature
    - when executing a series of signings during a single, continuous period of controlled system access, the first signing should be executed using all electronic signature components; subsequent signings can be executed using at least one electronic signature component that is designed to only be executable by the individual
    - when an individual executes one or more signings not performed during a single, continuous period of controlled system access, each signing shall be executed using all of the electronic signature components
    - administered and executed to ensure that attempted use of an individual's electronic signature by anyone other than its owner requires collaboration of two or more individuals
- required controls for passwords involve
    - maintaining uniqueness such that no two individuals have the same passwords
    - ensuring passwords nuances are periodically changed
    - use of safeguards to prevent unauthorized use of passwords, and methods to detect and report attempts of their unauthorized use

Alternatives

Several methods for systems integration were considered for this project. The alternatives to the proposed solution, including their advantages and disadvantages are outlined below.

Research Data Warehouse

A research data warehouse (RDW) traditionally aggregates routinely de-identified patient information from EHR, research systems, registries, and biorepositories for feasibility studies and identification of potential participants prior to initiating a study. This information can be abstracted by a data analyst provided a request for information from investigators. Access to information can also be achieved by creating or purchasing additional software applications that are more user friendly, allowing investigators to retrieve data of interest (Medicine, 2017).

To meet integration of EHR and research system's needs, an RDW can be developed to include an application that supports data relationship management supporting feasibility studies, study recruitment by identifying participants who meet specific eligibility criteria, and individual patient level data once a participant provides consent. Individual patient information is initially based on a pseudonymized record and remains invisible to study teams until a signed informed consent document is obtained. The security architecture for an integrated RDW that supports clinical data at an individual patient level will vary across organizations. However, secured messaging, profiles for end-user authentication, and mechanisms for specifying delegation credentials, authorization decisions and control policies are basic requirements (De Moor et al., 2015).

Advantages:

- Incorporates all research systems across the enterprise to aggregate and store data
- Potential to utilize current research systems

Disadvantages:

- Requires data analyst to pull discreet information or the purchase of user friendly software
- Requires significant changes to the information and technology architecture including the acquisition of new applications to support data relationship management
- Requires a nightly load to update the database
- Does not auto-populate research databases with relevant data

Web Services

Web services allow for the exchange of data between applications and software systems using open internet protocols and standardized XML messaging. This allows different software systems that may have different program languages to communicate information in terms of service requests and service providers. Software is made available by exposing a set of operations that can be performed with a task. Each task exposes a set of variables that can accept data passed into the set of operations, allowing for a data request ("Getting a Look at Web Services ").

There are several measures employed to secure web services using the enterprise application integration (EAI) domain within an intranet and across a corporate firewall. Authentication requires the service requestor and service provider to enter credentials and verify identity. The service requestor's credentials are confirmed using authorization. Web services also utilize nonrepudiation ensuring that the message sender is the same as the creator. However, a major security concern for the utilization of web services for the EAI domain is the rare use of data encryption and only one required level of authentication. There is a need for more sophisticated security as information is moved from application to application as XML documents increasing vulnerability to unauthorized access and malicious use (Samtani, 2002).

Advantages:

- Availability over the internet
- Built on top of open standards
- Provides application to application interaction
- Can use software written in various program languages and running on various platforms

Disadvantages:

- Does not auto-populate research databases with relevant data
- Insufficient security or authentication requirements for the exchange of PHI
- High vulnerability to breached information
- Requires the acquisition of additional software

Continue Current Infrastructure

The current infrastructure encompasses disparate research systems that require manual transcription from EHR data. The clinical data warehouse is used to conduct feasibility analysis for proposed research as well as cohort identification.

Advantages:

- Does not require additional resources
- Organizational familiarity with the research processes

Disadvantages:

- Does not auto-populate research databases with relevant data
- Requires extensive training for incoming staff
- Lack of consistent information across systems
- Decreases clinical productivity/workflow

Evaluation Matrix

The proposed solution and each alternative were rated on a scale from 1 to 5, with 1 being the worst and 5 being the best. The scores were totaled to indicate which solution best aligned with identified organizational infrastructure, in additional to organizational goals and strategies, and federal initiatives. Specific organizational goals, strategies, and federal initiatives are outlined in the business case referenced in the executive summary.

Criteria	Retrieve Form for Data Capture	Research Data Warehouse	Web Services	Current Infrastructure
Interoperability	5	3	4	1
Deployable with Current Infrastructure	3	2	3	5
Data Auto-population	5	1	1	1
Reduces Transcription Errors	5	1	2	1
Improves Workflow	4	3	3	1
Increases Productivity	4	3	2	1

Increases Efficiency	5	3	3	1
Decreases Long Term Research Costs	4	3	3	2
Initial Development Costs	1	1	1	5
Information Security & Privacy	5	5	2	5
BA Strategic Alignment	4	3	3	4
IA Strategic Alignment	4	4	2	5
TA Strategic Alignment	3	3	2	5
Total Score	**52**	**35**	**31**	**31**

Based on the evaluation matrix, the Retrieve Form for Data Capture method and a research data warehouse received the highest scores as options for systems integration. However, to achieve interoperability and the auto-population of data from EHR to research databases, the Retrieve Form for Data Capture method is the most suitable option. The foundation to deploy this solution is established within the organization, but will require major modifications and additions to currently utilized research systems to meet federal requirements for electronic records.

Chapter IV Discussion

Introduction

This chapter intends to summarize the project, discuss implications for public health and academic medical centers conducting clinical research, describe limitations and barriers of the proposed solution, and identify areas of improvements and future recommendations.

Summary

Clinical research is the gateway to developing and testing health treatments and interventions that impact public health as well as monitor long term health outcomes related to these treatments. As clinical research studies become more complex, study timelines and the cost of conducting research continue to rise. Related to increasing costs are extensive data clean up periods post study completion. Disparate research systems perpetuate this problem by requiring manual transcription and data duplication across multiple research systems. Several methods for systems integration have been developed, however options for integrating electronic health records with clinical research systems are limited. Creating a method for pre-populating systems with pertinent research data is an even greater challenge.

The Retrieve Form for Data Capture process including the Integrated Data Capture workflow is an integration approach developed by the IHE and CDSIC to achieve automated electronic transmission of relevant data elements from an EHR to electronic data capture systems. This method ensures that the transmission of data is federally compliant, meeting appropriate security measures to ensure health information is properly protected during transmission. However, to meet federal requirements for electronic records, major

modifications are needed to update the research systems. Evaluating alternatives to the Retrieve Form for Data Capture method such as the creation of a research data warehouse, establishing web services between organizational systems, and the continuation of disparate systems proved to be inadequate solutions for improving data quality and decreasing the need for manual data transcription.

Implications

Integrating electronic data capture and research management systems has the potential to improve the overall conduct of clinical research from study initiation to data analysis by shortening research timelines and improving the reliability of results. In doing so, public health benefits from the systematic testing of new and existing health interventions more quickly, thus providing the evidence needed for decision making and health policy implementation affecting population health. By integrating systems, long term costs associated with extended research timelines could be reduced, in turn decreasing the amount of federal spending for sponsored research.

Within the Emory enterprise, an academic medical center, although a heavy monetary and labor resource investment is initially needed, long term costs, benefits, and added value are expected to outweigh startup expenditures. Shortening research timelines, especially for investigator initiated studies, reduces the risk of losing funding prior to reaching study endpoints. As incoming funding plays a key role in conducting research within the organization and realizing the Emory mission, it is imperative to ensure adequate support is provided to keep research initiatives ongoing within the enterprise while remaining an innovative competitor in

the research stratosphere. Findings from this project also serve as an example of a potential integration architecture to other academic centers conducting clinical research within multiple systems.

Barriers, Limitations, and Recommendations

Although we have identified a solution to integrating EHR and research systems in efforts to reduce data duplication, transcription errors, and improve the research workflow within an academic medical center, there are several barriers to implementing the proposed solution within our organization and throughout the clinical research community. Major impediments to deploying the solution are the cost and time related to ensuring the integration of EHR and research databases is federally compliant. As these systems were initially developed to support research administrative activities and operations, they were not designed to interface with electronic health records, therefore currently lack the procedures and controls to maintain the confidentiality, integrity, and authenticity of the records as required by federal law. Updates needed to ensure the systems meet the requirements as outlined in the federal regulation 21 CFR part 11 require considerable monetary and time resources.

Implementation of the Retrieve Form for Data Capture method also has a significant impact on the workload for the organization's information systems staff, requiring development of script code to convert relevant data elements of the continuity of care document to a usable format for the receiving research systems. The specific data elements needed will vary by study. Protocol amendments may include the addition of study procedures, thus requiring the capture of additional data elements and modifications to script code. This is a

heavy time investment for systems architects and developers, not only for initial deployment, but for ongoing upgrades and maintenance. Deploying the solution may result in a workload shift from research study teams to information systems staff which may decrease study timelines from only one perspective. This may not equate to achieving complete organizational efficiency.

The extent of data quality improvement can also be argued as a limitation to systems integration. Utilizing the continuity of care document assumes that the electronic health record is an accurate source of truth. One inaccurate data element captured within the CCD has the potential to populate all research systems with false information. Although errors found within research records can be overridden once populated into a system, these mistakes must be found and manually corrected for systems to reflect identical information. If errors are not identified and corrected across systems, invalid data may perpetuate decreasing data integrity across systems.

Lastly, there is a need to identify one electronic data capture system to integrate with research management systems. There are many solutions available for electronic data capture and several are used throughout the organization. However, to ensure this project is successful, a commitment to one EDC compatible with an EHR interface should be chosen. This requires that the organization assess the usage of EDCs across the enterprise and evaluate which is best suited for integration purposes. Once chosen, the benefits of this project may be largest for investigator initiated studies, which comprise nearly 19% of Emory research studies, as industry studies typically mandate use of one contracted electronic data capture system.

Conclusion

Beyond development of a systems architecture using the Retrieve Form for Data Capture method, this project identified barriers to integrating EHR and clinical research systems exceeding challenges related to architecture. Although the Retrieve Form for Data Capture method aims to enhance the workflow and increase productivity for research and clinical staff, a large workload is passed to systems architects and developers as extensive code and validation workflows are needed to implement the solution. Time and cost for creating research systems that meet requirements and standards needed to share protected health information are major hindrances to integration, perpetuated by the development of siloed solutions to meet short term needs. Developing an integrated systems solution requires organizational commitment to enhancing the research workflow, reducing study timelines, and lowering long term costs despite the initial investment. More awareness surrounding the public health benefits of integrating EHRs and research systems as well as requirements for data sharing are needed.

Chapter V Executive Summary

Introduction

This chapter provides an overview of the thesis project presented as a formal business case for Emory University, an academic medical center which served as a model for project evaluation.

Emory University EHR and Clinical Research Data Integration Project: A Business Case

As one of the country's leading academic medical centers that conducts thousands of clinical research studies a year, the need to remain innovative and develop a streamlined, systems driven process for research conduct across the enterprise remains a need for the institution. Improving data integrity and decreasing research timelines in partnership with federal and private sponsors when conducting clinical research is crucial in maintaining relationships among research partners, establishing incoming funding for the organization, remaining a prestigious and competitive research institution, as well as contributing to the advancement of monitoring and improving population health.

Problem:

Increasing clinical research timelines consequently delay the evaluation of new treatments and therapies, limit the ability to accurately study long term health effects and accurately monitor disease prevalence, and postpone the creation of standards, policies, and programs aimed at managing population health and bettering outcomes. The lack of integrated research systems contributes to impairing the delivery of timely and accurate study information needed to discover new and better health solutions for emerging and existing conditions that affect population health.

Solution:

An enterprise wide integration of EHR data and clinical research systems that allows for the integration of data from electronic health records (EHR) to clinical research databases can help alleviate many issues that delay progression of discovering newer and better treatments, while furthering the reach of public health and decreasing long term costs for the institution. Such an architecture may also serve as an innovative guide for other academic medical organizations by providing a framework that solves many of the barriers to conducting more effective clinical trials.

Objectives:

This project aligns with the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) goal to support and encourage the use of seamless data exchange, including the re-use of information from health care systems to clinical research systems allowing for the one time, point of care, entry of data (Administration, 2014). The project also aligns with the President's Management Agenda (PMA) to increase the economic impact of Federally-funded research (Kalil, 2016).

- Assess whether the automatic transfer of data from EHRs to a clinical trials management system (CTMS) and EDC system is a feasible solution for improving data quality and reducing timelines for clinical research studies.

- Evaluate whether integrating these systems improves organizational workflow by reducing or eliminating the need for duplicate data entry.
- Describe barriers to systems integration and present an infrastructure to overcoming these barriers.

1 INTRODUCTION

1.1 PURPOSE OF BUISNESS CASE

This business case is intended to assist stakeholders in decision making regarding the viability of integrating EHR and clinical research databases within an academic medical center. The intended audience of the EHR and Clinical Research Data Integration business case is organizational senior leadership.

2 PROJECT AND PRODUCT OVERVIEW

The EHR and Clinical Research Data Integration Project aims to identify a method for improving data quality within research databases, maximize organizational workflow by reducing the need for data duplication by practitioners and research staff, and reduce research timelines and costs by decreasing the need for extensive data cleanup. The project will serve as a model for linking electronic health records (EHRs) with pertinent clinical research data that ultimately allows for cleaner and timelier information that can be analyzed more quickly by study teams and sponsors, while providing more accurate, real time administrative reporting. This project will take place within Emory University, an academic medical center, which conducts thousands of clinical trials annually across multiple locations. Initial cost for project is estimated at $1.79 million.

3 JUSTIFICATION

3.1 BUSINESS NEED

Billions of dollars are spent annually on the development and conduct of clinical research in the United States alone. Clinical trials, especially phase III trials, have gotten larger and become increasingly more complex with the average length of one trial increasing by 70% and the clinical research staff work burden by 67% from 1995 to 2005. This impact is thought to have grown even greater in recent years (Sampat, 2017).The exponential growth in the complexity of conducting clinical research has resulted in the development of niche research systems used to meet a variety of clinical and administrative research needs. These siloed systems often hinder organizational workflow by requiring transcription of data across multiple databases, threatening data quality, and extending research timelines by requiring prolonged data cleanup periods. Such a duplicative data entry process also detracts from clinical duties and productivity. A solution to addressing these issues is needed.

3.2 PUBLIC HEALTH AND BUSINESS IMPACT

The conduct of clinical research leads to innovative health solutions that are critical for improving health outcomes and providing scientific evidence needed to support and guide clinical and health policy decisions. Through the thousands of clinical studies that are conducted globally each year, scientists can contribute medical knowledge, develop medicines and devices, vaccinations, and lifestyle recommendations related to the treatment, diagnosis, and prevention of diseases and conditions that affect population health. Clinical research also provides imperative insight to the efficacy, safety, side effects, and long term outcomes from receiving a health treatment or intervention (Health, 2015).

Researchers have identified several causes to increasing clinical research timelines that consequently delay the evaluation of new treatments and therapies, hinder the ability to accurately study long term health effects and accurately monitor disease prevalence, and postpone the creation of standards, policies, and programs aimed at managing population health and bettering outcomes. Major contributors include extensive periods of data collection, monitoring, and data cleanup which also create loss of productivity within a research organization.

3.3 STRATEGIC ALIGNMENT

Goal	Project Response Rank	Comments
Scale: **H** – High, **M**- Medium, **L** – Low, **N/A** – Not Applicable		
Emory University Research Mission:		
Conduct innovative and collaborative research and integrate this knowledge into the practice of medicine	L	
Advance the early detection, treatment, and prevention of disease	M	
Emory University Office for Clinical Research Mission:		
Facilitate operational processes that support the efforts of the clinical research team in the timely initiation, management and completion of clinical trials at Emory	H	
Food and Drug Administration (FDA) Center for Drug Evaluation and Research Strategic Initiative:		
Support and encourage the use of seamless data exchange	H	
FDA Strategic Initiative:		
Advocate the need for capturing source data, including data originating in health care systems	H	
Department of Health and Human Services (DHHS) Strategic Goals:		

Goal	Project Response Rank	Comments
Goal 2: Advance scientific knowledge and innovation	H	
Goal 3: Advance the health, safety, and well-being of the American people	M	
President's Management Agenda (PMA) Strategic Goals:		
Increase the economic impact of Federally- funded research and development by accelerating and improving the transfer of new technologies from the laboratory to the commercial marketplace	H	

4 SCOPE

4.1 OBJECTIVES

The overall goal is to integrate EHR with clinical research databases allowing relevant study information to be auto-populated into appropriate electronic data capture systems. Having information flow from the EHR will reduce the need for data duplication, resulting in decreased data cleanup periods and overall study timelines, as well as reduce the administrative burden for research staff.

The objectives of systems integration are as follows:

- Decrease the error rate within clinical research databases
- Decrease data cleanup periods post study completion
- Reduce overall study timelines
- Improve long term follow up data quality for research studies
- Accelerate the lab to market process for new and improved health treatments

4.2 HIGH-LEVEL REQUIREMENTS

The following table presents the requirements that the project's product or result must meet in order for the project objectives to be satisfied.

Req. #	Requirement Description
1	Ability to access EHR data for pariticpants in clinical research studies
2	Ability to identify specified data elements for clinical research studies
3	Ability to de-identify protected health information for research participants
4	Design an infrastructure to populate EDC system with research data extracted from EHR
5	Remain 21 CFR part 11 compliant

Risks

Risks	Mitigation
Funding may be diverted to other projects	Seek organizational support and leadership buy in by proposing cost analysis and demonstrating added value
May pose risk to patient privacy	Ensure appropriate security measures are in place and patient has consented to usage of de-identified PHI

4.3 MAJOR DELIVERABLES

The following table presents the major deliverables that the project's product, service or result must meet in order for the project objectives to be satisfied.

Major Deliverable	Deliverable Description
Approval and funding	Obtain approval and funding to initiate the project
Data Elements	Define what data elements will be most useful for the Project
Data Infrastructure	Develop strategies for interoperability between information systems to support the transfer and population of data

4.4 BOUNDARIES

The project applies to the EHR and clinical research systems utilized at Emory University. The enterprise wide infrastructure proposed may vary among other academic medical centers that utilize different vendors and systems. There is also variation in EDC systems within the organization that study teams use for data collection. Data collected from EHR for research purposes will be limited to what individual participants allow in a signed informed consent document and HIPAA release form. De-identified participant data collected for future analysis will be required to remain separate from the EHR. The automated transmission of data will be unilateral, and will pre-populate from the EHR to clinical research systems only.

Cost for this project is an estimation based on a systems development calculator. Actual cost for completion of this project may differ from the estimated cost as requirements may change prior to completion.

5 ENTERPRISE ARCHITECT IMPACT

5.1 BUSINESS NEED AND SOLUTION

Billions of dollars are spent each year on the development and conduct of clinical research in the United States alone. Clinical trials, especially phase III trials have gotten larger and become increasingly more complex with the average length of one trial increasing by 70% and the clinical research staff work burden by 67% from 1995 to 2005. This impact is thought to have grown even greater in recent years (Sampat, 2017). The exponential growth in the complexity of conducting clinical research drives the need for a system that maximizes the use of staff time and existing data.

This project will propose a solution that allows for a reduction in data duplication and errors within clinical research systems and databases. Studies have reported that error rates between EHR and research databases as high as 27% (Goldberg et al., 2008). An integration that bridges EHR with clinical research allows for an innovative solution for academic medical centers to maximize productivity as well as deliver cleaner more accurate clinical research data. More reliable data has the potential to decrease research costs by reducing study timelines and time to market for health treatments by decreasing data cleanup periods post study completion. Integration also has the potential to improve long term follow up data quality if adherence to follow up study visits declines.

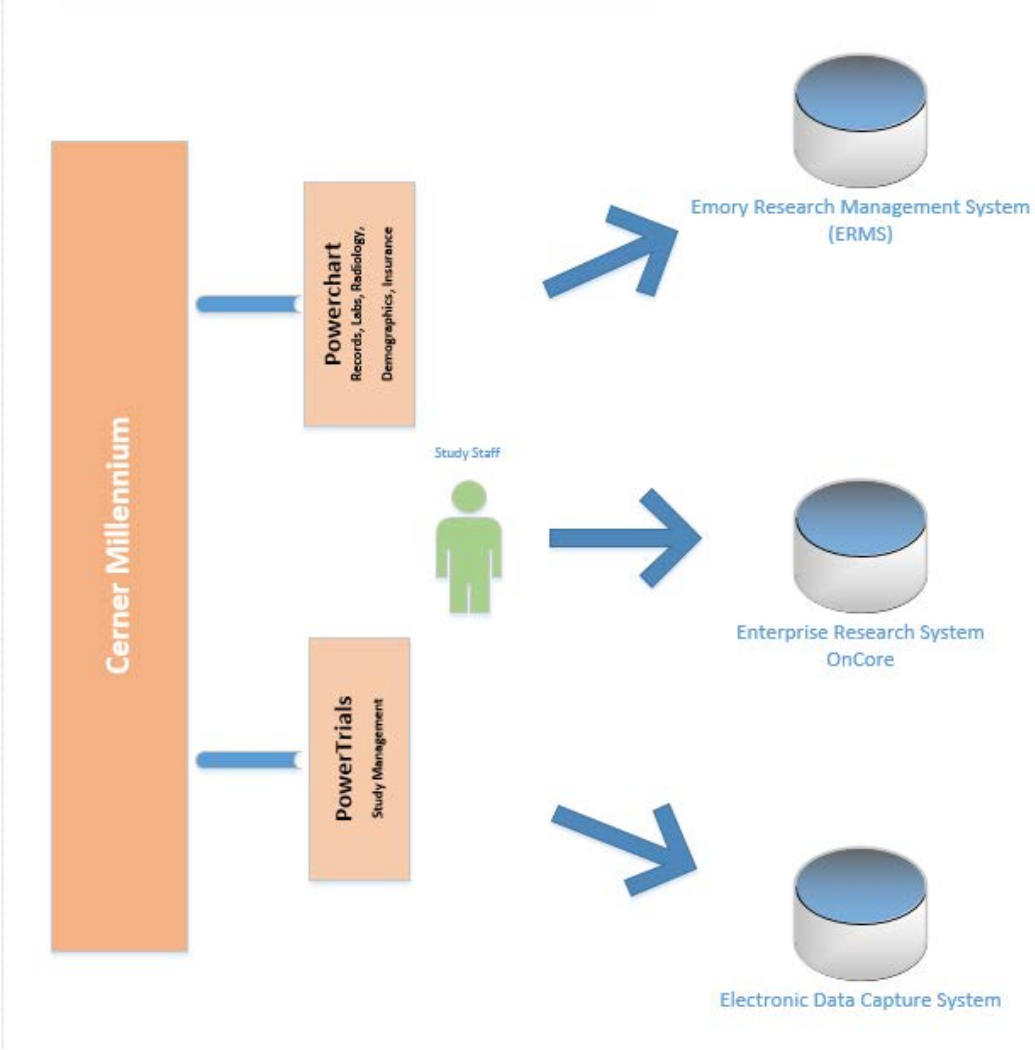
This project will contribute to organizational missions to conduct innovative and collaborative research and integrate this knowledge into the practice of medicine, and facilitate operational processes that support the efforts of the clinical research team in the timely initiation, management and completion of clinical trials at Emory. Through these missions, Emory University will have the ability to directly contribute to national initiatives to:

- Support the use of seamless data exchange
- Advocate the need for capturing source data, including data originating in health care systems
- Advance the health, safety, and well-being of the American people
- Increase the economic impact of Federally- funded research and development by accelerating and improving the transfer of new technologies from the laboratory to the commercial marketplace

Currently, information for patients who consent to research studies is entered into four major information systems; the hospital EHR which serves as the primary source of information, two research management systems, and an EDC system. Once information is dictated, scanned, or uploaded into the EHR, data is manually transcribed from the EHR into research systems that serve various purposes across the organization, including administrative data collection to facilitate operations, participant registration and tracking, and study data collection. Transcribing information across multiple systems

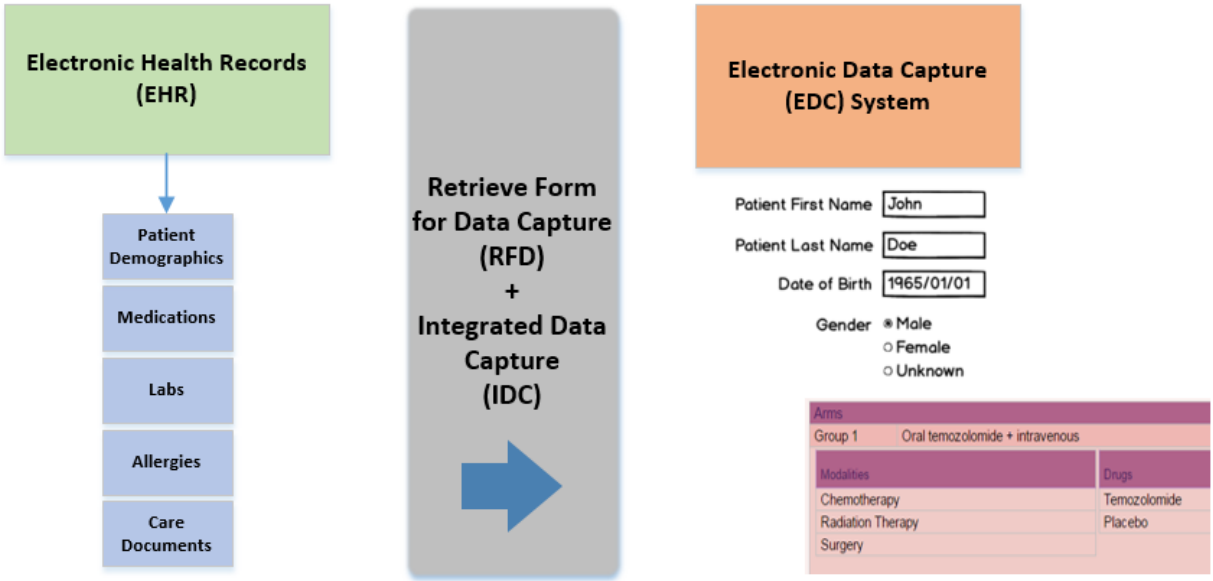
increases the risks for data discrepancy among systems leading to poor data quality, and interrupts the workflow for clinical and research staff (see Figure 1).

Figure 1:
Current organizational disconnection between EHR and research systems

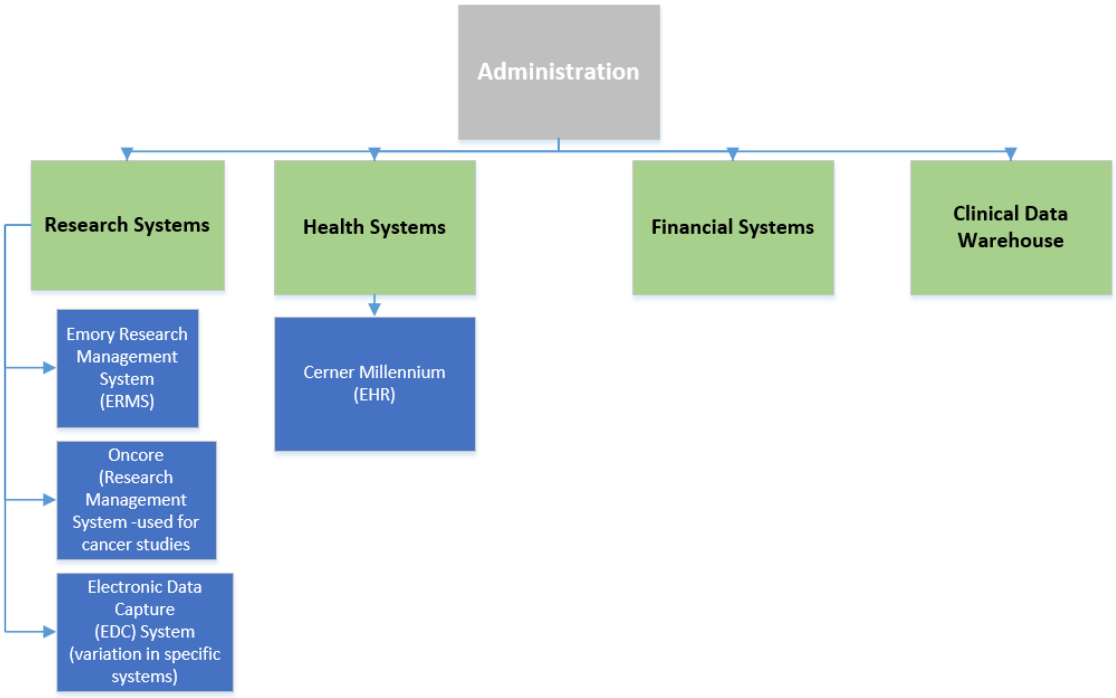


The proposed solution for this project is to integrate EHR and clinical research management systems and an EDC system through Retrieve Form for Data Capture (RFD) and Integrated Data Capture (IDC) to enable defined data elements to be auto-populated into relevant systems.

Retrieve Form for Data Capture (RFD) is an Integrating the Healthcare Enterprise (IHE) and Clinical Data Interchange Standards Consortium (CDSIC) method used for collecting data within a user's current application to support the pre-population of forms retrieved from an external source. Integrated Data Capture, an extension of RFD workflow, is the process that enables electronic transmission of data to research systems. By generating a continuity of care document from the EHR containing relevant data elements needed by a study, a script code transforms the document into a format that can be used by the research system by pre-populating the appropriate fields (Marsha Laird-Maddox; Susan B. Mitchell, 2014).



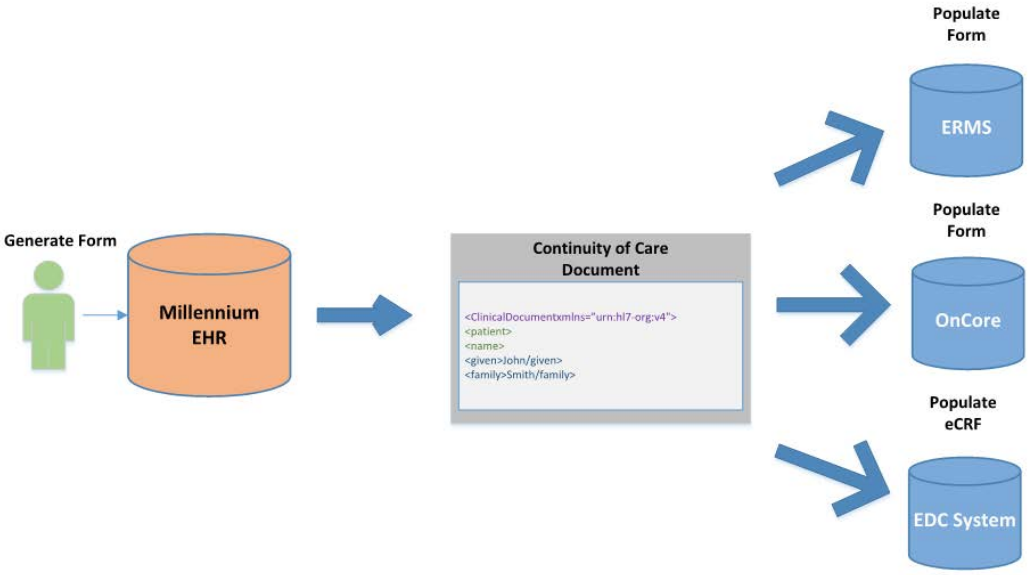
5.2 EMORY UNIVERSITY ORGANIZATIONAL ARCHITECTURE



The Emory University systems affected by implementation of the solution presented in EHR and Clinical Research Data Integration Project are as follows:

- **Health Information Systems:** The EHR will serve as the main source of data from which information will be pre-populated.
- **Research Systems:**
    - EDC System: Selection of one system will be required to be used for data capture of relevant data elements. This is a system that will receive data from the EHR.
    - ERMS: This system is used for research participant registration and tracking, as well as for adminstrative reporting. This is a system that will receive data from the EHR.
    - OnCore: This system is used for patient registration,the collection treatment information, adverse events, and long term follow up tracking.

Proposed Information Flow



5.3 ENTERPRISE ARCHITECTURE

The components affected by the implementation of the solution presented in the project are:

Business Architecture

- Integrating systems to improve organizational workflow and data quality across systems
- Enhancing the business, clinical, and research processes for administrative, clinical, and research staff
- Collaborating among administrative, research, and IT staff to develop the final product

Information Architecture

- Data: Pre-defined data elements will be identified to pre-populate research systems

- Integration: Data from within the EHR will be integrated into the research systems by various actors and transactions
- Applications: Cerner applications will continue to be utilized

Technology Architecture (Enterprise, 2011)

- Selecting options for the Integration Profile
- Complying with ITI TF-2x: Appendix V: Web Services for IHE Transactions and 21 CFR part 11 (U. S. F. a. D. Administration, 2016)

5.4 BUSINESS PROCESSES

Process	Description
Receive data from EHR	The research systems will need to be able to receive data from within the organization
Send data from EHR	EHR will need to be able to send data to research systems
Procurement	Procurement of additional hardware and software as needed
Use agreement for one EDC system	The organization will need to decide on one central EDC system
Implementation date	The organization will need to decide on an implementation date and for which departments
Patient Consent	Consents for research participants will need to be reviewed and revised as necessary to cover data usage terms
Data Elements	The data elements for pre-populating research systems will need to be decided per individual protocol

5.5 INFORMATION ARCHITECTURE (IA)

The following services will be impacted by the solution:

- **System and Network Management:** The network will be configured to include interfacing with a new source.
- **Security:** As the research systems will contain PHI, the data will need to be encrypted at rest and in motion.
- **User Interface:** A user new interface will be needed to support queried forms.
- **Data Interchange:** Data from EHR will need to be pushed into three research systems. These systems will need to be able to accept information.

5.6 TECHINCAL ARCHITECTURE

The following categories will be affected by the implementation of this project:

- **Service Requirements:**
  - Compliance: Regulations in effect to protect patient privacy regarding the use of electronic health records.
  - Authentication/Single Sign On: Access to EHR and research systems will be restricted to certain personnel and will require a sign on into each system authenticated by a user.
- **Component Framework:**
  - Digital Signatures: Regulations regarding digital signatures including EHR and research records will need to be followed.
  - Validation workflows: Regulations requiring procedures to ensure accuracy, reliability, consistent intended performance, and the ability to discern invalid or altered records
  - Data Interchange: Data exchange will be impacted by the exchange of information across systems.
  - Database Connectivity: Relevant data within the EHR will need to be connected.
- **Service Interface and Integration:**
  - Enterprise Application Integration: As the research systems will no longer stand alone, the enterprise will be affected.
- **Interoperability:**
  - Data Format/Classification: RFD is capable of leveraging industry standards. HL7 is the messaging format (Enterprise, 2011).
  - Data transformation: Data coming from the EHR may need to be transformed to an acceptable research format.

5.7 REQUIRED POSITIONS AND SKILLS

ROLES:

This project will require architects for enterprise, database, and technology architecture. An IT manager as well as a programmer, developer, and business analyst will also be required.

Experienced research staff will be needed to identify relevant data elements to be pre-populated into research systems. Research administrative staff will also be needed to identify relevant data elements.

SKILLS:

- Project management and change management skills will be needed to ensure this project is successful.
- Familiarity in information sharing standards, software engineering, security, and data management is needed.
- Legal expertise involving data protection laws, contracts, and HIPAA and HITECH standards is needed.

5.8 RISK MANAGEMENT

Risks have been identified and categorized.

Probability and Impact: Probability should be measured as the likelihood that the risk will occur. Impact should be measured in terms of the impact made on the organizational enterprise should the risk occur.

- **Probability Levels:** Certain, Expected, Likely, Possible, Unlikely
- **Impact Levels:** Very High, High, Medium, Low, Very Low

Mitigation Strategy: Identify what actions can be taken in order to reduce the probability of the risk.

Contingency Plan: Identify what actions will be taken when the risk materializes.

Organizational Risks

1. **Lack of Interest**

Description: Leadership may find other projects as priorities	Problem: Possible	Impact: Very High
	Mitigation Strategy: Explain the added value of integration and present data regarding long term costs and benefits	
	Contingency Plan: Reduce project scope and offer an alternative	

2. **Lack of Cooperation**

Description: Research teams may not provide collaboration and provide necessary data elements	Problem: Unlikely	Impact: High
	Mitigation Strategy: Explain the added value of integration as time saved and a reduction in effort	
	Contingency Plan: Have research committees provide required input	

3. **Lack of Funding**

Description: Adequate resources may not be available to cover the project costs	Problem: Likely	Impact: Very High
	Mitigation Strategy: Describe project importance and added value presented as return on investment	
	Contingency Plan: Reduce project scope and offer an alternative	

Project Risks

1. **Incorrect Form Generation (Enterprise, 2011)**

Description: User retrieves the incorrect form	Problem: Certain	Impact: Very Low
	Mitigation Strategy: Create a user-friendly interface and provide training	
	Contingency Plan: Users will have the ability to discard forms	

2. Missing Required Information

Description: User does not enter all required data fields	Problem: Certain	Impact: Very Low
	Mitigation Strategy: Create a user-friendly interface and provide training	
	Contingency Plan: Form validations will prevent the submission of forms with missing data	

3. Incorrect Information Auto-populated

Description: User notices incorrect data element auto-populated from source	Problem: Certain	Impact: Very Low
	Mitigation Strategy: Identification of data elements prior to implementation	
	Contingency Plan: User can override data entered into research databases	

4. Security Breach of Protected Health Information (PHI)

Description: PHI is accessed and/or distributed by unauthorized users	Problem: Possible	Impact: Very High
	Mitigation Strategy: Enable access controls and security	
	Contingency Plan: Utilize organization breach response plan	

5. Network Failure

Description: Instance where the network becomes unavailable	Problem: Expected	Impact: Very High
	Mitigation Strategy: Deploy a network monitoring tool to provide warning of unusual activity on the network and identifies the source	
	Contingency Plan: Implement organization network failure response plan	

6. Hardware Failure

Description: Instance where the hardware experiences failure	Problem: Expected	Impact: Very High
	Mitigation Strategy: Ensure current hardware is in working order	
	Contingency plan: Implement organization hardware failure policy	

7. Incorrect System Acting as Form Filler (Enterprise, 2011)

Description: The incorrect system populates a form	Problem: Unlikely	Impact: High
	Mitigation Strategy: Extensive code development to ensure accurate data is pulled	
	Contingency Plan: Enable policy controls to determine which systems may perform the form filler action	

8. Unauthorized Completion of Forms

Description: An unauthorized user completes the form filler actions	Problem: Unlikely	Impact: High
	Mitigation Strategy: Create a user-friendly interface and provide training	
	Contingency Plan: Enable policy controls to determine user access to form filler action	

6 ANALYSIS OF ALTERNATIVES

The major goal of this project is to develop a solution that allows for the pre-population of defined data elements from EHR to clinical research systems, based on current organizational infrastructure and workflow. Ultimately, the project aims to decrease research timelines and costs by reducing the need for manual data transmission across systems in addition to improving organizational workflow.

This project will aid in reaching the organization's goals to 1). Conduct innovative and collaborative research and integrate this knowledge into the practice of medicine 2). Advance the early detection, treatment, and prevention of disease and 3). Facilitate operational processes that support the efforts of the clinical research team in the timely initiation, management and completion of clinical trials at Emory. Through these goals and project implementation, the organization will contribute to federal initiatives to

- Support and encourage the use of seamless data exchange
- Advance scientific knowledge and innovation
- Advance the health, safety, and well-being of the American people
- Increase the economic impact of Federally- funded research and development by accelerating and improving the transfer of new technologies from the laboratory to the commercial marketplace

5.1 ALTERNATIVES

Several methods for systems integration were considered for this project. This section briefly outlines the strengths and weaknesses of those alternatives.

- **Research Data Warehouse**

  A research data warehouse aggregates routinely de-identified patient information from EHR, research systems, registries, and biorepositories.

  Advantages:

  - Incorporates all research systems across the enterprise to aggregate and store de-identified data
  - Potential to utilize current systems

  Disadvantages:

  - Requires data analyst to pull discreet information or the purchase of user friendly software
  - Requires significant changes to the information and technology architecture including the acquisition of new applications to support data relationship management
  - Requires a nightly load to update the database
  - Does not auto-populate research databases with relevant data

- **Web Services ("Getting a Look at Web Services ")**

  Web services allow for the exchange of data between applications and systems using open internet protocols.

  Advantages:

  - Availability over the internet
  - Built on top of open standards
  - Provides application to application interaction
  - Can use software written in various program languages and running on various platforms

  Disadvantages:

  - Does not auto-populate research databases with relevant data
  - Insufficient security or authentication requirements for the exchange of PHI
  - High vulnerability to breached information
  - Requires the acquisition of additional software

- **Current Infrastructure**

  Siloed research systems that require manual transcription from EHR data.

  Advantages:

  - Does not require additional resources
  - Organizational familiarity with the research process

  Disadvantages:

  - Does not auto-populate research databases with relevant data
  - Requires extensive training for incoming staff
  - Lack of consistent information across systems
  - Decreases clinical productivity/workflow

6 EVALUATION MATRIX

Federal goals and initiatives as well as organizational goals as presented in the strategic alignment section along with organization infrastructure were translated into criteria used to evaluate project alternatives. Each architecture was rated on a scale from 1 to 5, with 1 being the worst and 5 being the best. The scores were totaled to indicate which solution best aligned with the identified goals, initiatives, and infrastructure.

Criteria	Retrieve Form for Data Capture	Research Data Warehouse	Web Services	Current Infrastructure
Interoperability	5	3	4	1
Deployable with Current Infrastructure	3	2	3	5
Data Auto-population	5	1	1	1
Reduces Transcription Errors	5	1	2	1
Improves Workflow	4	3	3	1
Increases Productivity	4	3	2	1
Increases Efficiency	5	3	3	2

Decreases Long Term Research Costs	4	3	3	2
Initial Development Costs	1	1	1	5
Information Security & Privacy	5	5	2	5
BA Strategic Alignment	4	3	3	4
IA Strategic Alignment	4	4	2	5
TA Strategic Alignment	3	3	2	5
Total Score	**52**	**35**	**31**	**31**

Based on the evaluation matrix, RFD and an enterprise data warehouse received the highest scores. The infrastructure to deploy these solutions is already established within the organization, but will require modifications to the current applications or an expansion of the Clinical Data Warehouse. Ultimately, to achieve interoperability so that research databases are auto-populated from EHR, the Retrieve Form for Data Capture method is the best option.

8 COST ANALYSIS (Advice, 2017)

Estimated initial cost for the project are outlined below. These estimates are made based on software costs, the existing infrastructure including hardware and software, and needs to meet project objectives.

	Cost in U.S. Dollars
Annual Licensees	**180,000**
Installation and Setup Installing software, configuring databases	**220,000**
Year 1 Customization and Integration Customizing systems to meet integration needs	**150,000**
Year 1 Data Migration Migrating data from existing systems, format conversion and data scrubbing	**400,500**
Year 1 Training Webinars and custom training sessions	**30,000**
Maintenance and Support Annual maintenance contract, which includes support, ongoing updates, and patches for bugs and issues.	**360,000**
Hardware Applications and database servers, networking infrastructure	**250,000**
Labor IT staff compensation	**200,000**
TOTAL:	**$1.79 million**

APPENDIX B: KEY TERMS

The following table provides definitions for terms relevant to this document.

Term	Definition
EHR	Electronic health records
EDC	Electronic Data Capture
RFD	Retrieve Form for Data Capture
IDC	Integrated Data Capture
CTMS	Clinical trials management system
eCRF	Electronic Case Report Form

APPENDIX C: INITIATIVES, STRATEGIES, AND GOALS

- **Emory University Research Goals**
  URL: http://www.emory.edu
    - Conduct innovative and collaborative research and integrate this knowledge into the practice of medicine.
    - Advance the early detection, treatment, and prevention of disease.
- **Emory University Office for Clinical Research Mission**
  URL: www.ocr.emory.edu
    - Streamline the clinical trial process
    - Enhance Patient Safety
    - Increase protocol adherence and regulatory compliance
- **Food and Drug Administration (FDA) Center for Drug Evaluation and Research Strategic Initiative**
  URL: http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/.
    - Support and encourage the use of seamless data exchange
- **FDA Strategic Initiative**
  URL: http://www.fda.gov/downloads/drugs/guidances/ucm501068.pdf
    - Advocate the need for capturing source data, include data originating in health systems
- **Department of Health and Human Services (DHHS) Strategic Goals**
  URL:
    - **Goal 2-** Advance Scientific Knowledge and Innovation
    - **Goal 3-** Advance the Health, Safety, and Well-Being of the American People
- **President's Management Agenda (PMA) Strategic Goals**
  URL: https://www.performance.gov/node/3395?view=public
    - Increase the economic impact of Federally- funded research and development by accelerating and improving the transfer of new technologies from the laboratory to the commercial marketplace

References

(CDC), C. f. D. C. a. P. (2010). *CDC Unified Process Templates*. Retrieved from https://www2a.cdc.gov/cdcup/library/templates/.

Administration, F. a. D. (2015). *Performance Report to Congress for the Prescription Drug User Fee Act*. Retrieved from http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UserFeeReports/PerformanceReports/UCM497750.pdf

Administration, U. S. D. o. H. a. H. S. F. a. D. (2016). Use of Electronic Health Record Data in Clinical Investigations - Guidance for Industry Retrieved from http://www.fda.gov/downloads/drugs/guidances/ucm501068.pdf

Administration, U. S. F. a. D. (2014, 12/09/2014). About the Center for Evaluation and Research Retrieved from http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/

Administration, U. S. F. a. D. (2015, 12/18/2015). Source Data Capture from Electronic Health Records (EHRs). Retrieved from http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm464653.htm

Administration, U. S. F. a. D. (2016). CFR- Code of Federal Regulations Title 21. *21 CFR part 11.* Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=812

Advice, S. (2017). Total Cost of Ownership Calculator. Retrieved from http://www.softwareadvice.com/tco/

Collier, R. (2009). Rapidly rising clinical trial costs worry researchers. *Canadian Medical Association Journal, 180*(3), 277-278.

De Moor, G., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., . . . Coorevits, P. (2015). Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of Biomedical Informatics, 53*, 162-173. doi:http://doi.org/10.1016/j.jbi.2014.10.006

Dee Cantrell RN, B., MS, FHIMSS. (n.d.). *Healthcare Information Exchange*. Emory University Healthcare. Retrieved from https://hometownhealth.wikispaces.com/.../Provider+Community+-+Dee+Cantrell.pptx

DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics, 22*(2), 151-185. doi:http://dx.doi.org/10.1016/S0167-6296(02)00126-1

Emory. (2017). Research. Retrieved from http://www.emory.edu/home/research/

Enterprise, I. t. H. (2011). IHE IT Infrastructure Technical Framework Supplement 10 Retrieve Form for Data Capture (RFD) Trial Implementation.

Evaluation, O. o. t. A. S. f. P. a. (2014). Examination of Clinical Trial Costs and Barriers for Drug Development. Retrieved from https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development

Getting a Look at Web Services Retrieved from http://www.dummies.com/programming/net/getting-a-look-at-web-services/

Goldberg, S. I., Niemierko, A., & Turchin, A. (2008). Analysis of Data Errors in Clinical Research Databases. *AMIA Annual Symposium Proceedings, 2008*, 242-246.

Goodman, K., Krueger, J., & Crowley, J. (2012). The Automatic Clinical Trial: Leveraging the Electronic Medical Record in Multi-site Cancer Clinical Trials. *Current oncology reports, 14*(6), 502-508. doi:10.1007/s11912-012-0262-8

Health, U. S. N. I. o. (2015, December 2015). "Learn about Clinical Studies". Retrieved from https://clinicaltrials.gov/ct2/about-studies/learn

HealthIT.gov. (2015, Febraury 6, 2015). Meaningful Use Definition and Objectives. Retrieved from https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives

Kalil, T. a. W., J. (2016). Cross-Agency Priority Goal. Retrieved from https://www.performance.gov/node/3395?view=public

Markus Schmidt, U. S.-B., Nadia Harbeck, Nicole Harzendorf, Stefan Paepke, Marion Kiechle, Gudrun Zahlmann. (2009). EHR and EDC Integration in Reality. Retrieved from http://www.appliedclinicaltrialsonline.com/ehr-and-edc-integration-reality

Marsha Laird-Maddox; Susan B. Mitchell, M., RN; and Mark Hoffman, PhD. (2014). Integrating Research Data Capture into the Electronic Health Record Workflow: Real-World Experience to Advance Innovation. *Perspectives in Health Information Management*(Fall 2014), 1-10.

Medicine, V. U. S. o. (2017). Research Data Warehousing. Retrieved from https://medschool.vanderbilt.edu/dbmi/research-data-warehousing

Murphy, E. C., Ferris, F. L., & O'Donnell, W. R. (2007). An Electronic Medical Records System for Clinical Research and the EMR–EDC Interface. *Investigative ophthalmology & visual science, 48*(10), 4383-4389. doi:10.1167/iovs.07-0345

Prevention, C. f. D. C. a. (2016, May 26, 2016). Meaningful Use. Retrieved from https://www.cdc.gov/ehrmeaningfuluse/introduction.html

Register, F. (2015, 06/26/2015). Source Data Capture From Electronic Health Records: Using Standardized Clinical Research Data. Retrieved from https://www.federalregister.gov/documents/2015/06/26/2015-15644/source-data-capture-from-electronic-health-records-using-standardized-clinical-research-data

Research, O. f. C. (2016). *Active Clinical Reserach Studies per Investigator Report as of December 5, 2016*. Retrieved from https://secure.web.emory.edu/med/research/ocr/secure/November%202016%20Clinical%20Studies%20by%20PI%20and%20Department.pdf

Sampat, K. (2017). What is the detailed cost breakdown of an expensive clinical trial? Retrieved from https://www.quora.com/What-is-the-detailed-cost-breakdown-of-an-expensive-clinical-trial

Samtani, G. (2002). Top 10 Web Services Security Requirements. Retrieved from http://www.techrepublic.com/article/top-10-web-service-security-requirements/

Sung, N. S., Crowley, Jr, W. F., Genel, M., & et al. (2003). Central challenges facing the national clinical research enterprise. *JAMA, 289*(10), 1278-1287. doi:10.1001/jama.289.10.1278

Victora, C. G., Habicht, J.-P., & Bryce, J. (2004). Evidence-Based Public Health: Moving Beyond Randomized Trials. *American Journal of Public Health, 94*(3), 400-405. doi:10.2105/AJPH.94.3.400