Health | *Transformation*

MITRE's Healthcare Research Program

Fiscal Year 2012

MITRE

Contents

Executive Summary 1
Vision. 1
FY12 Research Portfolio Budget Breakdown 2
Project End-of-Year Reports 2
Connect Goal 3
hData 4
Efficient De-Identification Using Targeted Human Review 6
hArchitecture 8
ABLE: Identity Matching in Healthcare 9
Empower Goal 12
hReader 13
HealthAction Patient Toolkit 15
Kairon Consents: Trust Aware Consent Decisions 18
Equip Goal 21
Healthcare Analytics Roadmap 23
Privacy Testing in the Healthcare Environment. 24
TranScript: Detecting Discrepancies in Pharmaceutical Prescriptions 25
Eyes First 28
Analytics for Rehabilitative Motion Sensing (ARMS). 29
Laser-Enabled Gait Recording System (LEGS) Walter Reed Gait Collaboration 31
Making Predictions from Examining Health Data 32
Large-Scale Data Analytics for Medical Records 34
Automating Fact Extraction from Medical Records 36
Informatics for Integrating Biology and the Bedside (i2b2) Natural Language Processing (NLP) Challenge. 38
Unlocking Patient Data for Translational Medicine 40
Privacy-Preserving Data Mining (PPDM) 42
DrugOrNot 43
Align Goal 45
Healthcare Technology Investment Modeling 45
Summary, Conclusion, Future Directions 47

Executive Summary

Vision

As the nation's largest purchaser of health services, the federal government seeks to accelerate the transformation of the health sector to achieve higher quality and better public health outcomes at sustainable costs. The MITRE Health Transformation innovation investment area funds the research of principal investigators from across MITRE who apply their advanced knowledge of MITRE's core competencies to help our sponsors advance progress in the health domain. A robust research program is critical to our ability to perform the federally funded research and development center (FFRDC) mission for our sponsors.

Research targets are selected based on a multiyear roadmap for the health domain that will move the nation toward an integrated health system. At a high level, there are four goals in MITRE's Health Transformation strategy that will help our sponsors move the country toward that goal, namely:

1. **Connect** health information by moving from paper-based silos to interoperable, secure, and private electronic health records across the health sector;
2. **Empower** patients/consumers to manage their health and healthcare by extending information technology (IT) into personal health management;
3. **Equip** healthcare decision makers with the tools and information they need to set policy and ensure safe and high-quality care; accelerate research to move promising, safe, and effective new therapies and devices from "bench to bedside;" and
4. **Align** incentives, measures of quality, and payment systems for a sustainable national healthcare system.

Over time, the healthcare domain—and the challenges it presents—reveals itself more fully to MITRE personnel through growing sponsor work programs, progressive research, and engagement with the complex and diverse healthcare community. Thus, the MITRE Health Transformation innovation investment area topics also evolve over time, in concert with staff core competencies and capabilities, as MITRE refines its role in the health domain. For example, in previous years, "acceleration of research" has been a key topic in the research portfolio. In fiscal year 2012 (FY12), the portfolio made solid progress in providing tools and technology advances to researchers, thereby accelerating research. Thus, the "accelerating research" component of the goal was absorbed into the broader outcome of equipping decision makers.

Each Health Transformation research project addresses one or more barriers to achieving progress that require novel systems engineering approaches, new standards and techniques, or the application of advanced technologies pioneered in other domains for MITRE's FFRDC sponsors. Major trends, represented by the goals, are already under way in healthcare in the United States and other advanced countries. However, there are still key barriers that prevent progress toward achieving measurable results. Such barriers pose enterprise systems engineering, policy, and analytic challenges that MITRE can play a major role in solving.

MITRE works effectively with the Government principally because of the company's core competencies and unique position as the new Centers for Medicare and Medicaid Services (CMS) Health FFRDC operator working at the intersection of government, nonprofit organizations, for-profit healthcare organizations and insurers, health information technology (HealthIT) vendors, and public health authorities. *The Journal of the American Medical Association (JAMA)* recently cited systems engineering as one of three keys to countering the effects of an aging population and biotech advances, calling it "essential to spawning cost-moderating care delivery innovations more rapidly."[1] MITRE's enterprise systems engineering and research approach is intended to lower technical barriers while at the same time encouraging and stimulating community dialogue and cohesion around innovative technical approaches. In this way, the national health information community begins to move forward in concert with the White House strategy for digital government in the 21st century.[2]

FY12 Research Portfolio Budget Breakdown

Figure 1 offers an overview of the allocation by portfolio goal. The funding allocation for the Equip goal—a key, new thrust—was almost half of the portfolio investment, with the Connect goal in transition, the Empower goal gathering momentum, and the Align goal research only just beginning.

Project End-of-Year Reports

The following pages provide project descriptions, accomplishments, and plans for each FY12 Health Transformation MITRE Innovation Program (MIP) project, organized by overall portfolio goal: Connect, Empower, Equip, or Align.

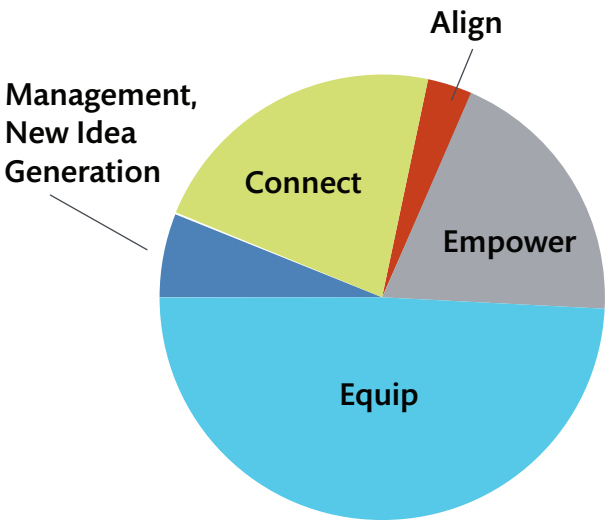

Figure 1. FY12 HCT Actuals by Portfolio Goal

[1] A. Milstein and S. Shortell, "Innovations in Care Delivery to Slow Growth of US Health Spending," Journal of the American Medical Association, Vol. 308, No. 14, pp. 1439–1440, Oct. 10, 2012.

[2] "Presidential Memorandum—Building a 21st Century Digital Government," The White House, Office of the Press Secretary, May 23, 2012. Available: http://www.whitehouse.gov/the-press-office/2012/05/23/presidential-memorandum-building-21st-century-digital-government.Accessed Oct. 16, 2012.

Connect Goal

A key goal of our Innovaiton Portfolio is to **Connect** health information by moving from paper-based silos to interoperable, secure, and private electronic health records across the health sector.

In furtherance of the Connect goal, hData efforts to define a lightweight standard option for health information exchange were recognized by the Object Management Group (OMG) adopting the hData RESTful Transport (HRT) Specification (also known as hData REST Binding for RLUS) and by Health Level 7 (HL7)—the leading international health information technology standards development organization—approving the hData record format as a Draft Standard for Trial Use (DSTU). hData has been noted by industry and VHA thought leaders as a key enabler of the transformation of health information exchange from data sharing to semantic-health-web enabling, described by hospital information systems expert Tom Munnecke as "a step towards a semantic approach to clinical data."[3]

Transition activities for projects furthering the Connect goal included the following:

- Using hData principles and research products with respect to granular data, the RESTful Health Exchange (RHEx) project, funded by the Office of the National Coordinator for Health Information Technology (ONC) Federal Health Architecture, successfully demonstrated the secure exchange of medical data. This project also enabled an even broader open source community through its exposure via the ONC Standards and Interoperability community(http://wiki.siframework.org/RHEx).
- MITRE is now a driving force in defining the medically relevant application task for Informatics for Integrating Biology and the Bedside (i2b2) 2013, based on experience in gathering community input for BioCreative Challenge Evaluations. The inspiration for an Advisory Group comes directly from prior experience in BioCreative.
- The MITRE Identification Scrubber Toolkit (MIST), an open source software system developed under the Efficient De-Identification Using Targeted Human Review MITRE-sponsored research (MSR) project, added to its user community and impact on the privacy-enabled sharing of complex narrative from clinical records for secondary use research (http://sourceforge.net/projects/mist-deid/). The team made improvements to the re-synthesis engine and released a stand-alone resynthesis engine to Salt Lake City Veterans Health Administration (VHA) partners. Additionally, the team provided several customizations to our research partners within the ONC Strategic Healthcare Advanced Research Projects, Area 4 (SHARP4) program (Mayo Clinic, Group Health Cooperative). MIST is the de-identification solution used for sharing data within the SHARP4 program. Finally, the team transitioned MIST to our research partners at Cincinnati Children's Hospital Medical Center, University of Michigan Health System, and Vanderbilt University.

[3] T. Munnecke, Feb. 27, 2012, OSEHRA, "Some background material for Semantic Web topics for Architecture Conference Call Feb 28." Available: http://www.osehra.org/blog/some-background-material-semantic-web-topics-architecture-conference-call-feb-28. Accessed Feb. 27, 2012.

hData

Principal Investigator: Dr. Mark Kramer

Project Description

hData is an emerging standard for lightweight, scalable clinical information exchange. hData focuses on system-to-system data exchange by creating network-facing interfaces based on the Internet's REpresentational State Transfer (REST) design model, the Atom Syndication Format—an XML language used for Web feeds—and the Atom Publishing Protocol, a simple HTTP-based protocol for creating and updating Web resources that can be plugged into existing clinical systems. The hData standard organizes clinical information for Web access, defines Web methods for consuming and producing data, standardizes metadata annotation of data, and enables popular authentication and authorization models, such as OpenID and OAuth 2.0. hData is helping sponsors evolve toward a patient-centric vision for continuing health. Because it leverages the Web to a greater extent than alternative health information technology (IT) standards, hData will be particularly important on mobile devices.

Research Objectives and Milestones

The main goal is to make adoption of health IT standards easier. Current standards are unnecessarily complex and constitute a significant barrier to entry. Transport via Simple Object Access Protocol (SOAP) and Simple Mail Transfer Protocol (SMTP) are impediments to mobile health. Mobile health has the potential to reduce costs of chronic disease management and aging. hData is easy to process on mobile devices and greatly enhances simplicity by offering one end-to-end standard. By separating the content model from exchange mechanics, hData allows a cleaner and better-defined standard. By disaggregating clinical information, hData achieves better performance and potentially achieves better privacy control, compared with copying the entire patient record from system to system. Finally, by utilizing Web standards, hData enables popular authentication and authorization models, simplifying security and deployment.

FY12 Accomplishments

hData achieved several significant milestones in FY12:

- The OMG adopted the hData RESTful Transport Specification (a.k.a. hData REST Binding for RLUS). The HRT defines the HTTP methods used to access information expressed according to the hData Record Format (HRF).
- Health Level 7 (HL7), the leading international health standards development organization, approved the hData Record Format as a Draft Standard for Trial Use (DSTU).
- Integrating the Healthcare Enterprise (IHE), an initiative of healthcare professionals and industry dedicated to improving the way standards are deployed and coordinated to achieve better patient care, developed a mobile heath profile designed to align precisely with hData.
- Several HL7working groups , including the Medication Working Group, developed hData Content Profiles (HCPs). An HCP is a specific instantiation of an hData Record Format that allows unrelated organizations to recognize and exchange information.

- The Services Task Force of the Continua Health Alliance, a nonprofit industry organization dedicated to establishing a system of interoperable, personal, and connected health solutions, selected hData as the basis of a next-generation end-to-end (device-to-server) solution.
- hData was presented at several conferences and meetings, including BioIT World 2012, mHealth Summit 2012, and OMG Interconnected Health 2012 (to name just a few).
- The hData team participated in numerous standards development meetings, particularly HL7, OMG, OASIS, Continua, and IHE. The hData team also hosted a Continua Health Alliance meeting at MITRE.
- The hData team participated in the Data Segmentation for Privacy (DS4P) working group of the Standards and Interoperability (S&I) Framework program on ONC for Health Information Technology, focusing on how hData disaggregates the medical record and allows granular access.
- A mobile hData client (hHub) was developed for Android, which demonstrated efficient and secure data browsing on a mobile device. The Patient Data Server established for this experiment was also used by hReader and medCafe.
- The hData team contributed two presentations to the MITRE Privacy and Security Technical Exchange Meeting, one about releasability and the other on the HL7 Privacy and Security Classification System.

Transition Activities

Many of the above activities could be considered transition activities. One project in particular deserves special mention. Funded by the Office of the National Coordinator for Health IT, Federal Health Architecture, the RESTful Health Exchange (RHEx) Project demonstrates medical data exchange using hData principles. The first goal of the project was to demonstrate that RHEx could provide secure transport for sensitive data, which was achieved using OAuth and OpenID Connect in a "single sign-on" environment. As a result of successfully demonstrating secure RESTful transport, RHEx recently became a full-fledged S&I Framework initiative.

Future Directions

- Final ballots for HL7 and OMG after incorporation of feedback from trial/beta periods.
- Promote adoption by Continua Health Alliance in next-generation standards profile involving hData over Message Queuing Telemetry Transport (MQTT).
- Support RHEx and incorporation of feedback from implementation.
- Create a content profile for HL7's Fast Healthcare Interoperability Resources (FHIR), which defines a set of resources that represent granular clinical concepts.
- Support DSTU for medication statement service.
- Continue support of ONC, S&I Framework RHEx and DS4P, IHE, and other sponsors, government organizations, and SDOs.

Efficient De-Identification Using Targeted Human Review

Principal Investigator: John Aberdeen

Project Description

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) defines a safe harbor method for creating releasable de-identified, unstructured (free-text) records by first removing 17 identifier types (name, date, location, Social Security number, etc.). Although current state-of-the-art automated de-identification systems (e.g., MITRE Identification Scrubber Toolkit [MIST]) can de-identify unstructured data at a high rate of accuracy, in practice, data stewards do not typically release such de-identified data under the HIPAA safe harbor policy. Rather, data stewards release data via limited-use, two-way agreements. While these two-way limited-use agreements have enabled some research activities (including the development of MIST), the lack of large amounts of widely available data (particularly patient-level longitudinal data) continues to create a bottleneck for the research community. The recent growth of research consortia (eMERGE, SHARP, Consortium for Healthcare Informatics Research [CHIR]) that deal with clinical data points sharply to the need to share data safely across institutions. Additionally, there is growing interest in deploying clinical natural language processing systems via secure cloud computing. For clinical researchers to take advantage of this, they need efficient de-identification that meets the safe harbor standard. A recent survey of Institutional Review Board (IRB) managers indicated that they would be much more likely to approve using Natural Language Processing (NLP) systems in the cloud if patient data were fully de-identified. Currently, de-identification is achieved very inefficiently by manual (human) redaction of each document.

Research Objectives and Milestones

MITRE's goal is to enable data stewards to share unstructured data for secondary use while maintaining patient privacy. The relative lack of such data has been a barrier to clinical and biomedical research for many years. Providing an efficient method of achieving de-identification either for limited use or, even better, to the HIPAA safe harbor standard, would enable wider exchange of such data for secondary use. The resulting larger quantities of unstructured data available would be a boon to the growing number of clinical research consortia. In addition, clinical natural language processing systems could then be deployed via secure cloud computing, processing patient data securely.

Efficient safe harbor de-identification of unstructured text is achievable by using prioritized human review only when it is truly needed. Prioritizing for human review of those sections and documents that most need it can improve the efficiency of de-identification. MITRE assesses performance by comparing the results of MIST plus targeted human review against the performance of manual human review.

FY12 Accomplishments

The team performed targeted review experiments through simulation using the Informatics for Integrating Biology and the Bedside (i2b2) de-identification corpus. MITRE showed that it

is indeed quite feasible to "bin" documents and sentences based on MIST's de-identification confidence score, and that the confidence bins have the predicted positive relationship with de-identification accuracy. We found further that the most efficient use of human annotators' time is to de-identify at the document level and target review at the sentence level.

Collaborating with researchers at Group Health Cooperative and Vanderbilt University, we conducted a preliminary study of the protective effect of resynthesized surrogates in de-identified text, or Hiding in Plain Sight (HIPS). All de-identification methods, whether automated or manual, leave behind a certain number of residual identifiers. Careful replacement of known identifiers with realistic but made-up surrogates (resynthesis), causes the residual identifiers to blend in, making them far less detectable. The experiment showed that approximately 90 percent of residual identifiers can be effectively concealed by this HIPS approach. Further, an analysis of errors points to areas where resynthesis can be improved to increase this protective effect.

MITRE released MIST 1.3 (open source) this year. It includes functionality to facilitate autotagging for similar strings, as well as numerous improvements to the experiment engine, server, input and output formats, and scoring.

Transition Activities

In addition to releasing MIST version 1.3, MITRE improved the resynthesis engine and released a stand-alone resynthesis engine to the Salt Lake City VHA partners. MITRE also provided several customizations for research partners within the ONC SHARP4 program (Mayo Clinic, Group Health Cooperative). MIST is the de-identification solution used for sharing data within the SHARP4 program. We have also transitioned MIST to research partners at Cincinnati Children's Hospital Medical Center, University of Michigan Health System, and Vanderbilt University. Finally, MITRE demonstrated MIST to representatives of the Internal Revenue Service (IRS) at the IRS Innovation Exchange.

Future Directions

In FY13, the team will seek further transition partners for MIST, including a stewardship partner capable of taking on the role of future development and maintenance of MIST. We will also transition MIST to a new partner (Marshfield Clinic) in support of its effort to integrate de-identification with note creation (*real-time* de-identification). Finally, we will continue our collaboration with Group Health Cooperative and Vanderbilt University investigating the protective effect of resynthesis (*hiding in plain sight*).

hArchitecture

Principal Investigators: M. David Allen and Dr. Marc Hadley

Project Description

The government's current goals for health IT, including big data analytics and outcome analysis, rely on the assumption that they can properly source and assemble information from thousands of independent and autonomous entities. There are many possible ways to architect a large, loosely organized system of data. Each method has its pros and cons, strengths and weaknesses, and immediate out-of-the-box benefits. However, the current approaches for health exchange focus on full Electronic Health Records (EHRs) (C32) across a Service-Oriented Architecture (SOA), which will not provide the necessary support for the government's identified goals. Given limited success with C32/SOA work, the community is currently piloting direct peer-to-peer connections using less-than-productive protocols.

Where different use cases and needs are apparent, different architectures should be chosen. This research investigates alternate architectures for health data exchange to propose ways to make an appropriate health data exchange architecture choice for a given set of clinical-data usage needs.

Research Objectives and Milestones

The problem with the status quo isn't a lack of technical excellence; it's a failure to temper those technical considerations with the organizational and cultural incentives that control how the larger complex system actually operates. To solve the problem, MITRE:

- Examined multiple architecture options (outside of the default conventional wisdom that SOA is best); and
- Defined an example framework for evaluating those options in light of their relative advantages and disadvantages to participants.

Because the healthcare system has so many different players there is no single best option that maximizes all participants' equities. MITRE investigated other architectural choices to determine how best to respect the system organizational "realities" while at the same time employing the best possible technical solution for those realities. In other words, we built "Consumer Report" white papers for government healthIT architectures. The first of these white papers, titled "Fit for Purpose: How to Select Data Exchange Standards," was delivered in FY12 for hArchitecture.

The hArchitecture Team met with MITRE subject matter experts (SMEs) in the health IT space in order to prioritize the efforts. Rather than examining all criteria associated with sponsors' architectural choices, the SMEs' feedback focused the project on a goal of helping sponsors with the problem of how to select data standards for a given challenge. Data standard selection predicated on developing characterizations of families of data standards, as well as creating and standardizing decision rules to guide sponsors' data architecture approach selection,

is a key piece of an overall information sharing architecture. The MITRE SMEs considered this a useful chunk of work that could help move sponsors forward if they were given good advice about how to go about making this key decision.

From a technical perspective, there are a number of challenges around how best to characterize standards and architecture options. Each sponsor's unique problems represent a tradespace of potentially dozens of different variables. Several MITRE projects focus on best strategies for arranging presented information about architectural options and data standards so that such information is actionable for sponsors.

FY12 Accomplishments

- Delivered the first white paper focused on data standard selections; and
- Developed lessons learned that focused on how to "tame the tradespace."

Transition Activities

- Delivered the decision framework into teams supporting data exchange decisions in our sponsor work program;
- Planned FY13 work with the Federal Health Architecture (FHA) Data Exchange and Interoperability (DE&I) Working Group; and
- Planned FY13 work with the FHA Managing Board.

Future Directions

- Continue to apply the concept and framework in our advice and guidance to sponsors challenged with the selection of data exchange techniques and standards components underlying health data exchange architectures.

ABLE: Identity Matching in Healthcare

Principal Investigators: Gail Hamilton and Keith Miller

Project Description

A major problem in healthcare today is that it is impossible to obtain a patient's entire health record without adequate patient identification and patient matching solutions. From a policy perspective, a national patient identifier is not an option. The ABLE project addresses the patient identification challenge by leveraging the extensive tools and expertise gained from name matching for national security and applying them to the healthIT domain. Unlike other domains, where "false negative" matches were the critical issue (e.g., missing a terrorist in a name-matching function), the patient matching domain has the challenge that "false positives," defined as associating a patient history with the wrong patient, could be harmful if not deadly. The end goal of this project is to address a challenge best codified in a finding from an ONC Privacy and Security Tiger Team: in healthcare "[t]here is little research on best practices for

matching."[4] This project has the goal of providing a framework to apply industry best practices and tools to enable measurement of effectiveness of healthcare identity matching solutions.

Research Objectives and Milestones

Although access to patient demographic data in healthcare is very tightly controlled due to HIPAA laws, one goal of ABLE is to create and characterize Personal Identity Information (PII) data without having continuous hands-on access to this data. As a result, one of the main goals of the ABLE project is to develop best practices, methods, and tools to characterize the salient features of the data and create a representative artificial data set. This representative artificial data set will then be used to assess how well commercial and open source identity matching programs and algorithms perform in matching patients' records under different circumstances, whether some matching programs and algorithms handle different matching challenges better than others, and whether some are better suited to certain use cases.

FY12 Accomplishments

During FY12, the team tackled many foundational issues related to delivering a framework to enable better nationwide patient matching, including: What are the best practices? What are the use cases? What are the key differences between healthcare and more traditional identity matching domains, such as national security? To answer these questions, a primary objective this year was to establish an ongoing collaboration with a group of SMEs and at least one healthcare entity already addressing this problem. Through this collaborative group, ABLE determined how the tools available through the MITRE Identity Matching Lab (IML) can help enable better patient matching in healthcare. We were successful in this goal, establishing ongoing communications that culminated in an intensive and extremely productive two-day workshop at the University of Texas Health Science Center at Houston (UTH) with representatives from academia and the medical community from UTH and private industry representatives from Argo Data. The outcomes included a request for a white paper to be presented at a national conference, detailed discussions, and a higher degree of understanding of the main issues that UTH, providers, and SMEs had encountered. These discussions focused on creating a taxonomy of various issues around each attribute used in patient matching and issues related to current matching techniques. MITRE's experience in providing guidance for sponsors tackling identity matching challenges, using the EAGLEs 7 framework, was key in establishing a defined step-by-step approach to solving this issue.

Collaboration/Data: MITRE obtained an agreement granting access to demographic data used by the FFRDC sponsor for matching patients across their enterprise. Access to this data will be a great help because it provides access to one of the largest data sets of demographic data used for healthcare in the United States. The agreement also establishes a path for other data sharing agreements, especially health information, through establishing secure processes and mechanisms for securely hosting this data. Most importantly, this collaboration will develop trust between the sponsor and MITRE in the area of identity matching and secure storage of sensitive data. This collaboration is the first step in helping a MITRE sponsor achieve better

[4] Health IT Policy Committee transmittal letter, Feb. 8, 2011. Available: http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_11673_953462_0_0_18/HITPC-transmittal-letter-Priv-SecTigerTeam-020211.pdf.

patient matching outcomes, particularly when retrieving health information from outside entities. It will also provide the foundation for other health transformation sponsors in the guidance they provide to their stakeholders for the industrywide challenge.

One of the first tools we are developing is a data characterization toolkit for healthcare data. This would allow for analysis of the health data by building a representative data set for patient matching testing in healthcare. This tool will help overcome one of the greatest barriers in most healthcare research—difficulty in gaining access to protected health information (PHI) or even PII from health institutions due to legal restrictions and complex privacy regulations. Through this data characterization toolkit, MITRE can gain an understanding of what a data set looks like without actually having hands-on access. MITRE can then use this information to build a good representation of the demographic data that healthcare institutions currently use for their own patient matching. We also plan to provide this toolkit to industry via open source to enable others to better characterize their data challenges, particularly those related to patient matching performance.

Transition Activities

FY12 transition work is largely aimed at the Department of Veterans Affairs (VA). VA is already engaged and enthusiastic about our work and the potential to help with patient matching, particularly with establishing tools, best practice guidance, and metrics to enable better patient matching results when getting patient information from private providers through Health Information Exchanges. These patient records are required to create a virtual lifetime electronic health record (VLER), a key initiative at VA.

Future Directions

This year's ABLE effort was small and exploratory, geared around building the foundation for future FY13 activities, which will include:

- The use of data characterization tools on health data at UTH and data from a major sponsor;
- Development of a data characterization analysis toolset; and
- Providing tools and a framework to enable measurement of performance of patient matchers.

Finally, at the request of the conference chair, we plan to submit a paper to the Summit on Clinical Research Informatics conference detailing major differences between identity matching in the healthcare domain and in the national security domain.

Empower Goal

A key portfolio goal is to **Empower** patients/consumers to manage their health and healthcare by extending information technology (IT) into personal health management.

The Empower goal has seen vibrant community engagement through its projects in FY12. The hReader project signed a Memorandum of Understanding with Boston-based Partners Healthcare for multiple hReader pilots' collaboration, furthering the ideas underpinning patient-centric health information hosting on mobile devices. The HealthAction Patient Toolkit team reached out to providers and patients to engage and receive feedback on the gaps in and approach for more effectively empowering patients with chronic conditions (and their caregivers) to manage their disease, establishing relationships with providers and researchers at Columbia University and the University of Virginia. HealthAction Patient Toolkit also has the potential to be leveraged as an interface with VA's online personal health record, MyHealtheVet, to increase veteran patient engagement.

Our granular patient consent framework, Kairon Consents, established a relationship with the Substance Abuse and Mental Health Services Agency (SAMHSA) and the Veterans Health Administration (VHA). Kairon was part of a demonstration of the possibilities for patient empowerment, consenting to share its private health data done in partnership with SAMHSA and VHA. The impact of the Kairon Consents work has gone a long way toward making SAMHSA more confident about the feasibility of powerful, patient-driven consent management, and they've engaged the MITRE team to advise them on their future plans for patient consent in the extremely sensitive areas of mental health and substance abuse data.

Transition activities for projects furthering the Empower goal included:

- The Kairon Consents: The Trust Aware Consent Decisions project established a relationship with SAMHSA, providing SAMHSA more confidence in the feasibility of powerful, patient-driven consent management. The project also created a relationship with Health Management Sciences, a commercial company that sells provider affiliations and credentials data. This company is now interested in selling its data for use in consent systems. MITRE project staff provided knowledge about requirements and data quality concerns. Finally, industry recognized the MITRE/Kairon Consents project and its contributions to open source (http://sourceforge.net/projects/kaironconsents/). For example, John Halamka, CIO and industry thought leader from Beth Israel Deaconess Medical Center, blogged, "Once data is segmented, we can then record patient privacy preferences for each segment. How do we do that? MITRE has created an open source patient consent policy management tool called Kairon Consents that is available for use now." [5]

[5] J. Halamka, Life as a Healthcare CIO, "Data Segmentation," Mar. 5, 2012. Available: http://geekdoctor.blogspot.com/2012/03/data-segmentation.html. Accessed Oct. 17, 2012.

hReader

Principal Investigators: Dave Hill and Harry Sleeper

Project Description

Currently, fewer than 7 percent of Americans have used a website to access their online health information and even fewer have used mobile health applications. Americans do not use mobile and Web technologies for these purposes due to lack of awareness, affordability, security, and access to their own or family members' health data, as well as health data for third parties for whose care they are responsible. It is currently very difficult for patients to access their healthcare information; many providers do not provide Web-based or mobile-based solutions. When provided, most mobile solutions lack the security to provide adequate privacy protections for a complete data set.

To address these issues and demonstrate the art of the possible to our health transformation sponsors and their stakeholders, we built the hReader iPad application and mobile health framework. hReader provides patients with all of their family's health information. The relevant information is provided in an easy-to-understand format and stored locally on the iPad, allowing patients to use that information during healthcare sessions to ask better questions and have an improved dialogue with providers. hReader strategically leverages other open-source health initiatives at MITRE, including hData, RHEx, and popHealth, promoting those related standards as well as providing an engaging prototype to give patients the information necessary to improve their health. hReader also implements an open applet architecture, allowing third parties to create hyperpersonalized applets that engage patients toward better health. hReader leverages MITRE's security expertise to provide necessary privacy and raise the bar on mHealth security.

Research Objectives and Milestones

The hReader research seeks to determine whether a mobile family health manager,together with relevant health information exchange standards and systems, can provide the tools to empower and engage families to improve their health and have better communication with their health providers. FY12 research tasks continuing into FY13 include:

- Design and implement an easy-to-use, secure mobile health management platform;
- Vet, document, and understand the mobile health family use cases;
- Establish collaboration with major healthcare providers to validate hypotheses and demonstrate the value of a true mobile health strategy;
- Enable secure, privacy-enabled mobile access to personal health data;
- Create an applet framework to integrate health application functionality;
- Provide documentation and support to third parties to write applets for hReader; and
- Pilot hReader connected to RHEx and hData Patient Data Server.

FY12 Accomplishments

- Designed and developed the hReader iPad application and applet framework along with more than 15 sample applets;
- Demonstrated our framework flexibility by integrating applets developed by other teams, including TBI Tracker and a custom Anduino-based Bluetooth temperature sensor;
- Publicly released hReader website (www.hreader.org) along with Google Group discussion forums and a social network presence (Twitter, Facebook);
- Signed a Memorandum of Understanding with Partners Healthcare for hReader pilot collaboration;
- Formed relationships with Beth Israel Deaconess Medical Center and Maine Health Information Exchange (HIE), creating possibilities for FY13 collaborations and pilots;
- Completed an hReader user interface review with SMEs and patients at Partners Healthcare;
- Prepared for initial pilots using synthetic patient data from Partners test systems and real patient PHI pilots at Partners using the RHEx exchange services;
- Completed a technical evaluation of the Partners Healthcare Patient Information Model;
- Completed INFOSec testing and validation to achieve DISA STIG-level security; and
- Raised the visibility of hReader to ONC, White House fellows, and the Maine HIE through the RHEx project.

Transition Activities

Through its interaction with the RHEx project, hReader has been discussed with ONC as well as White House fellows, Maine HIE, and Hitachi, Limited's HealthIT division. All have expressed interest in the program. The launch of the hReader website, along with the upcoming release of source code on Github will allow interested parties outside of MITRE to experiment with the code and use it to develop commercial products. The focus in FY13 will be to build a vibrant development community to promote hReader as a mobile health development solution, leveraging other MITRE health data standards and relevant open source community engagement.

Future Directions

The targeted activities for hReader in FY13 include:

- Completion of the synthetic patient pilot at Partners Healthcare;
- Third-party security review;
- Patient pilot at Partners Healthcare using real patient PHI data;
- Maine HIE Pilot;
- The building of a vibrant online community around hReader.

HealthAction Patient Toolkit

Principal Investigator: Kristina Sheridan

Project Description

Four in five healthcare dollars (78 percent) are spent on behalf of people with chronic conditions.[6] The Department of Health and Human Services (HHS) reports that by 2020, about 157 million Americans will be affected by chronic illnesses.[7] The HealthAction Patient Toolkit project addresses the financial and personal burden of chronic illness by increasing patient awareness of their health status, enabling patients to participate more fully with their care team, and improving their treatment compliance. The end goal of this project is to identify ways to engage patients more effectively in their care in order to reduce the financial and personal burden of chronic illness.

Research Objectives and Milestones

Many current healthcare improvement initiatives are focused on the provider. This research focuses instead on empowering the patient. The goal of the HealthAction Patient Toolkit project is to identify ways to increase patient engagement in their care in a way that will improve patient-provider interactions, increase treatment compliance, and lower costs.

To meet this goal, the team will develop best practices, methods, and tools to enable patients to apply these ideas, as well as execute trials to measure the impact of developed products on patients with chronic illness. The team will research gaps that currently prevent optimal patient-provider interactions, treatment compliance, and cost containment. The team will develop use cases and requirements that can be leveraged to build an IT solution to close the gaps. The team will also develop a mobile application prototype that will be used in trials to measure the impact of closing these gaps.

FY12 Accomplishments

During FY12, the HealthAction Patient Toolkit team researched the key gaps existing today that prevent chronically ill patients from receiving the best care at lower costs and maintaining compliance with their treatment plans. We developed use cases and detailed requirements that can be used to create IT solutions to close those gaps. Key gaps addressed by these requirements include:

- Tools to allow patients to identify a change in status or easily track their status between visits to ensure they provide a complete set of data during a provider visit;
- Interactive analysis of patient data to extract information for both the patient and provider (e.g., cause and effect relationships between medications and symptoms);

[6] G. Anderson and J. Horvath, "The Growing Burden of Chronic Disease in America," *Public Health Reports*, Vol. 119, pp. 263–270, May/June 2004,.

[7] G. Anderson and J. Knickman, "Changing The Chronic Care System To Meet People's Needs," *Health Affairs*, Vo. 20, No. 6, pp. 146–160. Available: http://content.healthaffairs.org/content/20/6/146.full#fn-group-1.

- Easy-to-use tools to increase patient treatment compliance to address complexity of integrated treatment plans, alerts to remind patient to comply, and information necessary to avoid medication errors or interactions;
- Smart access to targeted education, social networking, and disease-specific modeling;
- Tools with incentives and easy-to-use interfaces to engage patients on a regular basis to change behaviors; and
- An ability to combine patient-generated data with data currently existing in EHR systems to provide whole-patient view.

After identifying the gaps and developing use cases and requirements to address them, the team researched existing tools available to identify best practices, lessons learned, potential collaborators, and overlap. The team evaluated mobile IT platforms to identify an appropriate platform to build out a prototype that could be leveraged to measure the success of the methods being used to close these gaps. We chose the iPad based upon the results of this research and the adoption of that platform by the target population. In addition, the team developed best design practices to ensure the final prototype would be reusable to the maximum extent possible.

The development team split into two parts. The first group focused on the core toolkit, designing and building an initial version that will be completed in the early part of FY13 and will be leveraged during testing and evaluation. This core toolkit has been designed generically to be beneficial for all patients and is not specific to one disease. Functionality completed this year includes:

- Framework to allow multiple patients to be added to and managed by the tool;
- Framework to create forms, which has resulted in an ability to collect patient symptom and medication data;
- Data analysis framework that allows the symptom and medication data to be graphed over time to be used to evaluate causes and effects of medication and symptom changes by zooming and expanding to assess data;
- Framework to allow provider information to be stored and synced with the main iPad to avoid data being out of sync (this code will be reapplied to the calendar function); and
- Database structure that leverages existing HHS data standards to ensure compatibility with existing EHR systems.

The second group focused on addressing how to provide smart access to targeted education, social networking, and disease-specific modeling. For this second part, the team created a prototype leveraging other MITRE research in Social Radar for education and sharing on Lyme disease. The group created Navigate Lyme, an iPad application designed to educate patients about Lyme disease using online resources, including social media and an interactive simulation model. The team pulled together a list of frequently asked auestions with answers pointing to vetted websites augmented by search strategies to obtain more information. Online resources included pointers to websites and discussion groups, as well as searching news and publications feeds in real time. In addition, the team interfaced with the Twitter application-programming interface (API) to obtain tweets by topic.

Based on limited available data for Lyme disease, the team built an interactive simulation to investigate the probability of infection based on where a person lives (which U.S. state), how long a person takes antibiotics (2–4 weeks), whether a person lives in a rural or city environment, how much time a person spends in a wooded or grassy area, and whether or not a person has pets. The user runs the model after indicating answers to these questions. The model is run 50 times and the distribution of outcomes is displayed at the state and individual level. The goal of the interactive simulation is education of patients, possible patients, and their families.

We believe that this interactive and educational framework is applicable across a variety of chronic illnesses. The team is working on documenting lessons learned for developing an interactive iPad application for educating a user about an emerging disease. This part of the research resulted in a new MITRE research project focused on using Social Radar for healthcare.

In addition to Navigate Lyme, the second team created a generic tool to allow patients to create, save, and execute searches for the latest information (either publications or Twitter feeds) on any health issue. This feature is being integrated with the main toolkit.

In addition to completing the work above, the HealthAction Patient Toolkit team reached out to providers and patients to receive feedback on the gaps and the approach. The team established relationships with providers and researchers at Columbia University and the University of Virginia. The team also had significant engagement with other internal MITRE research teams and healthcare staff.

In order to evaluate the impact of these new methods, the MITRE team worked in partnership with the University of Virginia team to define a clinical trial with multiple sclerosis patients at UVA Medical Center. This trial, to be executed in FY 13, will evaluate the impact of engaging patients to evaluate the value of additional data about symptoms and medications, and will measure the correlation of the self-reported patient data to external health devices worn by the patient. In addition to this external clinical experiment, the team has identified MITRE staff to participate in a broader, internal evaluation of the toolkit, and has defined a set of evaluation criteria, mainly focused on usability and engagement of this nonchronic population.

Transition Activities

One of the primary sponsors being engaged as a transition for this research is the VA, since this research will provide methods to better engage veterans in their care and it has the potential to be leveraged as an interface with MyHealtheVet. HHS and the Department of Defense (DoD) are also potential users of this research. As an example, the Food and Drug Administration (FDA) has been focused on patient-centeredness and could leverage this work to increase patient engagement with providers and increase feedback regarding medications and devices. The research products that can be leveraged by these sponsors include best practices, use cases and requirements, open source code, and clinical trial results. The team has also started to identify a potential nonprofit organization to house the open source code from the prototype/reference implementation.

Future Directions

The FY12 work completed above built the foundation for future FY13 activities, which include:

- Expansion and testing of HealthAction Patient Toolkit:
  - Completion of functionality, including reports;
  - Agile sprints to address iterative user feedback;
  - Expanded research into user behavior: ensuring user compliance, flexibility, improved usability; and
  - Testing for specific use cases.
- Integration with other MITRE MIPs, including leveraging the hReader framework for security and the ability to push/pull to external systems;
- Partnering with the University of Virginia Broadband Wireless Access & Applications Center (BWAC) research team on a clinical experiment with multiple sclerosis patients to measure the impact of patient-reported data;
- Demonstrating capability to VA and HHS to enable the transition of the capability, best practices, and lessons learned; and
- Releasing the reference implementation and its documentation as open source.

Kairon Consents: Trust Aware Consent Decisions

Principal Investigator: Arnon Rosenthal

Project Description

Obtaining and enforcing consent is a major barrier to data sharing in the healthcare system. Inadequate sharing leads to suboptimal treatment, duplication of work, and slowed research progress. The current system for consents is incompatible with a fully automated health data exchange, with satisfying individual patients' wishes, with maximizing data for secondary use, and with protecting data when requests come over the Internet.

Kairon aims to support general, patient-centric policies in the complex, imperfect healthcare environment. MITRE seeks to enable patients to specify what they want, while minimizing the burden on data custodians, especially small practices and allied professions.

Research Objectives and Milestones

In the consent arena, the project addresses eliciting patient preferences, making them available to deal with requests, and obtaining evidence needed for decisions. MITRE seeks to enable the majority of patients with the ability to create appropriate policies without having either "all" or "nothing" shared or seeing complex Boolean logic, making it possible to mix the concerns of 50 states and multiple usage modes (e.g., emergency care). For trust, MITRE seeks to create a loosely coupled web of trust, combining classical delegation (i.e., trusted authority vouching) with explicit treatment of uncertainty.

FY12 Accomplishments

We have found a simpler way for patients to create appropriate rule sets. Originally, we thought to create user interfaces that could enable patients to write rules with appropriate Boolean conditions. We have since created a "power tools" approach based on two tactics. First, the patient is asked to specify factual "evidence" (e.g., identifying who treats them) and indicate a willingness to treat consultants, etc. as treating clinicians. Second, patients identify their sensitivities on various topics. This information is used as input parameters to an expert-created wizard to create policies that deal with all the subtle problems potentially encountered, including conflicts and uncertainties, without needing to train patients or providers in a new logic language. We also presented the idea of a "What If" interface to SAMHSA. As a result, SAMHSA requested a focus on user interface rather than enforcement.

We created a design paper for mixing preferences that come from patients, states, federal agencies, and organizations. It emphasizes standards and vendor efforts that focus on federal statutes only and do not describe solutions to the variability needed, especially support for defaults, given the reality of this complex ecosystem.

We developed *uncertainty* as a unifying construct, spanning multiple trust challenges, such as claims of emergencies, imperfect annotation of health records, authentication, and vendor data quality.

As part of the above efforts, we have discovered numerous areas where guidance is needed, for example, about prioritizing rules (across patient, state, federal, and organizational stakeholders), describing uncertainty, and validating rule sets, as appropriate.

Our system overview paper received good reviews at the *Journal of Medical Informatics Research*, but reviewers asked for more details. Also, since it was written, MITRE has found that the design philosophy (emphasizing modular, reusable capabilities) enables the team to meet challenges beyond those originally described. MITRE is rewriting the paper to emphasize these benefits.

In FY12, the second year of the project, the Kairon Consents reference prototype has been improved in numerous respects, notably user interface and standards conformance.

Transition Activities

The Kairon Consents reference prototype software was released as open source.

Kairon Consents was presented at Healthcare Information and Management Systems Society (HIMSS) 2012 Conference.

The team transferred knowledge within MITRE and collaborated with MITRE's RHEx project, helping them to design their user interface. We were very active in the Granular Access Flash TEM, presenting Kairon, co-writing the capstone framework for releasability, and advocating for simplifying frameworks. The team also created principles for data annotation.

The team participated in the HL7 Security Working Group and the S&I Segmentation Effort. Based partly on this work, there is wide awareness of the benefits of Internet-based consents and the importance of one rule set per patient. Based on Kairon annotation principles in collaboration with hData, the team quashed a nonevolvable approach proposed for HL7.

The team created a relationship with Health Management Sciences, a commercial company that sells data (provider affiliations and credentials). Health Management Sciences is interested in selling its data for use in consent systems. MITRE has educated Health Management Sciences on some of the requirements and on some data quality concerns.

Kairon Consents is drawing industry attention. For instance, John Halamka's blog states, "Once data is segmented, we can then record patient privacy preferences for each segment. How do we do that? MITRE has created an open source patient consent policy management tool called Kairon Consents that is available for use now."[8]

Future Directions

The initial patient interface results will be elaborated and informally validated in FY13. MITRE will write papers to document the major technical results. The papers will be briefed to sponsors and the technical community. MITRE intends to connect to additional sponsors, notably VA (which is interested in creating an open-source consent system). For ONC, MITRE will create a paper that brings together all of the study's observations about the need for guidance in data segmentation and granular consent.

[8] Halamka, John. "Data Segmentation," Life as a Healthcare CIO. March 5, 2012. http://geekdoctor.blogspot.com/2012/03/data-segmentation.html: 2013

Equip Goal

Our portfolio also aims to **Equip** healthcare decision makers with the tools and information they need to set policy and ensure safe and high-quality care, accelerate research to move promising, safe, and effective new therapies and devices from "bench to bedside."

In furtherance of the Equip goal, the Privacy Testing in the Healthcare Environment project successfully mapped a set of generic system privacy requirements and associated privacy tests to privacy controls in National Institute of Standards and Technology (NIST) Special Publication 800-53, *Security and Privacy Controls for Federal Information Systems and Organizations*, Appendix J, *Privacy Control Catalog*. MITRE then identified which specific healthcare-oriented privacy requirements and associated tests match requirements in healthcare-related laws (e.g., HIPAA). Following MITRE engagement with NIST discussing the integration of the tests into NIST guidance, NIST is considering adding the privacy tests to NIST Special Publication 800-53A, *Guide for Assessing the Security Controls in Federal Information Systems*.

Project TranScript developed a new relationship collaboration with a major New England provider network, Partners Healthcare, gaining access to a data set of prescriptions annotated for the presence of inconsistency errors. The MITRE TranScript analysis of both the prescription data and the cost implications of the domain revealed that discrepancy errors during prescription order-entry likely constitute a $4.6 billion/year problem. The TranScript project has effectively developed machine learning and natural language processing techniques to automatically detect inconsistency errors and will be working toward refining their algorithms to transition them to industry and potentially the VHA to address patient safety issues related to electronic prescription order entry.

The EyesFirst project focused on validating the EyesFirst automatic optical coherence tomography screening system in collaboration with clinical partners at the New England Eye Center while continuing development of classifiers and a clinical decision support interface. One key accomplishment included evaluating the initial retinal thickness and hard exudate classifiers. The retinal thickness classifier work was presented at the 2012 Association for Research in Vision and Ophthalmology (ARVO) Conference. Another key accomplishment was developing a clinical decision support interface that would be used to evaluate the consistency of the automated declarations with clinicians' observations and to evaluate the utility of the automated algorithms for clinical decision support in both early detection and disease progress measurement of diabetes. Finally, EyesFirst conducted a successful pilot study using the EyesFirst system instantiated at the New England Eye Center to evaluate the hard exudate classifier.

Engineering and scientific obstacles were overcome, clearing the way for the Analytics for Rehabilitative Sensing (ARMS) prototype to get under way in concert with the National Rehabilitation Hospital, a Medstar hospital focused on civilian and veteran rehabilitation. The ARMS project developed wearable inertial sensors that log and transmit data to a nearby Smartphone. Currently in its second revision, the ARMS sensor captures a wider spectrum of data at higher fidelity than similar consumer devices and includes both internal logging and Bluetooth transmission capability. The ARMS sensor also developed and tested a classification

system capable of translating the ARMS sensor data stream into a time series of functional/nonfunctional data points. The distinction between functional use (lifting a spoon to eat breakfast) of a prosthetic and "nonfunctional" use (arm swinging) is key to effective rehabilitation of arm amputees and is today done in laboratory settings by human observers. ARMS has the potential to transform (and inform with objective data) this element of rehabilitation by allowing patients to be evaluated in natural settings. Extensive validation testing of the system took place at the National Rehabilitation Hospital. Results have been documented in a paper that will be submitted for publication to *IEEE Transactions on Biomedical Engineering*.

The Healthcare Fraud Fluorescent Marker project was successful in developing a prototype for not just generating fluorid identities for use in fighting Medicare fraud, but also developed an entity resolution framework to test these synthetic identities that has already transitioned to CMS's major fraud prevention efforts through our direct work program in provider screening and fraud prevention.

Transition activities for projects furthering the Equip goal included the following:

- The Privacy Testing in the Healthcare Environment project continued their engagement with the NIST, the HHS Office of the National Coordinator for Health Information Technology (ONC), and healthcare standards bodies to enable the inclusion of healthcare-related privacy requirements and associated tests in standards and guidance documents that are used by the healthcare industry, ultimately promoting broad adoption of privacy testing activities within the healthcare industry.
- EyesFirst deployed its automated retinal image analysis platform to the New England Eye Center to evaluate the system in a clinical environment.
- Both the National Rehabilitation Hospital (NRH) and Washington University have collaborated on the ARMS project experiments and plan to use the ARMS system in the future for a longitudinal study to further validate the classification algorithm for meaningful daily use of prosthetic arms—key to optimizing amputee rehabilitation.
- The Healthcare Fraud Fluorescent Marker project successfully transitioned the entity resolution framework built to validate synthetic identities to both MITRE's CMS Center for Program Integrity Lab and to the new CMS Fraud Prevention Partnership. The entity resolution framework developed was used to create the CoFraud data set, a comprehensive provider risk data set that is being actively used by CMS in both of its fraud prevention efforts: automated provider screening (APS) and the fraud prevention system (FPS).

Healthcare Analytics Roadmap

Principal Investigator: Kim Warren

Project Description

MITRE's healthcare sponsors (e.g., CMS, VA, FDA, ONC) have enormous needs for data analysis to improve decision making and generate innovations across a variety of areas. These innovations have a strong potential to "bend the curve" of healthcare costs for government agencies while preserving or enhancing quality and access to care. But the potential can be realized only if the organizations have the capacity to leverage their data and successfully apply the analytic results into their complex environments.

Research Objectives and Milestones

While there are many potential focus areas for relevant research in healthcare data analytics, the goal of this project was to identify the state of the practice, the state of the art, and the gap areas to inform MITRE investment in equipping decision makers.

FY12 Accomplishments

This project delivered a report that surveyed the state of the practice and state of the art in healthcare data analytics, identified gap areas, and proposed potential areas for MITRE investment. This report is currently used in multiple ways, including to:

- Guide investment decisions;
- Educate potential proposal teams about important gap areas;
- Provide a basis for understanding portfolio-level outcomes (as opposed to project-specific outcomes alone); and
- Identify areas in which to make strategic hires.

Transition Activities

The roadmap has been used internally to guide investments and engage potential researchers. Elements of the paper have been incorporated into presentations to our healthcare sponsors and their stakeholders to describe our overarching Equip goals, the value of data analytics in healthcare, and a high-level framework for classifying health data analytics challenges.

Future Directions

This report will be updated as part of the portfolio strategic planning for Health Transformation moving forward.

Privacy Testing in the Healthcare Environment

Principal Investigators: Julie McEwen and Julie Snyder

Project Description

During its second year exploring the value of privacy testing frameworks for healthcare, MITRE continued to work with sponsors to encourage them to design privacy protection into information systems as they are built. This is known as "Privacy by Design." The term "privacy testing" refers to specific system tests that are performed to ensure that privacy is implemented correctly in systems. This is an important step to ensure that systems appropriately protect Personally Identifiable Information (PII), and it is especially vital for systems that process large amounts of Protected Health Information (PHI) to reduce the likelihood of errors in care, ensure compliance to privacy regulations, and reduce the overall cost of errors in providing healthcare services. So privacy testing is *true* Privacy by Design. A privacy team at MITRE has engaged in groundbreaking work with a sponsor since FY11 to integrate privacy testing into their existing systems development process. However, there has not yet been a broader effort to address using privacy testing to verify that basic privacy controls are correctly implemented within the healthcare environment.

Including privacy testing in the system development process will:

- Provide better privacy protection for PHI because organizations will be able to more fully identify and address privacy issues;
- Embed privacy into the system development process;
- Address previous audit findings and possibly prevent future ones;
- Help organizations to potentially avoid privacy incidents;
- Directly validate proper implementation of privacy requirements for systems;
- Aid in identifying privacy issues prior to production, including those that may not have been apparent in the system design. This will also help reduce the cost of mitigating privacy issues; and
- Establish confidence with stakeholders and demonstrate due diligence.

Ultimately, including privacy tests in guidance and standards used by the healthcare industry will advance the state of privacy protection for PHI.

Research Objectives and Milestones

The objectives of this research are:

1. Identify a set of privacy requirements and associated healthcare-oriented privacy system tests based on privacy controls issued by NIST, the HIPAA Privacy Rule, and other federal requirements pertaining to specific types of sensitive health information; and
2. Identify other verification techniques for those privacy controls that cannot be verified by privacy system tests, such as code reviews and document reviews used to verify that privacy controls are implemented correctly.

FY12 Accomplishments

National Institute of Standards and Technology (NIST) risk management standards and tools are often referenced in the healthcare environment, so guidance for privacy controls issued by NIST was used as the foundation for this research. MITRE successfully mapped a set of generic system privacy requirements and associated privacy tests to privacy controls in NIST Special Publication 800-53, *Security and Privacy Controls for Federal Information Systems and Organizations*, Appendix J, *Privacy Control Catalog*. MITRE then identified which specific healthcare-oriented privacy requirements and associated tests matched requirements in healthcare-related laws (e.g., HIPAA) and guidance. MITRE engaged with NIST to discuss how to integrate the tests into NIST guidance. As a result of the MITRE engagement, NIST would like to add the privacy tests to NIST Special Publication 800-53A, *Guide for Assessing the Security Controls in Federal Information Systems*. NIST also agreed that the specific privacy tests that apply to the healthcare environment could possibly be placed in a healthcare privacy overlay so that it is easier for users in the healthcare environment to identify the privacy tests that apply to their systems.

Transition Activities

MITRE will engage again with NIST, as well as with ONC and healthcare standards bodies, to continue discussing the inclusion of the healthcare-related privacy requirements and associated tests in standards and guidance documents that are used by the healthcare industry. The ultimate goal is to promote broad adoption of privacy testing activities within the healthcare industry.

Future Directions

In FY13, MITRE will revise an existing Web-based privacy risk management tool developed at MITRE so that it can be used for privacy testing efforts in the healthcare environment. The objective of that work will be to make it easier for the healthcare industry to integrate privacy testing into its existing system testing processes.

TranScript: Detecting Discrepancies in Pharmaceutical Prescriptions

Principal Investigator: David Tresner-Kirsch

Project Description

TranScript will improve the safety and accessibility of prescription instructions through the application of automated language processing and artificial intelligence. Typically, medication instructions are conveyed between clinicians, pharmacists, and patients in natural language, which is difficult for automated systems to reason with. The MITRE team is designing and prototyping a novel system (TranScript) that can extract a structured knowledge representation from those natural language instructions and subsequently reason with or transform that representation. The TranScript design includes modules capable of checking for inconsisten-

cies in the extracted, structured prescription data against the natural language description and then generating clear and unambiguous text instructions. It is also designed to be controlled through a Web-based API for ease of integration with existing prescription order-entry client applications.

Research Objectives and Milestones

The project's objective is to prototype a system capable of analyzing natural language text in prescription instructions and automatically detecting a class of errors that exist as a side effect of the introduction of electronic prescribing systems. Descriptive research about these errors has been previously published,[9,10] but no work has been undertaken about using automated processes to detect and prevent the errors.

To address this gap, MITRE prototyped and tested several approaches to computational reasoning systems: a rule-based system, a statistical machine learning system, and a hybrid system:

- The rule-based system leverages work done for an earlier research project (Multimodal Medical Data Capture and Representation) about tagging and labeling medication-related concepts in text via regular expressions. The team adapted this system and added a post-process that parses the tagged text, attaching a semantic canonicalization to each. The team also prototyped a reasoning system that compares these values (extracted from free text) with existing structured metadata, and asserts the presence or absence of inconsistency.
- The statistical system applies the Online Balanced Winnow algorithm to gold-standard data (labeled for the presence and category of present discrepancies) to build a model capable of identifying similar discrepancies in unlabeled data. As a feature space, we used the cross-product of metadata field values with character- and word- n-gram substrings of the free text directions. The training framework leverages code previously developed by MITRE research.
- The hybrid system uses the same training framework as the statistical system but takes the output of the regular-expression concept tagger as additional input. In this case, rather than building features by pairing metadata fields with arbitrary substrings of the directions, the hybrid system uses the specific and categorized substrings found by the tagger.

The team evaluated intermediate implementations of these systems on a gold-standard development set and found that all three implementations identify discrepancies with nontrivial accuracy, showing that the approach adds value to the prescribing process.

---

[9] H. Singh, S. Mani, et al., "Prescription Errors and Outcomes Related to Inconsistent Information Transmitted Through Computerized Order Entry: A Prospective Study," *Archives of Internal Medicine*, Vol. 169, No. 10, pp. 982–989, May 25, 2009. Available: http://archinte.jamanetwork.com/article.aspx?articleid=773518.

[10] M. Palchuk, E. Fang, et al., "An unintended consequence of electronic prescriptions: prevalence and impact of internal discrepancies," *Journal of the American Medical Informatics Association*, Vol. 17, pp. 472–476, July 2010. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995649/.

FY12 Accomplishments

During project TranScript's first year, the team negotiated access to a data set of prescriptions from the Partners Healthcare network that were previously annotated for the presence of inconsistency errors.

The team performed an analysis of both the prescription data and the cost implications of the domain. The analysis revealed that discrepancy errors during prescription order-entry likely constitute a $4.6 billion/year problem.

In addition to developing the approaches mentioned in the previous section, the team also sequestered a set of test data (not yet inspected) on which evaluations will be run. In order to obtain a truly robust evaluation of the prototypes, it is necessary to perform some additional data stewardship. Thus, the team also developed annotation guidelines for adding gold-standard concept tags and is in the process of annotating that data set. Additionally, the team is in the process of adding discrepancy class annotations that were not covered by the original dataset received from Partners Healthcare network. The earlier annotation schema allowed only one error type to be assigned to any given prescription, whereas many records contain multiple types of errors concurrently. These missing, concurrent discrepancy labels function as noise in the current training and evaluation. For a true prototype deployment, it is necessary for these discrepancy-finding back-end systems to be capable of integration with front-end e-prescribing interfaces. MITRE has taken a first step toward deployment by implementing a Web-based API wrapper for the rule-based reasoning system.

Transition Activities

In FY12, the team's attention to transition focused primarily on building relationships with sponsors and other stakeholders. In particular, the team sought feedback and collaborative relationships with the Air Force Medical Service via the AFMS Chief Information Officer and Deputy Chief Medical Information Officer and with the Veterans Health Agency (VHA) via VHA Pharmacy Benefits Management and researchers at VHA Houston.

Future Directions

There are several goals for FY13:

- Integration: Pilot with a computerized order-entry prescribing system, either with a sponsor system or a vendor system;
- Research: Develop adaptable methods capable of providing higher value and of being applied to new domains. This research includes the investigation of unsupervised methods for training discrepancy-finding models on data without gold-standard labels; and
- Transition: Improve prototypes and release code to sponsors and community.

Further into the future, the team also intends to apply the TranScript methodology to other semistructured, semiredundant data types in the healthcare domain (e.g., patient encounters and billing).

Eyes First

Principal Investigator: Salim Semy

Project Description

Challenges in effectively treating chronic disease are early detection combined with appropriate treatment to reverse, arrest, or retard the disease process. Current diagnostics are expensive, invasive, specialized, and often inconclusive until the late stages of the disease. As a result, screening techniques are not routinely used, disease detection is significantly delayed, and treatment options are limited.

The EyesFirst system is an automated image analysis system to detect the early stages of multiple chronic diseases through retinal imaging.

Research Objectives and Milestones

The objective of this research is to develop and validate an open source platform (the EyesFirst system) for automated retinal image analysis to detect the early signs of multiple chronic diseases. This platform includes:

- A database of retinal images of healthy eyes and diseased eyes manifesting a variety of disease symptoms. Subsets of images in the database are used to train algorithms and quantify algorithm performance;
- Image feature extraction and statistical classification algorithms for automated disease detection; and
- A clinical decision support interface to support the use and evaluation of the automated algorithms in a clinical environment.

This platform has been co-engineered with SMEs at the New England Eye Center. The initial use of the platform is for diabetic retinopathy screening and management.

FY12 Accomplishments

In FY12, this research focused on validating the EyesFirst system in collaboration with our clinical partners at the New England Eye Center while continuing development of classifiers and clinical decision support interface. Key accomplishments include:

- Evaluating the initial retinal thickness and hard exudate classifiers. The retinal thickness classifier was presented at the 2012 Association for Research in Vision and Ophthalmology (ARVO) Conference;
- Developing clinical decision support interface to evaluate the consistency of the automated declarations with clinicians' observations and to evaluate the utility of the automated algorithms for clinical decision support; and
- Conducting an initial pilot study using the EyesFirst system instantiated at the New England Eye Center to evaluate the hard exudate classifier.

Transition Activities

The initial transition was focused on deploying the EyesFirst platform to the New England Eye Center to evaluate the system in a clinical environment. This initial transition helped inform the research team of considerations in use of the system in a clinical environment and showcased initial validation for future pilots and transition opportunities.

In addition, the team initiated discussions with multiple federal agencies to discuss potential sponsor transition opportunities. Specific sponsors include the Office of National Coordinator for Health IT on its Ocular Imaging Challenge and the VHA on its Tele-Retinal Imaging Program.

Finally, the results of the classifier development, specifically the retinal thickness classifier, were presented at the 2012 ARVO Annual Meeting.

Future Directions

In FY13, the team will determine the viability of using an optimal coherence tomography-(OCT) based retinal imaging system and automated feature extraction and classification algorithms for diabetic retinopathy screening and management. The team will be completing development of two additional algorithms started in FY12 (retinal fluid volume quantification and fundus photo and OCT registration). These algorithms, plus algorithms developed previously for retinal thickness analysis and hard exudate detection, will be used in a pilot at the New England Eye Center to screen patients for diabetic retinopathy.

A second area of focus in FY13 will be pursuing direct engagement with potential sponsors for additional piloting and transition opportunities. Furthermore, the team will continue to broadly share lessons learned from this research, including technical publications providing the results of our classifier development and recommendations on implementing DICOM to support better interoperability.

Analytics for Rehabilitative Motion Sensing (ARMS)

Principal Investigator: Elaine Bochniewicz

Project Description

Patient rehabilitation for stroke victims and recent amputees is extremely expensive. One of the primary reasons for the high cost involves the labor- and equipment-intensive, slow, and sometimes outright subjective process of evaluating patient recovery. Making the process cheap, fast, accurate, and precise would bend the rehabilitation cost curve significantly downward (for the VA, federal and private-sector payers, and providers). A MITRE research team believes this can be accomplished by analyzing data from inexpensive sensors mounted on the patient. Classifying usage of an upper limb into functional and nonfunctional categories in this manner would give therapists an outcome analysis tool that would allow for tailored therapeutic regimens, remote monitoring of patient recovery, and objective validation of existing therapeutic techniques. Collecting a constant stream of bio-kinetic data, mining it to

extract quantitative measures of recovery, and finally presenting those measures to a therapist in a useful form, is also an exercise in Big Data Analytics that MITRE is uniquely qualified to solve.

Research Objectives and Milestones

The objective of this research is to develop an inexpensive method of persistently evaluating how much and how often a patient's upper limb is being utilized. Current methods for estimating this data are either very expensive or very inaccurate. However, miniaturized sensors, modern communications infrastructure, and powerful wearable computers may present all the components needed for a system that can collect this data cheaply, accurately, and reliably.

The project's approach focused on real-time classification of inertial data taken from a wrist-mounted Inertial Measurement Unit (IMU). The anticipated result is a system that provides arm usage statistics as a patient goes through daily routines outside of a healthcare facility, rather than requiring lengthy testing in a laboratory environment.

FY12 Accomplishments

During this project, a number of engineering and scientific problems were solved to prototype the ARMS system:

- Develop wearable, inertial sensors that log and transmit data to a nearby smartphone. Currently, on its second revision, the ARMS sensor captures a wider spectrum of data at higher fidelity than contemporary devices used in the field, and includes both internal logging and Bluetooth transmission capability. Future revisions will focus on extending battery life.
- Develop and test a classification system capable of translating the ARMS sensor data stream into a time series of functional/nonfunctional data points. Extensive validation testing of the system took place at National Rehabilitation Hospital (NRH). Results have been written into a paper that will be submitted to *IEEE Transactions on Biomedical Engineering*.
- Develop mobile and server applications to collect data streaming from the ARMS sensor, run the classification model on a mobile device, and stream results in real-time to an aggregating service which then visualizes the data and makes it available to healthcare workers.

Transition Activities

Over the course of the ARMS project, a research agreement between NRH and MITRE has allowed joint research to take place. Both NRH and Washington University plan to use the ARMS system in the future for a longitudinal study to further validation of the classification algorithm. Finally, the results of the classifier research have been documented in a journal paper that will be submitted for peer review and available to the public.

Future Directions

Plans for FY13 involve improving the ARMS system in several ways:

- *Expanded validation:* Collecting data from, and testing classification methods on, a larger group of both impaired and unimpaired patients. Collecting and annotating data collected "in the wild" from patients outside a laboratory setting;
- *Improved hardware:* The ARMS sensor will undergo a third revision to decrease size, increase data quality, and improve battery life;
- *Model development:* Improving the classification model toward a system that does not require patient-specific calibration; and
- *Integration:* Streaming of results to a patient medial record access tool, such as MedCafe.

Laser-Enabled Gait Recording System (LEGS) Walter Reed Gait Collaboration

Principal Investigators: Michael Fine and Adam McLeod

Project Description

MITRE will collect a large corpus of human kinematic data at the Walter Reed National Military Medical Center that demonstrates natural human ambulation with an emphasis on *transitional* and *nonmodal* movement. If *modal* human movement is defined as periodic ambulation (walking, running, etc.) plus standing (which can be thought of as periodic movement having a velocity of zero and an infinite period), then *transitional* and *nonmodal* movement can be thought of as the set of natural movements that interrupt or happen in between *modal* activities:

- *Transitional movement*: Movement that occurs when changing gaits, e.g., from walking to running or vice versa, or when transitioning between standing and a periodic gait.
- *Nonmodal movement*: Periodic movement that occurs during natural ambulation or standing. This includes changing direction while walking or running, translating the center of mass less than one stride in any direction while standing, or turning in place to look behind oneself.

Research Objectives and Milestones

MITRE developed a system to programmatically guide amputees through the transitional and nonmodal movements that occur during walking, running, and standing on a flat surface by projecting a laser dot onto the floor. The dot will be generated by a laser mounted to a computer-controlled pan/tilt mount and will trace curved paths spanning the motion capture system area (to encourage periodic, transitional, and nonmodal movement) and semicircular arcs around the participant (to encourage aperiodic, nonmodal movement).

Paths for the laser will be generated by the Laser-Enabled Gait recording System (LEGS). This tool is comprised of three main subcomponents:

1. An experiment designer generates paths within the motion capture environment in a constrained random fashion. The user specifies what and how many tasks are required for an experimental session and the experiment designer generates laser paths based on this information, shuffling the order of the tasks and randomizing their location within the motion capture volume.
2. A calibration utility guides the user through activities that give the system enough information to accurately place laser paths.
3. An experiment runner instructs two pan/tilt units to trace the generated laser paths along the floor.

FY12 Accomplishments

During FY12, both the hardware and software components of the GAET system were developed. Data collection is scheduled to take place at Walter Reed in early FY13.

Transition Activities

Near the end of FY12, the prototype LEGS system was installed in the Walter Reed motion capture facility.

Future Directions

Data collection at Walter Reed is expected in FY13, as well as a Memorandum of Understanding between Walter Reed and MITRE, which will make future collaboration much easier. Co-authorship on papers utilizing the LEGS system and possible collaboration on future grants are also possible paths for further establishing MITRE's relationship with Walter Reed.

Making Predictions from Examining Health Data

Principal Investigator: Alexander Yeh

Project Description

The U.S. government has a strong incentive to control medical costs. Detecting medical conditions early, or even better, preventing them, can keep such conditions from becoming more serious and costly. As part of the record keeping and payment processes, organizations like CMS and the VA receive large amounts of healthcare data. This data has secondary uses, including the analysis of the data in order to build systems for the automatic detection or predication of medical conditions. This project is about using patient clinical data to build systems that either detect when an individual patient has a condition or predict when (s)he is at risk of getting that condition and medicine that can specifically affect the prediction.

Research Objectives and Milestones

The Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC2) intensive care unit (ICU) data set, gathered at Beth Israel Deaconess Medical Center and later de-identified and managed by Massachusetts Institute of Technology (MIT), is a rich data set that includes both

waveform and nonwaveform data. Such data would allow MITRE to combine various types of both waveform and nonwaveform medical information, such as blood pressure waveforms, what medications are given, etc. An initial task to test is the ability to use information in MIMIC2 to detect hypovolemia (low blood volume) in ICU patients. Hypovolemia is present in some forms of shock and is a common cause of death in healthcare settings. We will build an interactive annotation aid for MIMIC2 and clean the data set so that medical experts can determine and annotate when patients in the data set have hypovolemia.

We are using MIMIC2 to develop and compare several hypovolemia detection systems:

- A system using waveform data alone will be developed by extending work on variability measures that medical research have found to increase in the arterial blood pressure and plethysmogram waveforms as blood volume decreases.
- A system to use nonwaveform data alone will be developed using Support Vector Machines, a machine-learning technique. The system will use the nonwaveform data as inputs together with the output and intermediate features of existing systems like SAPS and SOFA that use nonwaveform data to estimate general patient health.
- A system to use both waveform and nonwaveform data together will be developed by feeding the waveform data alone system's output and intermediate features as additional input to the system that already handles nonwaveform data input. The team is also working on interactive data annotation aids and filtering/cleaning-up the MIMIC2 data set (both feature values and waveforms).

FY12 Accomplishments

The project acquired and stored the MIMIC2 data set. The database storing the feature-value data has been cleaned up to make it more useful for this and other projects. The cleanup added indices, views, and tables to increase the database performance and simplify the queries that will be written. In addition, the cleanup dealt with mapping between many related dictionary codes. For example, there are 165 different yet similar codes for "urine," including slightly different spellings, units, etc.

The MIMIC2 waveform data was determined to have sections of malformed data. For the waveform of interest, arterial blood pressure, causes of bad data include disconnecting the sensor line, flushing the line (to prevent blood clots), and injecting fluids into the line. Routines were built to automatically detect and remove malformed data in this waveform. The measurement methods developed in other studies were extended by handling malformed data and automating the measurement methods for longer-term monitoring. In addition, a larger variety of patients produced a larger variety of waveform patterns. Other studies considered only patients with breathing controlled by respirators. In addition to respirator patients, MIMIC2 includes patients breathing on their own with more varied breathing patterns.

A preliminary version of an annotation interface has been built. The team has started manual annotation of the data set to learn where the annotation process can be improved.

Transition Activities

Work is still in a preliminary stage. A possible future transition activity is to give our MIMIC2 clinical (feature-value) data cleanup back to the MIMIC2 administrators at MIT for broader dissemination to the research community. Projected transitions for algorithms are planned for the clinical and research community based upon results.

Future Directions

After the initial hypovolemia detection experiments using waveforms only, feature values only, and combined waveform and feature-value information, the MITRE team will improve on the initial system performance by:

- Iteratively analyzing results, proposing and testing improvements; and
- Interactively generating more training data by iteratively:
  - Running the initial system on unmarked data; and
  - Manually repairing the system output to generate more training data.

Depending on the experience with the manual hypovolemia annotation, one direction would be to further extend the system using MIMIC2 data to detect the cause of hypovolemia and its opposite, hypervolemia, a condition in which too much blood volume is present in the body. Another direction would be to search for the availability of other data sets, especially already annotated ones.

Large-Scale Data Analytics for Medical Records

Principal Investigator: Zohreh Nazeri

Project Description

The growing availability of electronic health data provides new opportunities for analysis and discovery of health safety, health claim fraud, and other issues. More and more health organizations launch initiatives to seek scalable analytical solutions. The Foundation for the National Institutes of Health (FNIH) has launched the Observational Medical Outcomes Partnership (OMOP) to research and identify the most reliable methods for analyzing huge volumes of data drawn from heterogeneous sources. However, key information could be lost if the data is analyzed in small subsets, separately. To take advantage of the data as a whole and to perform holistic analyses, new analytical approaches are needed. This project establishes a new approach to analyzing observational health data using the OMOP simulation made available to OMOP researchers.

Research Objectives and Milestones

Medical records could be transformed into multiple one-dimensional signatures and/or multidimensional image-like representations, and the images analyzed for patterns of interest, anomalies, and other issues of concern. The colors and shades of pixels in the images would represent levels of presence (or lack of presence) of various values in the data. For example,

observational data in insurance claims and medical records for each patient would be transformed into an image representing the patient's exposure to different drugs at different intervals over a period of time. It would then be possible to look for patterns of concurrent drug exposures and effectiveness of the drugs as well as any adverse effects. With this approach, the potential collective impact of various data parameters could be captured without losing information. This research is aimed at developing a method for transforming the data elements into signatures or image pixels and analyzing the transformed data.

FY12 Accomplishments

In FY12, the MITRE team used simulated observational medical records provided by the OMOP. We established a process for transforming the observational data records into one image for each patient. We used the first occurrence of a targeted medical condition as the reference date for each patient and used this date as a reference point. Images for all patients were normalized by considering a time window in each image, built around the reference point, and lining up the images according to this point. MITRE then performed spatio-temporal, cross-correlation clustering, and used the heatmap approach to analyze the images. In the heatmap approach, a matrix of drug-condition pairs was generated where rows represented medical conditions and columns represented the drugs the patients were exposed to. Any particular entry corresponded to a condition-drug pairing, and its magnitude represented a degree of confidence. Evaluation of the results, using the ground truth matrix, provided with the initial set of simulated data, showed promising results. The initial set of data did not include actual names of the coded drugs and conditions. Further discovery and evaluation of the findings, using the domain knowledge, will be possible when using the new set of data that is improved and includes the drug and condition names. The team plans to use this new data in future activities.

Transition Activities

This work will demonstrate the effectiveness of the developed data analytic approach as a solution for effective analysis of observational health data. The methodology developed in this research will enable a proactive approach to identifying safety issues in observational data.

Future Directions

Next steps include exchanging ideas and sharing the work with the FNIH, the OMOP, and the rest of the healthcare research community. Working on this problem with a group of researchers in the medical outcomes community will provide MITRE with a valuable opportunity to conduct peer-reviewed research in an area of vital importance to the public.

Automating Fact Extraction from Medical Records

Principal Investigator: Dr. Cheryl Clark

Project Description

Critical medical information is buried in the free text of electronic medical records. Current medical extraction systems identify certain medical concepts well, but serious gaps exist in these technologies that limit their ability to extract the relevant information. The MITRE team is developing algorithms that analyze linguistic context to determine the relevance of medical concepts mentioned in clinical reports. In previous years, MITRE focused on the detection of negation and uncertainty, building an assertion status classification system that could determine whether a medical problem mentioned in clinical text is present in the patient, absent, uncertain, hypothetical, present only under certain conditions, or associated with someone other than the patient. This system tied for first place in the medical problem assertion task of the Fourth i2b2/VA Natural Language Processing Challenge. In FY12, MITRE further developed the assertion status classification system to expand its functionality and enable its integration with a widely used open source clinical information extraction system.

Research Objectives and Milestones

The research objective is to build a system for automated fact extraction from medical free text that leverages the functionality of existing medical concept extraction systems and supplies missing information that is crucial to understanding meaning. Such a system should provide the kind of information needed to support the secondary use of EHR data in applications such as phenotyping and patient cohort selection, comparative effectiveness studies, and discovery of various disease-treatment relationships. In FY12, the goals included:

- Expanding the domain of assertion status assignment beyond medical problems to include medications, procedures, and other medical concepts;
- Providing more precise assertion information; and
- Developing algorithms to identify change in medical concept status (e.g., discontinuation of a medication, or an increase or decrease in the amount of a medication).

To ensure the research addresses real clinical research problems, MITRE has worked with several collaborators and sponsors, including:

- Massachusetts Veterans Epidemiology Research and Information Center at Boston VA Medical Center;
- CHIR, a consortium of eight VA sites across the country;
- i2b2, a National Institutes of Health-funded National Center for Biomedical Computing, Partners HealthCare; and
- SHARP4, a SHARP research initiative sponsored by ONC. Area 4 is a collaboration of 14 academic and industry partners aligned to develop tools and resources that influence and extend secondary uses of clinical data.

FY12 Accomplishments

In FY12, the team expanded the assertion status system's domain of concept interpretation from just medical problems to include medications, procedures, and anatomical sites. This expansion enables a concept extraction system to distinguish, for example, medications the patient is taking from medications the patient is no longer taking or might be taking in the future, or to distinguish treatments a patient has received from treatments the patient has declined or refused.

The team made architectural changes to the assertion status system so that it assigns independent assertion values for multiple assertion attributes rather than a single assertion category. This capability will allow a concept extraction system to provide a more accurate and complete representation of the meaning of clinical text.

The team enhanced the document zoner so that it now identifies not only document sections but also subsections. The section location of a medical concept can be the sole indicator of whether a medical problem is current, part of a patient's medical history, or part of the patient's family history. Section and subsection membership can also serve to disambiguate otherwise ambiguous abbreviations. For example, "CVA" typically means "cardiovascular accident" (i.e., "stroke") when it occurs in a Medical History section of a clinical report, but "costovertebral angle" when it occurs in the Physical Exam section.

The team made presentations to the following audiences:

- MITRE Center for Transforming Health (CTH) Health and Population Health Advisory Board, February 8;
- MITRE CTH leadership, March 26;
- SHARP4 Face-to-Face Conference, June 11–12; and
- CHIR teleconference, September 18.

Transition Activities

FY12 transition work focused on modifying the MITRE Assertion Status Tool for Interpreting Facts (MASTIF) to enable integration into clinical NLP pipelines used by our sponsors and collaborators. MASTIF was configured as a UIMA analysis engine for integration into the clinical Text Analysis and Knowledge Extraction System (cTAKES), which is an open source clinical extraction engine developed by Mayo Clinic. The output of the assertion engine was made to conform to SHARP4 NLP requirements so that it could be integrated in the SHARP4 cTAKES-based clinical NLP pipeline:

- MASTIF made publicly available with an open source license. The source code can be accessed from SourceForge.
- The assertion status code was integrated into the SHARP4 Clinical NLP pipeline.
- The assertion status code was integrated into the cTAKES and released as part of cTAKES 2.5.

- The team donated the assertion status code to the Apache Software Foundation as a component of cTAKES, which has become an Apache incubator project. The Apache release serves a strategy to make MITRE software accessible for adoption as part of an open source solution or via commercial adoption.

Future Directions

A major accomplishment for this research was integration into cTAKES and into the SHARP NLP pipeline. Future directions for this project include the following activities:

- Extension of assertion status coverage to relations between concepts;
- Delivery of new versions of MASTIF for incorporation in new releases of cTAKES;
- Incorporation of this technology in sponsor pipelines (CMS, VHA, other) and/or within a commercial product; and
- Evaluation of the accuracy of the system using available annotated data sets and evaluation of its impact on clinical applications in which MITRE expects to embed it, such as automatic cohort selection systems.

Informatics for Integrating Biology and the Bedside (i2b2) Natural Language Processing (NLP) Challenge

Principal Investigator: Lynette Hirschman

Project Description

Challenge evaluations play a critical role in defining research objectives, building a community, developing infrastructure and resources, and advancing the state of the art. The i2b2 challenge evaluations, begun in 2006, invigorated the field of medical NLP by making available realistic (de-identified) patient records, including training sets and "gold standard" data for evaluation to support research about the secondary use of electronic medical records applied to translational medicine. These evaluations are run by the i2b2 National Center for Biomedical Computing, a multi-institution collaboration run by Professor Isaac Kohane, Harvard Medical School, in cooperation with Partners Healthcare. The MITRE team has been part of the organizing committee for the i2b2 Challenge Evaluations since 2011 and has organized related challenge evaluations in biomedical informatics to drive the research community toward development of sponsor-relevant capabilities in text mining and NLP.

Research Objectives and Milestones

The project goals are:

- Participate in the definition of the Challenge Tasks and the application of MITRE technologies and expertise in the areas of de-identification and extraction of negation-uncertainty and temporal information. MITRE was a co-organizer for the conference evaluation task for 2011 and is co-organizer for the 2012 evaluation on extraction of temporal relations in clinical records.

- Influence the direction of BioNLP evaluation and infrastructure, including BioCreative (MITRE is a co-organizer with NSF funding), BioNLP evaluations, and resource preparation.
- Serve as liaison and catalyst to bring together the biomedical information and clinical informatics communities.

FY12 Accomplishments

MITRE participants (John Aberdeen, Cheryl Clark, Lynette Hirschman) supported the 2011 i2b2 Challenge Evaluation and participated in the workshop at AMIA in November 2011.

MITRE advised on various aspects of the 2012 temporal annotation task and will attend the 2012 i2b2 Challenge Evaluation Workshop (Q1 FY13). These activities are synergistic with the research being performed under Clark and Aberdeen's Automated Fact Extraction MSR project. For this task, MITRE worked with the co-organizers to define a medically motivated "application task" that would make use of the temporal annotations. Specifically, we explored a possible medication time alignment task: "binning" patient medications into preadmission, during hospitalization, and posthospitalization periods using the 2012 temporal annotations. This task was not pursued for a variety of reasons, but there was consensus to begin planning for a clearly defined "capstone" medical application task for November 2013. In concert with MITRE's recommendation, the first steps were to set up an Advisory Committee, including three external MD/PhD advisors to define the task.

Hirschman was an invited speaker at the NLM workshop (April 23–24) on "Natural Language Processing: State of the Art, Future Directions and Applications for Enhancing Clinical Decision-Making."

Hirschman continues as co-organizer of the BioCreative Challenge Evaluations (Critical Assessment for Information Extraction in Biology); activities this year (under NSF funding) include co-organizing a BioCreative workshop on text mining for the curation workflow as well as planning for BioCreative IV (April 2013). Future BioCreative evaluations will start to focus on translational medicine applications.

Transition Activities

MITRE is now a driving force in defining the medically relevant application task for i2b2 2013, based on experience in gathering community input for BioCreative Challenge Evaluations. The inspiration for an Advisory Group comes directly from prior experience in BioCreative.

Future Directions

- Attend i2b2 Workshop at AMIA 2012; and
- Oversee the organization of a medical application task for i2b2 2013.

Unlocking Patient Data for Translational Medicine

Principal Investigators: Lynette Hirschman and John Aberdeen

Project Description

Electronic medical records provide a rich set of time-stamped observations about individual patients. This project will "unlock the patient record" by linking computable information extracted from patient records to the research literature and to biomedical databases. Patient clinical data, coupled with increasingly available genetic data, support a growing range of translational applications, including identification of genetic components of disease (genotype/phenotype correlations), genetically based response to drugs (pharmacogenomics), and biologically based models of disease. The team is leveraging its expertise in curation of the biomedical literature and medical fact extraction to support these translational applications. The team is also collaborating with medical providers for data access and driving use cases, including the VHA, Harvard Medical School, Children's Hospital Boston, Mayo Clinic, the ONC SHARP4 Consortium, Group Health Cooperative, Vanderbilt University, and Cincinnati Children's Hospital.

Research Objectives and Milestones

The project goal is to support end-to-end translational medicine applications of clinical relevance to MITRE sponsors, particularly the VHA and HHS ONC.

Translational applications involve a multistep process. The first step is to identify patients based on clinically observable characteristics extracted from patient medical records. The second step is to mine characteristics of these patients for correlations (e.g., between disease and genetic variation, or between adverse drug reaction and patient genotype). The third step is to vet candidate correlations against the research literature and curated biomedical databases to determine plausibility of the correlation and to understand possible mechanisms of disease or drug interaction. These correlations can be used to elucidate new mechanisms of disease or to determine appropriate therapies for specific patients with a particular genotype and phenotype.

A key challenge is to make accessible (and computable) the massive volume of new experimental data appearing in the literature. This has been the focus in year one, building on experience in curation of the biomedical literature from the NSF-funded BioCreative project. The team is now exploring issues related to cost of curation and capture of computable information. The team has simultaneously been building partnerships to link this information to information extracted from patient records, leveraging work from the Automated Medical Fact Extraction MSR project.

FY12 Accomplishments

Collaborating with researchers from Harvard Medical School and University of Maryland, Baltimore County (UMBC), MITRE developed a prototype that uses crowdsourcing for curation of novel biomedical relations reported in the literature, such as gene-mutation-disease rela-

tions. This prototype was driven by the need for an up-to-date open repository of information on human genetic variants and phenotypes to support translational applications. Typically, manual curation of gene-mutation-disease relations is labor-intensive and requires expert curators, who are in short supply.

The initial prototype focused on gene-mutation relations. It relies on automated entity extraction for genes and mutations and uses crowdsourcing (via Amazon Mechanical Turk) to elicit human judgments on the correctness of candidate gene-mutation relations. Each candidate gene-mutation pair constitutes a "hit"; mentions of "gene" and "mutation" are highlighted in the abstract for presentation. Five "turkers[11]" judged each "hit" with results returned in 36 hours. Initial weighted results evaluated against a gold standard of expert-curated gene-mutation relations achieved 85 percent accuracy; the best turker achieved an accuracy of almost 95 percent. Cost averaged a little over 50 cents per abstract (per reviewer).

These results were presented at the Data Integration in Life Sciences Conference and published in the conference proceedings. The team is now studying the complex trade-offs in this space, including expertise (and availability) of curators/annotators; the need for curator/annotator training to achieve a desired level of accuracy; the intended applications of the curated/annotated data, and finally, the value of the curated/annotated data throughout its life cycle. This work was presented in an invited talk on "Cost of Curation" at the upcoming integrating Data for Analysis, Anonymization, and Sharing Annotation Workshop at University of California, San Diego (UCSD), held on September 29, 2012.

Transition Activities

- Contribution to Eastern Michigan University (EMU) source code: MITRE will contribute its modifications to a public code repository for EMU, including mention identification;
- Refinement of the EMU "gold standard" gene-mutation-disease database: MITRE made significant contributions to the refinement of the expert-curated database of gene-mutation-disease relations developed at UMBC; and
- National Center for Biotechnology Information (NCBI) collaboration: NCBI researchers would like to use the MITRE prototype software for gene-disease-mutation extraction to support existing NCBI databases.

Future Directions

- Repeat the gene-mutation experiment to explore cost/accuracy tradeoffs through prequalifying turkers and more sophisticated methods of aggregating turker results.
- Extend the system to extract gene-mutation-disease relations in partnership with NCBI; explore extraction of gene/drug/disease relations for pharmacology and pharmacogenomics experiments.

[11] "Turkers" are 'workers' that perform requested tasks via Amazon Merchanical Turk Crowdsourcing Internet Platform.

Privacy-Preserving Data Mining (PPDM)

Principal Investigator: James Davidson

Project Description

Several of MITRE's sponsors have a desire to "bend the cost curve" in healthcare. Toward that end, agencies like CMS are interested in outcome analysis—identifying treatments and interventions that have a high likelihood of resulting in positive outcomes. Identifying which treatments are most effective is a classic big data analytics problem, one that CMS is beginning to pursue based on claims and quality reporting data. However, for such a data mining operation to be maximally successful, CMS needs access to a richer set of information. Unfortunately, protected health information is distributed across a large number of provider organizations. No single source contains all of the information about a single patient, nor does CMS have the authority to collect microdata: to do so would be a serious privacy violation. Moreover, ONC is encouraging organizations like CMS to conduct data analyses without collecting microdata.

To address this complication, researchers have recently developed privacy-preserving data mining algorithms that reveal only the result of the data mining operation. The underlying protected data are not revealed. Thus, the goal of this exploratory project is to determine the extent to which a particular privacy-preserving data mining algorithm performs on a specific outcome analysis problem in order to better understand the applicability of privacy-preserving data mining in this domain.

Research Objectives and Milestones

The primary objectives of this research were to identify CMS areas that would benefit from PPDM; assess the extent to which existing techniques directly meet CMS's needs; and identify gaps that MITRE needs to close so that CMS can conduct effective outcome analyses.

The team's approach began with performing a complete survey of existing privacy-preserving data mining techniques. The team then conducted a proof-of-concept analysis requiring privacy-preserving data aggregation (a component of most data mining algorithms) to demonstrate the potential of privacy-preserving techniques to sponsors. In parallel, the team interviewed stakeholders to identify which CMS areas would benefit from PPDM. The analysis indicated that Clinical Quality Measures (CQM) was a likely candidate, so the team devoted its investigation into understanding CQM. Thorough investigation included discussions with SMEs to identify their data mining needs and privacy constraints.

Finally, the team generated two reports. The first report was a rich and detailed description of current PPDM techniques. The second report outlined the current CQM framework and explained why there is a lack of applicability of PPDM to the existing CQM framework. This second report concludes with an analysis of alternate frameworks that may be implemented in the future and how PPDM may provide value if such frameworks are eventually implemented.

FY12 Accomplishments

The accomplishments for FY12 were primarily gaining expertise in PPDM and CQM. The MIP team is now well-versed in PPDM. The team has read more than 50 papers, manually traced 16 algorithms, and presented details of 13 algorithms. This project also investigated CQM, and the team now has a deep understanding of CMS's current framework and has identified numerous enhancements to the CQM framework, which includes integration of PPDM techniques. Over 14 SMEs were extensively interviewed to learn about privacy and security requirements in the sponsor space. The research team probed SMEs for details about analytic methods being performed (or desired) by CMS.

Transition Activities

The lack of applicability of PPDM to CQM led the team to seek to transition the significant knowledge gained to other MITRE experts researching low disclosure methods. To this end, the team organized a Brown Bag between interested parties to further discussion on this topic. Team members also met with various internal MITRE health-related projects and other related MIP projects investigating privacy and security to advise them on the potential benefits of PPDM to their research.

Future Directions

Team members will continue interaction with the fraud prevention partnership and other CMS "trusted third-party" projects to gain insight and guidance about PPDM and its potential value in semitrusted environments.

DrugOrNot

Principal Investigator: Erika Darling

Project Description

Robust drug supply is the lifeline to America's healthcare system. FDA defines drug shortages as "a situation in which the total supply of all clinically interchangeable versions of an FDA-regulated drug is inadequate to meet the current or projected demand at the patient level."[12] The drug shortages doctors and pharmacists are experiencing every day are a symptom of a larger, systemic problem that poses a substantial threat to the lives of Americans and the national emergency preparedness of the country. It is a problem that shows warning signs that it is still growing despite the substantial measures and tools the FDA currently uses to address drug shortages. "Some recent shortages have involved drugs for life-threatening conditions and, in some cases, the product in shortage has been the only product available for the

[12] Center for Drug Evaluation and Research, Food and Drug Administration, *Manual of Policies and Procedures (MAPP)* 6003.1. Available: http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ManualofPoliciesProcedures/ucm079936.pdf.

patient's condition. While most drugs do not experience shortages, this is a significant public health problem, one that deserves the concerted attention of government and industry."[13]

Research Objectives and Milestones

Our objective was to understand why drug shortages increased for the fifth straight year in 2011 despite increased measures taken to prevent them. Specifically, we wanted to understand the political, operational, economic, and technical (POET) barriers to eliminating drug shortages.

FY12 Accomplishments

The team conducted an environmental scan of previously published reports and literature noting in particular the Government Accountability Office (GAO) Report published in November 2011, "A Review of FDA's Approach to Medical Products and Shortages" published October 31, 2011, and the HHS Assistant Secretary for Planning and Evaluation report "Economic Analysis of the Causes of Drug Shortages" published in October 2011.

During March 2012, the team then conducted interviews with government personnel, drug manufacturers, pharmacy benefit managers, and healthcare providers about their opinions of the POET barriers to eliminating drug shortages. While many are aware of the POET barriers within the U.S. prescription drug market, it appears possible that the dynamics between such barriers were not fully comprehended. This conclusion caused us to utilize the method of System Dynamics to better understand behaviors within the market and perhaps reveal the best paths to prevent drug shortages. System Dynamics modeling provides a formal approach for understanding complexity and nonlinearity within systems.

Employing System Dynamics modeling, the team gained a much better understanding of the causes of drug shortages and developed recommendations to prevent them. One recommendation of this study is to perform modeling and simulation to explore changes government could make to strike the right balance between enough financial gains for the market to be attractive to manufacturers and still enough regulation to ensure drug safety. Another recommendation is to utilize data analytics in the near term to better anticipate shortages.

Transition Activities

The internal MITRE report informed dialogue with the FDA sponsor and its stakeholders.

Future Directions

With new legislation driving better information exchange, MITRE will watch the measures of performance.

[13] "A Review of FDA's Approach to Medical Product Shortages," Oct. 31, 2011. Available: www.fda.gov/DrugShortageReport.

Align Goal

The portfolio has the goal to **Align** incentives, measures of quality, and payment systems for a sustainable national healthcare system.

In furtherance of the Align goal, the Healthcare Technology Investment Modeling project developed, applied, and extended data-analytic methods in MITRE's Healthcare Services Modeling Independent Research and Development (IR&D) to advance knowledge and know-how. This work will be targeted at specific payment model innovation programs as a robust tool in our MITRE toolkit to support data-driven decisions by CMS and its stakeholders.

Healthcare Technology Investment Modeling

Principal Investigators: Kristin Lee and Ken Hoffman

Project Description

The U.S. healthcare enterprise encompasses public and private entities with significant government involvement in policies and programs that have widespread interactions and effects at all levels of the system—a complex adaptive system. Improved healthcare service processes and supporting technologies are a major contributor to improved medical outcomes and effectiveness. At the same time, they are a major factor in the cost of healthcare services and were the focus of this research.

Programs and policies have impacts on healthcare services at multiple scales. Hybrid models are required to represent and analyze various aspects and impacts of new delivery methods and technologies on health services to affect outcomes and costs. Diabetes management is a major challenge and provides a robust demonstration case because of the growing costs and the outcomes in the progression of unmanaged diabetes to serious medical conditions, including circulatory and cardiac conditions. The hybrid modeling structure includes agent-based, systems dynamics, discrete event, and economic analysis methods.

Research Objectives and Milestones

Provide data-analytic approaches and methods to:

- Represent healthcare service activities patient to outcome as a data-analytic architecture—the Reference Health Services System (RHSS);
- Incorporate healthcare investments (in preventive services and medical technology) explicitly in healthcare modeling and economic analyses; and
- Demonstrate the data-analytic RHSS framework for diabetes management.

This research builds on MITRE's systems engineering capabilities that have been successful in activity-based analytics for services and in significantly improving the characterization and treatment of technologies in existing models that now lack these details. The approach has

been demonstrated for a wide range of infrastructure technologies and for selected healthcare processes. The multiscale modeling framework applied incorporates essential health services with process and technology detail across 10 health sectors of the U.S. economy using the RHSS framework. The research draws upon existing cost data and projections of healthcare technologies from published sources (Harvard-Kennedy School of Government, MIT, Brookings) and applies this information within an economic modeling framework in current use by HHS.

FY12 Accomplishments

Developed, applied and extended data-analytic methods in MITRE's Healthcare Services Modeling IR&D, including:

- Published a multiscale hybrid modeling capability that can be linked with an economic policy model (INFORUM-LIFT) for policy analysis and research and development planning;
- Described proposed healthcare programs and supporting technologies and processes in the RHSS, including telemedicine and the electronic medical record;
- Documented cost elements of current and selected emergent services and technologies within the U.S. healthcare system for information management, diagnostic, and therapeutic purposes;
- Compared the costs of new methods and technologies with those being replaced;
- Applied the RHSS and LIFT to analyze potential impacts of technologies;
- Formulated a healthcare services data-analytic framework based on the RHSS;
- Mapped budget elements for VA and CMS into the RHSS to support planning, analysis, and evaluation of programs; and
- Presented our modeling framework at the International Systems Engineering Conference.

Transition Activities

The project extended MITRE capabilities in modeling and simulation of complex systems for the healthcare sector with a capability to model the overall, national healthcare enterprise as well as diverse healthcare services. Application to "what if" scenario analysis is a valuable contribution to understanding the future healthcare cost curve as impacted by government programs and policies.

Through publications and briefings, there is increased recognition of potential contributions to the planning, analysis, and evaluation of HHS/CMS and VA programs and policies. Project staff presented a paper on multiscale hybrid modeling methods for the healthcare enterprise at the International Systems Engineering Conference in March 2012. The paper described a data-analytic architecture that can be applied to integrate diverse data sets and analytical models of healthcare delivery in a comprehensive framework for planning, analysis, and evaluation of programs.

The project team also briefed MITRE staff on the research to support the VA and HHS/CMS programs and demonstrated systems engineering methods for investment decision making, healthcare modeling, and impact assessments. Work products included:

- The multiscale approach uses the RHSS framework applied to a specific health service—diabetes management; and
- The data-analytic tools and methods described in briefing materials and papers published for dissemination to MITRE program staff.

Future Directions

The impact of policies and new technologies could be analyzed using a hybrid modeling approach including linkages of the following methods to analyze healthcare services as a nested System of Systems (SoS), including:

- Agent-based methods describing the population to be treated;
- Systems dynamics methods to represent the influence of policies and technologies on health service delivery;
- Discrete event state-space descriptions of the progression of the disease with and without management; and
- Inter-industry and macro-economic models to capture the impacts on the healthcare sector in the context of the U.S. economy, including 12 healthcare-related sectors of the U.S. economy that account for 18 percent of gross domestic product (GDP) and numerous job opportunities.

Summary, Conclusion, Future Directions

As the MITRE commitment to Health Transformation in the coming decade continues to grow and our responsibilities as the new CMS Health FFRDC grow with it, so too will the goals and objectives of the Health Transformation Innovation Portfolio. The Innovation Portfolio will continue to play a key, lead role in our work program and impact shaping activities, both applying emerging technologies and concepts as well as leveraging those proven in other domains in new and novel ways to the advantage of MITRE's Health Transformation Sponsors. We anticipate more emphasis on advanced techniques to Equip decision makers (operational and clinical, provider and payer) and Empower patients through data and prevention services.

MITRE and its diverse principal investigators are well positioned in the heart of the U.S. healthcare region—New England—as well as across key Washington, D.C. and Baltimore sites. We are positioned to leverage not only domain subject matter expertise but also advanced technology as it emerges from the New England cradle of U.S. academia and research, the Washington policy-setting environment, and the Baltimore daily interaction with HHS/CMS sponsors. Strategic leverage will be applied in the adoption and adaptation of transformational health technologies to these geographically present strengths to shape MITRE work programs for the benefit of our sponsors and their stakeholders. For example, several FY12 IR&D and MSR programs are leveraging access to Boston-based Partners Healthcare and Washing-

ton-based National Rehabilitation Hospital while others are leveraging academic institutions, including Massachusetts Institute of Technology (MIT), Boston University (BU), Johns Hopkins University (JHU), University of Maryland, Baltimore County (UMBC), University of Virginia, and federal healthcare providers, including the VHA.