The University of Michigan Honest Broker: A Web-based Service for Clinical and Translational Research and Practice

Andrew D. Boyd, MD, Paul R. Saxman, MS, Dale A. Hunscher, MS, Kevin A. Smith, MS, Timothy D. Morris, MS, Michelle Kasten, Frederick Bayoff, MA, Bruce Rogers, Pamela Hayes, Namrata Rajeev, PhD, MBA, Eva Klime-Rogers, MS, RN, NP, Kim Eagle, MD, Daniel Clauw, MD, John F. Greden, MD, Lee A. Green, MD, MPH, Brian D. Athey, PhD

Abstract

For the success of clinical and translational science, a seamless interoperation is required between clinical and information technology. Addressing this need, the Michigan Clinical Research Collaboratory (MCRC) was created. The MCRC employed a standards-driven Web Services architecture to create the U-M Honest Broker, which enabled sharing of clinical and research data among medical disciplines and separate institutions. Design objectives were to facilitate sharing of data, maintain a master patient index (MPI), deidentification of data, and routing data to preauthorized destination systems for use in clinical care, research, or both. This article describes the architecture and design of the U-M HB system and the successful demonstration project. Seventy percent of eligible patients were recruited for a prospective study examining the correlation between interventional cardiac catheterizations and depression. The U-M Honest Broker delivered on the promise of using structured clinical knowledge shared among providers to help clinical and translational research.


Introduction

A key lesson underlying recent advances in the biomedical and life sciences is that research must often cross disciplinary and institutional boundaries to best understand inherent scientific and sociological complexities.1,2 The adoption of clinical research data processing systems began in earnest only in the past two decades, and their adoption in academic health centers has been slower than in the pharmaceutical industry.3 Most current-generation clinical trial management systems are typically structured as monolithic, stand-alone applications characterized by proprietary data structures for interprocess communications and persistence of data, and the ad hoc encoding of data elements. Data sharing and integration involving these systems typically require logging into multiple systems from which datasets are extracted, impeding cross-disciplinary and interinstitutional collaboration.

One element of the National Institutes of Health (NIH) Roadmap program,4 initiated in 2004 by the National Heart, Lung, and Blood Institute (NHLBI), and later moved to the National Center for Research Resources (NCRR), was a Broad Agency Announcement (BAA) contract mechanism entitled “Re-Engineering the Clinical Research Enterprise.” The BAA was charged with the discovery of new tools and methodologies for the conduct of multidisciplinary, multi-institutional clinical research projects, to decrease the cost of such projects and to facilitate the extension of their reach into the community at large. In response, the Center for the Advancement of Clinical Research (CACR) and the Depression Center at the University of Michigan proposed and received funding to create the Michigan Clinical Research Collaboratory (MCRC).

The technical mission of the MCRC was to develop and deploy a system (the Honest Broker, or HB) to facilitate seamless and secure sharing of clinical data and research data within an organizational collaboration between geographically dispersed healthcare providers and research entities. Design objectives were to facilitate sharing of data, maintain a master patient index, deidentify data in conformance with HIPAA limited dataset HIPAA safe harbor, and route the data to preauthorized destination systems for use in clinical care, research, or both.

This paper describes the architecture and design of the Honest Broker and the results of a successful demonstration project.
project between U-M cardiologists, U-M Depression Center psychiatrists, and family medicine physicians in primary care settings outside the U-M Health System. The demonstration showed that the HB could (1) effectively support clinical and translational research in the challenging setting of busy full-time primary care practices and link it to academic resources, and (2) achieve unusually high levels of patient participation in C&T studies that involve more than one medical discipline or a diversity of care provider and research sites.

**Background**

There has been much discussion in the biomedical informatics literature regarding data brokers and health information exchanges. In other domains, moving large quantities of data from one Internet host to another is common practice. Health information, however, imposes much greater responsibilities upon sharers for privacy and security. In addition, when health data are used in research, issues of informed consent must also be addressed. The need for secure, reliable mechanisms for the automated transfer of biomedical research data are widely acknowledged, but maintaining privacy, security, and consent are daunting challenges. Additional challenges include correctly matching individuals’ identities across disparate systems, and ensuring that terminologies and data elements used in the various systems can be correctly associated across system boundaries (lexical compatibility). Meeting these challenges also requires assurance that the appropriate legal agreements and regulatory safeguards are in place and their terms respected. These activities are not easily automated.

Earlier efforts at the implementation of automated health information data sharing have fulfilled some subset of the functional requirements delineated above; such as Atemris, CaBIG, i2b2. These are detailed in the Online Supplement: “Previous Automated Data Sharing Efforts.” None of these previous efforts fully integrates all the privacy concerns, security concerns, consent management, identity management, and lexical resolution implemented in the MCRC Honest Broker project.

**Design Objectives**

The Honest Broker is an automated tool for the secure, reliable, institutionally authorized transmission of biomedical data among a multidimensionally heterogeneous network of clinical and research entities. The Honest Broker is intended to provide the means to accelerate translational research while simultaneously addressing the very real concerns surrounding patient privacy, information security, informed consent, lexical equivalence, and identity matching. In addition to the stringent industry-independent information security requirements for the handling of sensitive data of any type, such a tool should provide features specific to medical practice and biomedical research. These include (1) features that combine, filter, and/or transform datasets during the transmission process, to ensure that destination entities received all the information they need, and only the information they need; (2) a master patient index (MPI), eliminating the need for data sources and sinks to know the unique identifiers of patients in external organizations with which they are exchanging patient-specific data; (3) the ability to deidentify datasets in conformance with definitions of HIPAA limited or HIPAA safe harbor datasets provided in the human subjects protections regulations 45 CFR Part 164.501 and 164.514, implementing the Health Information Portability and Accountability Act of 1996 (HIPAA); and (4) the ability to route data intended for use in clinical practice, research, or both, in compliance with regulations.

The functional requirements of our HB project can be succinctly categorized into data sharing, identity management, and research data HIPAA safe haven deidentification or HIPAA limited dataset generation.

**Data Sharing**

As research crosses institutional boundaries, the data collected in individual silos needs to be transmitted to the other silos. Our design goal was not to replicate the databases of our collaborators, but instead to facilitate the transmission of relevant data to the appropriate clinical databases. The HB stores no clinical information, but instead acts as a router of clinical information. If a patient consents to participate in a research trial, the relevant research related data are routed to the research database.

**Identity Management**

A master patient index (MPI) is required to associate patient identifier information across disparate systems, to facilitate interchange of personal data among the various participating computer systems. Unlike Microsoft HealthVault, Google Health, or Dossia, the Honest Broker holds no medical data. The MPI is designed to hold system identifiers for the patient at all the interconnected medical systems, simplifying the integration between systems. This also allows for additional integration of other datasets without having to identify the same individual using multiple methods. This design also decreases people’s concerns about the security of another entity holding the medical records.

While not initially stated, one objective of the Honest Broker was to support human confirmation of potential matches of patient identities. Automatically associating patient identities across medical record systems raises the concern of false positive matches. These occur with several different matching algorithms, and there are substantial ethical and liability issues involved in the potential corruption of patient records with incorrect data. Additionally, identity theft is a rare but serious reality in today’s medical system. In light of these concerns, until a human confirmed a match between any two clinical systems, no data were to be permitted to flow between the systems.

**HIPAA Limited Dataset and Safe Harbor**

While medical records keep identity data with medical data, research projects usually attempt to separate out the identifiable demographic data from the research data with a secure and separate crosswalk file. The HB was designed to separate out the patient and research subject identifiers from both the clinical and research data. Depending on the research project, a HIPAA limited dataset or HIPAA safe harbor dataset can be constructed. The HB is a separate information system with a more rigorous security and auditing infrastructure and strictly limited user access, thereby securing the master patient index. A more detailed description of the security surrounding the Honest Broker is given in another paper.
Demonstration Project

The MCRC conducted a small prospective observational study of comorbidities between interventional cardiac catheterization events and clinical depression in community-based primary care practices. Several community hospitals and primary care practices in two medium-sized metropolitan areas (Grand Rapids and Lansing, MI) were involved, along with the University of Michigan Depression and Cardiovascular Centers and the Great Lakes Research Into Practice Network (GRIN, http://www.grinpbrn.org), a practice-based research network (PBRN) of some 300 primary care practices in Michigan.

System Description

The Honest Broker is a suite of software components, including a core system and a set of interfaces that are each associated with an individual clinical or research software system. These components communicate via the Web in a star, or hub and spoke, network pattern that is implemented as a Services Oriented Architecture (SOA). The hub of the network, the Honest Broker Core (HBC), is responsible for routing messages and clinical data between the other systems of the network. In addition, the HBC provides services for identifying patients; manages patient identifiers; provides a mechanism for partial, rules-based deidentification of protected health information; can convert messages between independent, structured formats, simplifying the integration of external systems; and ensures that all data are dealt with securely by enforcing system authentication, authorization, and message encryption.

The Honest Broker interface components are light-weight client-server systems that manage access to specific clinical data stores and communicate with the HBC securely over the Internet using standard, Web-based protocols (e.g., https) and Web Services (See online supplement: Honest Broker interface components, for more details).

In the feasibility study, the Honest Broker Core and interface components were deployed to facilitate interoperability between three previously unconnected clinical care and research systems, and a clinical research management system, each of which is operated by different departments and different institutions. The interconnected systems include (see Figure 1 for diagram):

- ClinfoTracker\textsuperscript{17}: A clinical prompt and reminder system developed by the U-M Department of Family Medicine, recently commercialized as Cielo Clinic (Cielo MedSolutions, Inc, Ann Arbor, MI), that is deployed at multiple primary care sites within the State of Michigan.
- Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2)\textsuperscript{18}: A system for capturing and storing deidentified interventional cardiac catheterization outcomes data developed at the U-M Cardiovascular Center as part of the Michigan Cardiovascular Outcomes Research and Reporting Program (M-CORRP).
- M-Strides: A depression screening and monitoring system that was developed by the U-M Depression Center and Department of Psychiatry. This computer system supports a collaboration called Michigan Depression Outreach and Collaborative Care (M-DOCC), where clinical care providers help patients complete their screenings via phone calls.
- Velos eResearch\textsuperscript{19}: A commercial clinical research management system (Velos, Inc, Fremont, CA) operated by the Michigan Institute for Clinical and Health Research (MICHR), successor organization to CACR.

The Honest Broker allowed an interventional cardiac catheterization patient’s cardiology and depression screening data to be shared, from BMC2 and M-Strides respectively, with one or more ClinfoTracker instances serving family practices where the patient received postdischarge follow-up care. In the primary care setting, patients were screened, consented, and enrolled in the prospective observational study using ClinfoTracker. The consent management process is unique to both ClinfoTracker and the HBC.

While many EHR’s will have notes stating the patient is enrolled in a specific study, the EHR is not the system that sends the message to the clinical trial management system. After the clinical research coordinator-consented the patient, the information was entered into ClinfoTracker. From ClinfoTracker, a message is generated to the Honest Broker to enroll the patient into Velos. Upon enrollment, previously sent percutaneous coronary intervention (PCI) data from BMC2 becomes requested by HBC to be sent to Velos. If patients decline enrollment, their clinical data are still passed through HBC to ClinfoTracker for their physicians’ use.

The use of SNOMED CT allowed HBC to maintain semantic interoperability between all the systems. The cardiology data collected in Velos for the study was a subset of the data sent to the ClinfoTracker and passed to practices for clinical use. The SNOMED CT codes were passed to the individual systems. However, ClinfoTracker maintained the SNOMED CT codes and Velos did not directly store the codes. The limitations of Velos were due to technical difficulties. The internal vocabularies of the different systems in this network were beyond the control of the research project. HBC maintained XLST transformation to migrate the proper subsets of information from the BMC2 data to both the ClinfoTracker and Velos. The benefit of using SNOMED CT allowed all systems to maintain the same meaning across systems and domains. The primary care doctors and ClinfoTracker did not need to learn the data definitions of the BMC2 data collection tool.

A benefit of the HB system is in the event of an adverse event or withdrawal from the study. The project was fortunate that no adverse events were reported or patients withdrew. However, upon an adverse event, the data transmitted to Velos could be turned off for all participants, until time of reconsent of all the research subjects. Disenrollment of the patient from this study would still permit all the follow-up clinical data to continue to be routed but no additional research data would be collected. This provides another level of protection for research subjects to meet the requirements of participating in a clinical trial, and enables the remote collection and management of research data.

The Honest Broker supports multiple security schemes to ensure the safety of patient data, while leveraging pre-
existing network-based security infrastructure (See online supplement: Security Infrastructure, for more details).

The HBC was developed using best-practices-based software engineering industry tools, design strategies, and development life cycle methodologies; which resulted in real-time updates about the quality and completeness of the software system (See online supplement: Software engineering methodology, for more details).

The development of the Honest Broker followed a multi-staged, iterative process that included the development of a clinical research domain model (see Figure 2); the identification of the messages required to support the feasibility study (see Figure 2); the writing and refining of the software; and testing of the system (see online supplement: Domain Model, for more details).

The HBC was designed as an n-tier software application, with discrete layers for information presentation and capture, business logic, and data persistence, and retrieval. The presentation layer, the topmost tier, is divided into two subcomponents, a Web application interface for human-computer interaction, and a Web Services interface for computer-computer interaction. The human-computer component of the presentation tier uses a model-view-controller (MVC) architecture, with separate components (controllers)
for handling requests and rendering user interfaces from HTML templates (see online supplement: Honest Broker core, for more details). The middle, business tier encapsulates all business logic operations of the Honest Broker, including technical operations, such as data validation, information routing, transaction handling, and auditing, as well as the functional operations required for the project, including subject enrollment, patient identity management, limited PHI de-identification, and patient searching. The persistence tier of the HBC was designed to employ the data access object (DAO) software design pattern (See online supplement: Honest Broker core, for more details).

One feature requiring additional discussion is the transaction support. This feature of the HBC was built following a model of events and subevents, where a transaction fault, e.g., a failure to route information to one of several destinations, causes the system to notify a system administrator via e-mail and to signal the event initiator to retrigger the event at a later time, usually 24 hours later. Once an event is retriggered, only incomplete subevents are executed, which alleviates the need for systems outside of the HBC from having to individually implement transaction support or monitor for duplicate information being sent. All events processed by the HBC, both successful and failed ones, are logged for auditing purposes, with each log entry including an identifier for the event initiator, the event status, and the time of the event.

Standards-based Message Formats and Common Data Elements

One of the stated goals of the project was to demonstrate that existing technologies were sufficient to enable interoperability between disparate system architectures, programming languages, message formats, and data encoding schemes. The use of Web Services as the transport layer accomplished the first two aspects of interoperability. Message formats and data encoding schemes would define the “payload”, and desired use existing standards.

To address the need for a standardized message format, we evaluated several alternatives, and ended up choosing the Clinical Document Architecture (CDA), a constrained subset of the Health Level 7 Version 3 (HL7v3) standard, for transmission of most messages containing clinical data. We chose to employ a mix of project-specific data structures defined by the development team for messages containing no clinical data. All messages used the XML standard as the syntactical base, and validation was enabled using definitions conforming to the XML Schema standard. Use of XML allowed us to easily translate incoming messages into a common internal XML format, which was then transform-
able into outbound formats after internal processing, filtering, and combining operations were performed. One benefit of using HL7-CDA was forward planning for future development. While all future systems messages would need to be validated, using HL7-CDA would permit faster interoperability.

The CDA message format is rich, heavily nested, complex, and more than adequate to meet our requirements. Its complexity was reflected in a steep learning curve, especially for developers with little or no prior XML experience. Our internally developed XML document types tended to be much simpler than CDA documents, primarily because we were able to define single-scenario data structures vis-à-vis the CDA mandate to serve a multitude of use cases. Of particular value to us were the CDA facilities for nested qualifiers on observational metadata, necessary to express in computer-readable form such constructs as “myocardial infarction event in the context of interventional cardiac catheterization procedure.”

Selection of standards for data encoding narrowed quickly to SNOMED CT and LOINC. For organizational identifiers, we chose the international Abstract Syntax Notation One (ASN.1) standard. Since our clinical problem domain was cardiology, we were unable to leverage the extensive work funded by the NCI under its caBIG21 initiative. While examination of other terminologies occurred, with SNOMED CT, we found an ontology fully capable of expressing the concepts in the project’s problem domain, with a few minor exceptions related to experimental assay results.

Downstream licensing issues were a factor considered in the evaluation. SNOMED CT is freely available for use in noncommercial contexts, owing to an agreement between the College of American Pathologists (CAP) and the National Library of Medicine (NLM). LOINC has similar licensing terms. One project survey instrument, the Patient Health Questionnaire 9 (PHQ-9),22 is licensed by Pfizer. Use of the PHQ-9 has grown rapidly as a screening and monitoring tool in psychiatric and nonpsychiatric settings. We worked with Pfizer and Regenstrief Institute, curators of the LOINC standard, to incorporate the PHQ-9, thus extending LOINC as part of the project. Our other cardiovascular survey instrument was proprietary, and its owners did not wish to consider its inclusion at this point.

The use of interoperable standards did impose a significant up-front cost, both in terms of learning curves and programming effort. Moreover, complex multipurpose data structures and encoding schemes sacrificed a degree of readability for computational accuracy and semantic expressiveness.

**Status Report**

Institutional Review Board (IRB) permission was obtained to begin the study from the four separate IRBs involved. The software and hardware for ClinInfoTracker was then deployed in two GRIN member practices in one metropolitan area, one urban and one suburban; and two additional deployments in another area serving uninsured populations, largely of “working poor”, one urban and one rural.

Patients could enter the study through one of two ways: (1) having a diagnosis of CHD made by their primary-care physician, or (2) having an acute-event coronary catheterization procedure at one of the participating BMC2 hospitals. Patients were consented for the study at their primary care office. Ninety-three patients with CHD were screened for depression; 88 enrolled in the study. The BMC2 database has 334 total entries for acute coronary events during the study period from the participating hospitals, of which 61 matched to patients in the primary care practices participating in the study. Most of the matches were historical; only 2 patients had acute coronary events during the study period. Of the 61 matches, there were no false negative or false positive matches; each match was verified by a member of the primary care clinical staff team. Of the 93 patients screened for depression, 23 were managed by the depression disease management program (M-DOCC), with the data forwarded to the primary care system.

The total number of messages handled by the Honest Broker system was 1,528. This number is relatively low, but demonstrates the feasibility of sharing data across institutions for clinical and research purposes. Originally, we envisioned using only dually authenticated SSL to secure the sharing, but some primary care sites preferred a VPN over SSL. Both methods were used and proved reliable for safe and secure data transmission.

**Discussion**

The increasing pace of development of Health Information Exchanges (HIEs) and Personal Health Records (PHRs) illustrates the building pressure for safe and secure sharing of medical records. Some HIEs, however, are creating monolithic databases to store relevant clinical data. In our experience, the federated network concept is more acceptable to collaborators than the creation of yet another entity that would store patient data, especially when the data has clinical and research purposes. It also appears to be more compatible with the long-term clinical and translational research goals of NIH. The link between depression and cardiovascular disease has been strongly established clinically,23 and federated data sharing networks like the HB will help to address the challenge of treating such co-occurring diseases. Synchronization of multiple monolithic data sources is already problematic, given the fragmented nature of the United States healthcare system; perpetuating the problem in the design of hybrid patient care/research data repositories is unacceptable when federation offers a practical and cost-effective alternative.

A key feature that has thus far been omitted from the requirements for HIE and PHR implementations is the ability to link relevant clinical data for use in research projects. High fidelity clinical data can be collected and reused for clinical research purposes with the appropriate consents and permissions obtained. At present, obtaining multiple IRB approvals and cooperation of healthcare IT staff requires significant effort, but this additional cost for both HIE and PHR systems should be considered a necessity as we seek to advance medicine through clinical and translational research.

Although the success or failure of HIEs, as with commercial Electronic Data Interchanges,24 is most likely governed by the traditional dimensions of interorganizational power and trust, we found the positive factors related to trust to be particularly relevant for an academic, federated data-shar-
Internet requires a higher standard for software engineering and data security management practices. While many were the adherence to commercial-grade software engineering expectations, and negotiating resource allocations to meet revised effort and timing estimates.

Working with data standards such as LOINC and SNOMED CT helped us achieve interoperability between the separately developed information systems. We believe that once the standards are implemented on both ends of a transaction, they greatly increase the likelihood of a successful and secure interchange of data. However, this project brought to light a potentially serious risk of disparities of cost and benefit in the deployment of an interoperable, standards-based biomedical information systems infrastructure merging clinical operations and clinical research informatics systems. Clinical operations applications are often maintained by legacy programming staff, many of whom have limited knowledge of research-related requirements for interoperability. Moreover, the missions and cultural norms of operations and research IT organizations are not often harmoniously aligned. These issues must be addressed globally if the national and international biomedical research enterprises are to make significant progress toward the free exchange of data in a regulatory-compliant health information infrastructure (HII).

Data exchange between institutions is notoriously challenging. This project benefited from the fact that there was no adversarial relationship(s) between participating hospital and community-based practices. While each discipline involved in our research protocol used a separate lexicon to share data, including SNOMED CT, LOINC, and ICPC, the systematic mapping of domain vocabularies helped ensure mutual understanding. Currently, there is no deterministic solution for semantically correct information exchange between domain-specific vocabularies. Collaboration between communities of generalists and specialists could facilitate the harmonization of domain terminologies.

A unique feature of the HB compared with other projects was the adherence to commercial-grade software engineering and data security management practices. While many successful programs can be developed with limited programming skills, sharing sensitive patient data over the Internet requires a higher standard for software engineering quality and formal data security planning than for sharing in an Intranet. This results in higher up-front costs, but the potential liabilities involved in security breaches are far greater and justify these up-front costs. This requirement must be addressed at the highest levels of the sponsors and the universities conducting biomedical research.

While the HIE and PHR projects are significantly larger in scope, this small pilot project demonstrated that incorporating using clinically structured data in a clinical workflow can be done successfully, as well as extracting the data for clinical research purposes. The primary care clinics did not view the recruitment or confirmation of patients from their clinic as a large burden. A recruitment rate of 70% of eligible patients in a primary care clinic is remarkable compared with typical published rates of 18%. While technology is not sufficient to study a distributed population of community-based patients, it does lower the bar to begin to study these populations in a cost-effective manner. Moreover, the substantial increase in the size of the potential subject pool will make possible research that is currently difficult or impossible to undertake, both from logistical and cost perspectives. In talking with primary care providers, the human confirmation of patients’ identities was considered necessary, since ClinInfoTracker uses the data to generate clinical reminders and decision support as well as store medication history. Therefore, the matches must be correct at the individual level; even a very high percentage correct at the aggregate level would mean some erroneous matches that would erode trust in the system and willingness to share data. While this could potentially limit the initial scalability of the system, once the links are established they are maintained. As increasing the number of patients in the system would also increase the number of primary care physicians available to confirm matches, the system as architected does scale without increasing burden on providers for match confirmation. It is routine in practice for a new patient to fill out a several page form, or a returning patient is often asked to describe any changes in their medical condition; so the alternative of human confirmation of matches followed by automatic data transfer may not significantly increase overall effort.

The HB approach does suggest solutions to many outstanding issues related to data sharing and clinical and translational research. This method, as currently implemented, will not scale nationally with a single Honest Broker instance, but a single instance could serve a large number of patients within a geographic region. With an increasingly mobile society, development of additional data structures and messages to facilitate sharing data on a national scale will be required for scaling beyond regional use. Moreover, a pressing need for improving the efficiency of the difficult process of obtaining multiple IRB approvals for data sharing projects is apparent. Further study will also be needed to fully understand the feasibility of the human-confirmed match process in various scenarios.

The next steps in this research are to see if this model of engaging community-based practices can be employed in other communities. Engaging other investigators and study participants that would benefit from the involvement of community-based practices is another step. The current Honest Broker needs additional development and refinement to become universally applicable. This software was developed on an NIH contract, and will be shared at no cost.
to any nonprofit entity for clinical research purposes. The Honest Broker is actively maintained at UM. The software is being used currently for another project involving Federally-Qualified Health Centers in Michigan, another project: the Multi-Modality Multi-Resource Environment for Physiological and Clinical Research (a CTSA Informatics Pilot Project), and is a key part of the infrastructure for two pending grant proposals. Commercial development is being explored but no decision has been made on commercializing it at this writing. Necessary license agreements will be developed to allow others to collaborate and expand the concept of the Honest Broker. While this was a small pilot demonstration, as academic medical centers and their partner practices begin to share and integrate clinical and clinical research datasets, the Web Services-based approach demonstrated in this project should be considered as a cost-effective means to share data. Novel paradigms such as this approach are the key to obtaining the maximum possible benefit to the public health in the translation of the findings from clinical research into clinical practice.

References