



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): Architecture

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters.

Citation	Mandl, K. D., I. S. Kohane, D. McFadden, G. M. Weber, M. Natter, J. Mandel, S. Schneeweiss, et al. 2014. "Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): Architecture." <i>Journal of the American Medical Informatics Association : JAMIA</i> 21 (4): 615-620. doi:10.1136/amiajnl-2014-002727. http://dx.doi.org/10.1136/amiajnl-2014-002727 .
Published Version	doi:10.1136/amiajnl-2014-002727
Accessed	February 16, 2015 4:53:08 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:12717616
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)



OPEN ACCESS



Open Access
Scan to access more
free content

For numbered affiliations see end of article.

Correspondence to

Dr Kenneth D Mandl, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA; kenneth_mandl@harvard.edu

Received 18 February 2014

Accepted 8 March 2014

Published Online First

12 May 2014



- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002864>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002747>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002758>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002764>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002744>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002746>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002743>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002827>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002740>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002751>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002745>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002759>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002756>



CrossMark

To cite: Mandl KD, Kohane IS, McFadden D, et al. *J Am Med Inform Assoc* 2014;**21**:615–620.

Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): Architecture

Kenneth D Mandl,^{1,2,3,4} Isaac S Kohane,^{1,2,3,4} Douglas McFadden,^{2,3} Griffin M Weber,^{2,5} Marc Natter,^{1,4} Joshua Mandel,^{1,4} Sebastian Schneeweiss,⁶ Sarah Weiler,³ Jeffrey G Klann,⁷ Jonathan Bickel,^{1,4,8} William G Adams,^{9,10} Yaorong Ge,¹¹ Xiaobo Zhou,¹² James Perkins,^{13,14} Keith Marsolo,¹⁵ Elmer Bernstam,¹⁶ John Showalter,¹⁷ Alexander Quarshie,¹⁸ Elizabeth Ofili,¹⁹ George Hripcsak,²⁰ Shawn N Murphy^{7,21}

ABSTRACT

We describe the architecture of the Patient Centered Outcomes Research Institute (PCORI) funded Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS, <http://www.SCILHS.org>) clinical data research network, which leverages the \$48 billion dollar federal investment in health information technology (IT) to enable a queryable semantic data model across 10 health systems covering more than 8 million patients, plugging universally into the point of care, generating evidence and discovery, and thereby enabling clinician and patient participation in research during the patient encounter. Central to the success of SCILHS is development of innovative 'apps' to improve PCOR research methods and capacitate point of care functions such as consent, enrollment, randomization, and outreach for patient-reported outcomes. SCILHS adapts and extends an existing national research network formed on an advanced IT infrastructure built with open source, free, modular components.

INTRODUCTION

The Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS, pronounced 'skills') is one of 11 clinical data research networks (CDRNs) funded by the Patient Centered Outcomes Research Institute (PCORI) in 2014. PCORI, a non-governmental organization created under the Patient Affordable Care Act seeks to build an information technology (IT) backbone to support comparative effectiveness research at a national scale across both CDRNs and also patient powered research networks (PPRNs).

SCILHS engages patients, clinicians, health systems leadership, and key healthcare stakeholders as collaborators to build on an existing network of hospitals and health systems that have already adopted a common clinical and translational research IT and regulatory framework. SCILHS, comprising 10 health systems (box 1), is a step toward answering the Institute of Medicine's call for a learning healthcare system (LHS)^{1–2} to 'generate and apply the best evidence for the collaborative healthcare choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care'.

Fifteen years ago, SCILHS informatics leaders began a quest to develop informatics infrastructure and regulatory innovation that would convert the emerging electronic health record (EHR) into a research tool for improving patient outcomes. All of our work and open source toolkits have been supported by grants from the National Institutes of Health, Centers for Disease Control and Prevention, and Office of the National Coordinator of Health Information Technology (ONC). First, we built Indivo,^{3–4} the first personally controlled health record, which gave patients their data, and apps to make those data useful. Then, i2b2 (Informatics for Integrating Biology and the Bedside)^{5–7} created an open source analytic platform to the EHR, to fuse and analyze data produced by the delivery system, and identify research cohorts. i2b2's flexible common semantic data model readily accommodates a variety of clinical data. Our next advance was SHRINE (Shared Health Research Information Network),^{8–10} a tool enabling investigators to query i2b2 nodes in real time across multiple sites for collaborative population research. i2b2 has been successfully implemented at more than 100 sites across the USA, thereby enabling investigators to use delivery system data to identify patients with specific illnesses and clinical characteristics. A recent PCORI survey of all PCORnet sites revealed that 37% of the existing CDRN nodes and 31% of the PPRN nodes already used i2b2. Finally, we built SMART (Substitutable Medical Applications, Reusable Technologies)—a platform to enable any developer to contribute to an 'App Store for Health and Research' compatible with i2b2-SHRINE instances or compliant EHRs.^{11–13}

These informatics tools and associated research policy advances have already contributed to transformation in the clinical research enterprise—real-time, collaborative population health research is now enabled across SHRINE member sites distributed nationally—but they have yet to yield substantial improvements in the health of our patients. Now, in establishing PCORnet, PCORI has catalyzed a new national research dialog to answer patient-oriented questions and improve human health. We directly address this challenge via a strategy intended to avoid prior mistakes of large-scale, top-down, costly software infrastructure efforts that

Box 1 Alphabetical list of Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS) sites

Beth Israel Deaconess Medical Center
 Boston Children’s Hospital
 Boston Health Net (Boston Medical Center and Community Health Centers)
 Cincinnati Children’s Hospital Medical Center
 Columbia University Medical Center and New York Presbyterian Hospital
 Morehouse School of Medicine/Grady Memorial Hospital (Research Centers in Minority Institutions)
 Partners HealthCare System (includes Massachusetts General and Brigham & Women’s Hospital)
 University Mississippi Medical Center
 The University of Texas Health Science Center at Houston
 Wake Forest Baptist Medical Center

failed to scale (eg, caBIG¹⁴), instead building SCILHS with open source, free, modular components^{5 15} with vibrant user and software developer communities that have already spread virally to scale across heterogeneous health systems.

Here, we detail the informatics approaches taken by SCILHS to identify large cohorts of patients and engage them for research. Our technology strategy links lockstep to processes for regulatory innovation, development of robust governance

constructs and policies, and local adoption by hospital leadership and institutional review boards.

THE SIDECAR APPROACH

SCILHS adopts and extends a strategy of establishing a freely accessible health data ‘sidecar’ warehouse to the EHR, effectively leveraging existing data collected by EHRs during routine care while avoiding costly, time-consuming EHR integrations (figure 1). Developed intensively over the past 5 years at Harvard Medical School, this approach employs vendor agnostic, free, open source, scalable, and interoperable technologies to produce the only research-based, shared repository of EHR data that can be queried in real-time. Of already proven value in the research ecosystem, these components support a cost-effective and sustainable research network of >8 million patients.

We consider the heterogeneity of collaborating institutions to be a key measure of success; via adoption of the sidecar approach, we enable any institution to join our SCILHS network. Specifically, a primary goal is inclusion of diverse populations within our CDRN network, thereby enabling capture of the genetic, genomic, and socioeconomic variation that exists beyond insured populations in managed care settings alone. Further, by freely sharing the processes and software that have been developed and supported by Harvard, we hope to catalyze the formation of many other new networks across heterogeneous health systems and institutions, and involve new

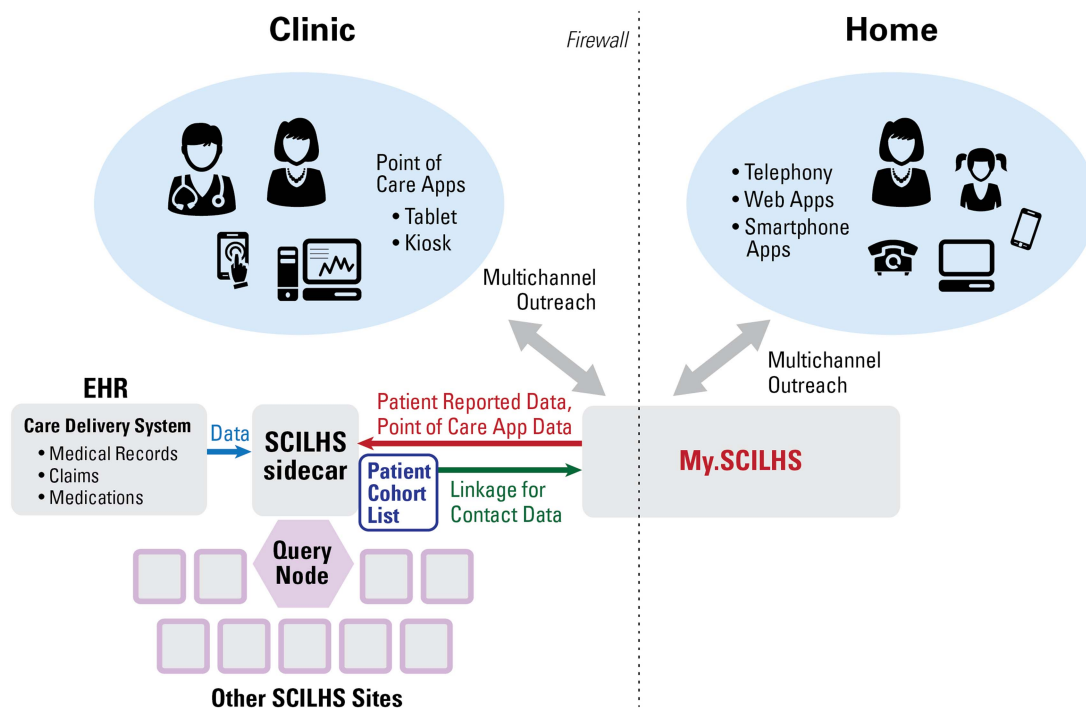


Figure 1 Each site will install the Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS) Sidecar, for identifying and reviewing cohorts and the mySCILHS suite to: (a) manage linkage of contact data to the de-identified Patient Cohort list produced by the multisite Shared Health Research Information Network (SHRINE) query; (b) administer and store consent documents; (c) outreach to patients through web-based survey and telephony; and (d) promote ongoing patient engagement through outgoing messaging, including (in the future) return of research results to patients. The web-based survey will be administered using REDCap and Indivo technologies, and will be accessible either by patients at home, or at the point of care, through tablet/kiosk-based interaction. Once completed, patient-reported data will have subject identifiers encoded; its standardized survey metadata will then be loaded into the corresponding SCILHS sidecar (i2b2 node), enabling semantic data linkage with electronic health record data via SHRINE/i2b2, while preserving subject confidentiality. These software platforms will be provided to sites as self-contained, pre-configured virtual machines, enabling rapid dissemination of these technologies while minimizing administrative and software development overhead at each site.

partners in improving our core components, common data models, and ontologies.

The sidecar infrastructure is composed of the following:

- ▶ *i2b2* (Informatics for Integrating Biology and the Bedside). Data analytic platform employed for EHR data analytics and clinical research at >100 academic medical centers worldwide (NIH funded).
- ▶ *SHRINE* (Shared Health Research Informatics Network). Federated query and response system that enables investigators to discover EHR data housed in *i2b2* nodes across multiple independent institutions (NIH CTSA funded).
- ▶ *SMART Platforms*. First described in the *New England Journal of Medicine*,¹² SMART has programmatic interfaces and applications that transform both EHRs and their sidecars into platforms that run substitutable iPhone-like apps.¹¹ SMART enables a national scale ‘App Store’ for PCOR for rapid cycle innovation of PCOR methods (ONC funded).
- ▶ *Indivo*. The original personally controlled health record^{3 4 16 17} links patients to clinical and research settings. Used by hundreds of thousands employees of Dossia’s founding companies (Wal-Mart, Intel, and AT&T), Indivo was also the initial software codebase for Microsoft’s HealthVault platform (NIH, CDC, and ONC funded).
- ▶ *REDCap* (Research Electronic Data Capture). Electronic data capture tool^{18 19} with 757 institutional partners, used to survey patients online (NIH CTSA funded).

DATA MODELS AND ONTOLOGIES

SCILHS will combine EHR data with payer claims to facilitate longitudinal tracking of patients over time and across sites of care. The sidecar approach provides the capability to implement new data models without transforming all of the stored source data—a key element in the scalability and interoperability of our platform (table 1). By enabling well-designed, cross-mapped ontologies that support a PCORnet common data model, this approach incorporates otherwise disparate clinical data sources into an easily-queried system that stores data in a flexible format. Data are stored in *i2b2* using an entity–attribute–value model,^{20 21} employing a central ‘fact’ table based upon Kimball’s *Star Schema*²² wherein each row stores a flexibly defined, atomic ‘fact’ or observation for a patient.⁵ Much of *i2b2*’s versatility arises from its focus on a semantic definition of patient observations that can represent various existing and newly defined data elements: claims, EHR, genetic and imaging data, as well as patient reported outcomes and demographics. Analogous to a capacious warehouse with adjustable shelves and bins, *i2b2* accommodates various nomenclatures for data elements, and supports robust tags of associated modifiers and values. This approach enables database indexing of facts and observations to support high performance execution of expressive queries and filters.

i2b2 employs an ontology-based approach that supports flexible, on-the-fly incorporation of new data elements and coding systems. Terminologies such as ICD, NDC, and LOINC may be pre-loaded as hierarchical concept trees; new or ad-hoc terminologies including patient-reported outcome measures or locally defined data dictionaries readily coexist and may be cross-mapped in *i2b2*. Concepts may be grouped using simple hierarchies and then optionally re-mapped into other reference coding systems and data models (eg, Observational Medical Outcomes Partnership (OMOP) data model).²³ In this way, *i2b2* accommodates diverse real-world coding systems while maintaining a straightforward query interface for its users.

The *SHRINE Adaptor Cell* maps local *i2b2* terminologies into a common, standards-based SHRINE ontology. This enables a common shared ontology for federated queries while allowing individual *i2b2* instances within institutions to retain local hierarchies and terminologies. The Adaptor transforms a federated SHRINE query into a query that runs on the local *i2b2* database. The Adaptor then converts the result of that query back into the common SHRINE message format, using well-maintained standards including RxNorm, ICD9, and LOINC. In addition, SHRINE includes tools for ontology mapping and ontology-based data mining. Simple SHRINE customizations enable use of other query systems, for example the QueryHealth distributed query system (ONC) uses PopMedNet to query *i2b2*.^{24 25}

SUCCESS TO DATE

SHRINE and *i2b2*-based research includes characterization of rare morbidities of common diseases,²⁶ very rare diseases such as peripartum cardiomyopathy (discovered in SHRINE and published in *Nature*²⁷), detections of drug–drug interactions,²⁸ and measures of quality and clinical efficacy across self-organized SHRINE networks in Europe, the University of California healthcare systems, and a just-in-time network to study the prevalence of complication rates of type 1 and type 2 diabetes in hospitals across this country. Others have used SHRINE to characterize and track the rising incidence of colorectal cancer²⁹ and further characterize it, and to identify and optimize practice variation in inflammatory bowel disease and intervene to change that practice.³⁰ *i2b2* and SHRINE have been implemented as the base infrastructure for a variety of enhanced chronic disease registry-based research efforts.³¹ The Childhood Arthritis and Rheumatology Research Alliance uses the SHRINE/*i2b2* registry framework to federate clinical care data and patient-reported data from 62 academic medical centers in the USA and Canada^{32 33} and is currently piloting consensus treatment protocol trials.^{34–37} The Harvard Inflammatory Bowel Disease (IBD) Longitudinal Data Repository employs the same infrastructure.³¹ ImproveCareNow³⁸ utilizes *i2b2* as its centralized data warehouse for IBD-related quality improvement development at 50 centers.

Table 1 Approaches to scalability and interoperability

Sidecar approach	‘Community-extensible ontologies’
EHR data are managed in a sidecar, readily established at any institution, regardless of EHR vendor product (Epic, etc)	All schemas and ontologies we produce are open source, free, and already widely adopted
<i>i2b2</i> uses a simple data model (Star Schema) greatly simplifying the Extract, Transform, and Load procedure. These ETL procedures are established for all major EHR products	Ontologies can be imposed on the data after the fact, enabling a hospital in our network to readily adapt to any ratified PCORI Common Data Model
SMART platform specifications enable any app developer to create substitutable PCOR apps without knowing details about the underlying hospital systems	For example, there are existing transpositions between OMOP and <i>i2b2</i> and PopMedNet can query <i>i2b2</i>

EHR, electronic health record; OMOP, Observational Medical Outcomes Partnership; PCORI, Patient Centered Outcomes Research Institute; SMART, Substitutable Medical Applications, Reusable Technologies.

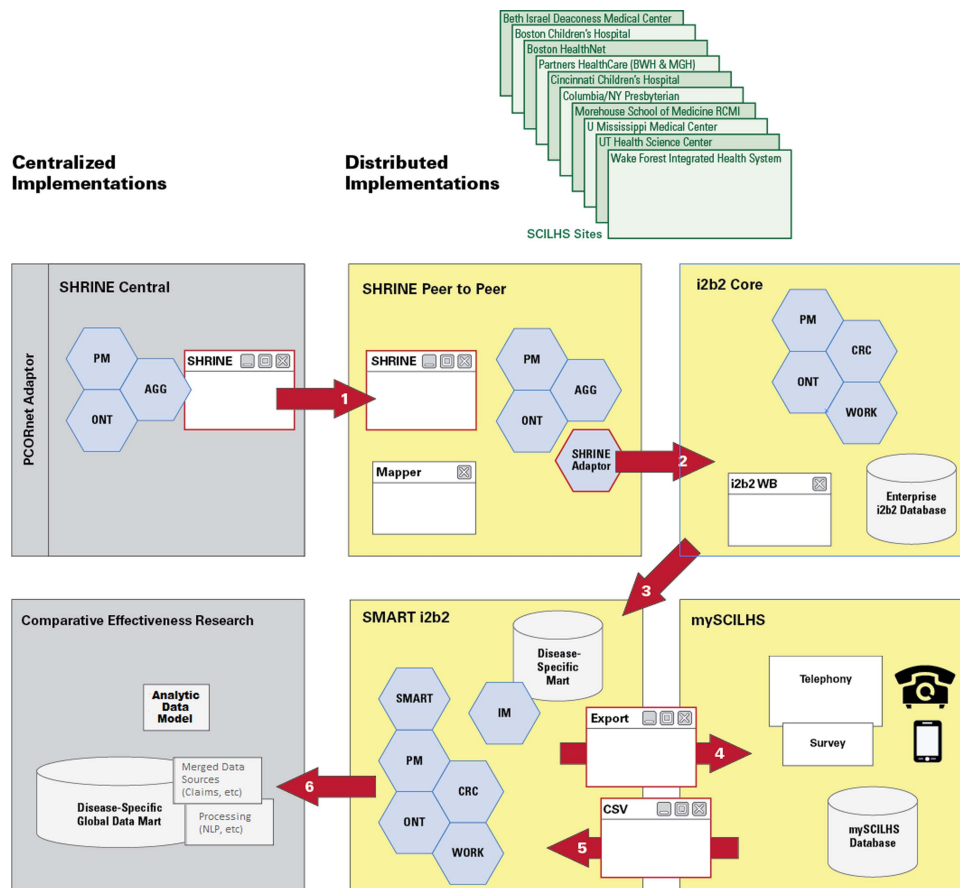


Figure 2 The Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS) data workflow. We present here a general workflow. There will be important variations depending on the nature of the study, whether in-person consent is required, and whether patient identifiers are needed. Shared Health Research Information Network (SHRINE) architecture implemented as a modular framework. Using a mapper toolkit, each site exposes a common queryable data model, implemented in the ontology (ONT) cell. The ONT cell manages the vocabulary of the data model and is one of several cells in the i2b2 architecture, including the broadcaster-aggregator cell (AGG, broadcasts the query across all i2b2 nodes in the SHRINE peer-to-peer network and aggregates the results), the Identity Management Cell (IM, used for authentication), the Clinical Research Chart (CRC, manages the clinical data), the Workplace Cell (WORK, manages the workflow), and the Substitutable Medical Applications, Reusable Technologies (SMART) Cell (manages the SMART API). We implement the following workflow. A query from a Patient Centered Outcomes Research Institute (PCORI) approved study is translated to a SHRINE central node query either manually, or by a PCORNet adaptor, the specifications for which are still to be determined. The SHRINE Central Node broadcasts the query across the true peer-to-peer network (ARROW 1). i2b2 nodes containing coded data are queried at each site to identify appropriate patients returning obfuscated, aggregate patient counts (ARROW 2). Patient identifiable data remains at each site where investigators can use SMART Apps to review records prior to aggregation (ARROW 3). Also, see [figure 3](#). The patient list is passed to mySCILHS for outreach to patients via apps, survey, or telephony (ARROW 4). Patient generated data are imported into i2b2 via simple input formats (CSV, for example) and placed into the i2b2 data model in a flexible schema that allows these to become first-class queryable data objects (ARROW 5). The adjudicated patient data (reviewed by investigators using SMART Apps and confirmed as valid) from each site, including patient-reported data can be added (ARROW 6) to a research data mart in one of several analytic data models (including the PCORI Common Data Model) with a level of identifiers appropriate to the level of consent obtained. Additional, outside data such as Centers for Medicare and Medicaid claims can be added in this step.

PATIENT ENGAGEMENT

The health systems that have joined SCILHS reflect the American demographic—an essential requirement for reaching statistically valid, clinically meaningful, and patient-centric conclusions about therapies across the diverse spectrum of all healthcare consumers. In order to achieve the comprehensive, patient-centered outcomes infrastructure called for by PCORI, we introduce a new, patient-centric platform (mySCILHS) based on the Indivo system and incorporating the REDCap electronic data capture tool.

mySCILHS will support the Blue Button REST API for standards-based interactions with PPRNs and other patient-selected tools. This API exposes up-to-date, structured clinical summary data for each participating patient. Via a consumer-friendly workflow based on web standards including OAuth2,

patients can authorize third-party apps and services, including PPRNs, to access their clinical data.

ANTICIPATED WORKFLOW

[Figure 2](#) shows the workflow from an initial query through the analytic phase in a comparative effectiveness study. Each node in the network maintains an instance of i2b2 containing claims and de-identified electronic medical record data. SCILHS is a true peer-to-peer network, meaning that any SHRINE-based node can initiate a query, using a common ontology, that aggregates results from all participating sites. After the initial query, the investigator can automatically pass the query to each site where duly authorized local site investigators may review individual subject data for study eligibility using i2b2 SMART apps ([figure 3](#)). The final patient list is transmitted to the mySCILHS

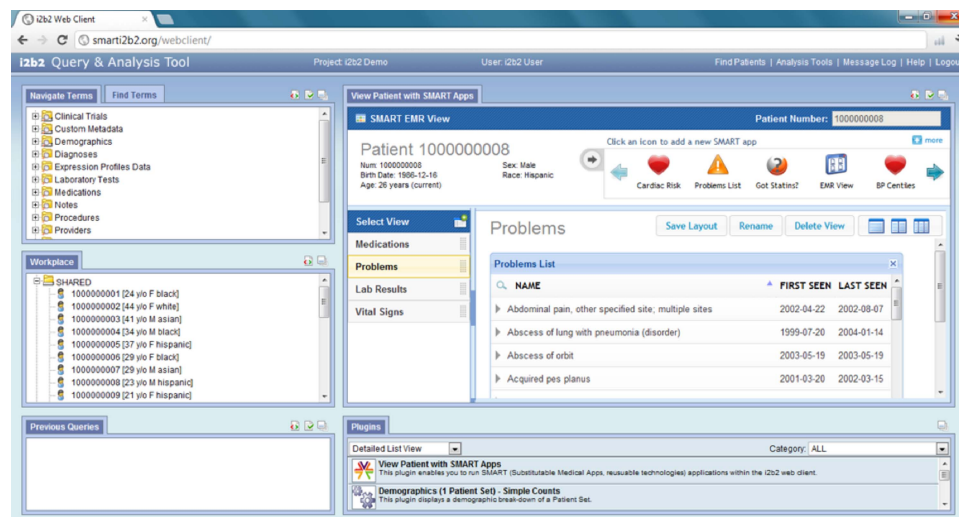


Figure 3 A Substitutable Medical Applications, Reusable Technologies (SMART) Platforms HTML5 App running on i2b2, providing a richly featured electronic health record-like view of the data.

patient-facing software. The mySCILHS research contact management module links de-identified i2b2 records to patient demographics and contact information. Patients are engaged by web survey, telephony, or SMART apps; patient-reported data are returned to i2b2 and are then transferred into a secure comparative effectiveness (CE) study environment for analyses. In the CE environment, further transformations may occur, supporting many other analytic tools and processes. We anticipate that PCORnet-level queries, which may launch against the full complement of 11 CDRNs and 18 PPRNs, will be initiated at the PCORnet adapter. We anticipate that natural language processing (NLP) of provider notes will play an important role for adding complete longitudinal coded data to the hospital-based record.³⁹ Early findings demonstrate that NLP of hospital-based EHR notes provides quite complete longitudinal data even when compared with Centers for Medicare and Medicaid Services claims data (personal communication, Katherine Liao, Brigham and Women's Hospital, 2014). Using NLP on hospital and clinic notes will complement our strategy of concatenating EHR data with external sources such as claims and pharmacy data.

IMPLEMENTING AND SCALING

SCILHS includes 10 legally and financially independent institutions whose CEO or equivalent senior institutional official has committed to active participation in governance, policy development, data sharing, and sustainability planning. Each member has pledged to invest additional personnel and resources to ensure the network meets local patient and clinical stakeholder needs. By harmonizing informatics infrastructure, data models, regulatory processes and policies, and patient participation within and across member institutions, we anticipate that SCILHS will achieve and remain a successful model for inter-institutional PCOR. Utilizing the innovative SCILHS sidecar IT approach to EHR access, we minimize local informatics burden, further enabling a sustainable and adaptable PCOR infrastructure.

Author affiliations

¹Children's Hospital Informatics Program at Harvard-MIT Health Sciences and Technology, Boston, Massachusetts, USA

²Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA

³Harvard Catalyst, Harvard Medical School, Boston, Massachusetts, USA

⁴Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

⁵Biomedical Research Informatics Core, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

⁶Division of Pharmacoepidemiology, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA

⁷Laboratory of Computer Science, Massachusetts General Hospital, Boston, Massachusetts, USA

⁸Information Services Department, Boston Children's Hospital, Boston, Massachusetts, USA

⁹Boston University School of Medicine/Boston Medical Center, Boston, Massachusetts, USA

¹⁰Boston University Clinical and Translational Sciences Institute, Boston, Massachusetts, USA

¹¹College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

¹²Department of Radiology, Center for Bioinformatics & Systems Biology, Wake Forest University Health Science, Winston-Salem, North Carolina, USA

¹³Clark Atlanta University, Atlanta, Georgia, USA

¹⁴Research Centers in Minority Institutions Translational Research Network, Data Coordinating Center, Jackson State University, Jackson, Mississippi, USA

¹⁵Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

¹⁶Division of Biomedical Informatics, Biomedical Informatics and Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, Texas, USA

¹⁷University of Mississippi Medical Center, Jackson, Mississippi, USA

¹⁸Department of Internal Medicine, Community Health and Preventive Medicine and Clinical Research Center, Morehouse School of Medicine, Atlanta, Georgia, USA

¹⁹Department of Internal Medicine, Clinical Research Center, Morehouse School of Medicine, Atlanta, Georgia, USA

²⁰Department of Biomedical Informatics, Columbia University, New York, New York, USA

²¹Partners HealthCare Systems, Information Systems, Charlestown, Massachusetts, USA

Acknowledgements We acknowledge the invaluable contributions of the many SCILHS investigators, leaders, and supporters and specifically call out those most involved in designing the network in the early phases: Barbara Bierer, Susan Edgeman-Levitan, Jonathan Finkelstein, Alison Goldfine, Jennifer Haas, John Halamka, Manny Hernandez, John Hutton, Ann Klibanski, David Ludwig, Joshua Metlay, Mary Mullen, Lee Marshall Nadler, Andrew Nierenberg, Harry Orf, Patricia O'Rourke, Eric Perakslis, Lee Schwamm, Daniel Solomon, Herman Taylor, Patrick Taylor, Aaron Waxman, Laura Weisel, and James Wilson.

Collaborators SCILHS Network.

Contributors The authors all: made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and drafted the work or revised it critically for important intellectual content; and gave final approval of the version to be published; and agreed to be accountable for all aspects of the work in ensuring that questions related to the

accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This work was supported by the National Institutes of Health: National Library of Medicine R01LM011185 and U54 LM008748; National Institute of General Medical Sciences R01GM104303; National Center for Advancing Translational Sciences 1KL2TR001100, UL1TR000454; National Institute on Minority Health and Health Disparities U54 MD007588; the Office of the National Coordinator of Health Information Technology SHARP Program Contract 90TR0001; and by Contract CDRN-1306-04608 from the Patient Centered Outcomes Research Institute (PCORI).

Competing interests SS is consultant to WHISCON, LLC and to Aetion, Inc., a software manufacturer in which he also owns shares.

Provenance and peer review Commissioned; internally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Etheredge LM. A rapid-learning health system. *Health Aff (Millwood)* 2007;26:w107–18.
- Olsen L, Aisner D, McGinnis M. *The learning healthcare system: workshop summary. roundtable on evidence-based medicine*. Washington, DC: Institute of Medicine of the National Academies, 2007.
- Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. *BMJ* 2001;322:283–7.
- Mandl KD, Kohane IS. Tectonic shifts in the health information economy. *N Engl J Med* 2008;358:1732–7.
- Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124–30.
- Goryachev S, Sordo M, Zeng QT. A suite of natural language processing tools developed for the I2B2 project. *AMIA Annu Symp Proc* 2006:931.
- Murphy SN, Gainer V, Mendis M, et al. Strategies for maintaining patient privacy in i2b2. *J Am Med Inform Assoc* 2011;18(Suppl 1):i103–i08.
- McMurry AJ, Gilbert CA, Reis BY, et al. A self-scaling, distributed information architecture for public health, research, and clinical care. *J Am Med Inform Assoc* 2007;14:527–33.
- Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 2009;16:624–30.
- McMurry AJ, Murphy SN, MacFadden D, et al. SHRINE: Enabling nationally scalable multi-site disease studies. *PLoS ONE* 2013;8:e55811.
- Mandl KD, Mandel JC, Murphy SN, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. *J Am Med Inform Assoc* 2012;19:597–603.
- Mandl KD, Kohane IS. No small change for the health information economy. *N Engl J Med* 2009;360:1278–81.
- Bosl W, Mandel J, Jonikas M, et al. Scalable decision support at the point of care: a substitutable electronic health record app for monitoring medication adherence. *Interact J Med Res* 2013;2:e13.
- Board of Scientific Advisors Ad Hoc Working Group. An assessment of the impact of the NCI cancer biomedical informatics grid (caBIG®). 2011. <http://deainfo.nci.nih.gov/advisory/bsa/bsa0311/caBIGfinalReport.pdf> (accessed Mar 2014).
- Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. *J Am Med Inform Assoc* 2012;19:181–5.
- Adida B, Sanyal A, Zabak S, et al. Indivo x: developing a fully substitutable personally controlled health record platform. *AMIA Annu Symp Proc* 2010;2010:6–10.
- Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. *J Am Med Assoc* 2008;15:1–7.
- Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
- Obeid JS, McGraw CA, Minor BL, et al. Procurement of shared data instruments for Research Electronic Data Capture (REDCap). *J Biomed Inform* 2013;46:259–65.
- Nadkarni PM, Brandt C. Data extraction and ad hoc query of an entity-attribute-value database. *J Am Med Inform Assoc* 1998;5:511–27.
- Johnson SB, Friedman C. Integrating data from natural language processing into a clinical information system. *Proc AMIA Annu Fall Symp* 1996:537–41.
- Kimball R. *The data warehouse toolkit: Practical techniques for building dimensional data warehouses*. New York: John Wiley & Sons, 1996.
- OMOP (Observational Medical Outcomes Partnership) Common Data Model to i2b2 (Informatics for Integrating Bench to Bedside) Transformation Process. 2011. <http://omop.fnih.org/sites/default/files/OMOP%20to%20i2b2%20Transform%20Nov%202011.pdf> (accessed Mar 2014).
- Klann JG, Murphy SN. Computing health quality measures using Informatics for Integrating Biology and the Bedside. *J Med Internet Res* 2013;15:e75.
- Klann JG, Buck MD, Brown J, et al. Query Health: Standards-based, cross-platform population health surveillance. *J Am Med Inform Assoc* 2014;21:650–6.
- Kohane IS, McMurry A, Weber G, et al. The co-morbidity burden of children and young adults with autism spectrum disorders. *PLoS ONE* 2012;7:e33224.
- Patten IS, Rana S, Shahul S, et al. Cardiac angiogenic imbalance leads to peripartum cardiomyopathy. *Nature* 2012;485:333–38.
- Tatonetti NP, Denny JC, Murphy SN, et al. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin Pharmacol Ther* 2011;90:133–42.
- Miksad R. Harnessing the power of SHRINE to identify new cancer trends: Colorectal cancer rising under age 50. Presentation at 2012 National SHRINE Conference, 2012. https://cbmi.med.harvard.edu/shrine2012/presentations/shrine_miksad.pdf.
- Crandall WV, Margolis PA, Kappelman MD, et al. Improved outcomes in a quality improvement collaborative for pediatric inflammatory bowel disease. *Pediatrics* 2012;129:e1030–41.
- Natter MD, Quan J, Ortiz DM, et al. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. *J Am Med Inform Assoc* 2013;20:172–9.
- Ringold S, Beukelman T, Nigrovic PA, et al. Race, ethnicity, and disease outcomes in juvenile idiopathic arthritis: a cross-sectional analysis of the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Registry. *J Rheumatol* 2013;40:936–42.
- Wallace CA, Ilowite NT. *Comparative effectiveness research in pediatric rheumatic diseases: leveraging CARRA for improved child health*. Seattle, WA: Seattle Children's Hospital, NIAMS/NIH, 2010.
- DeWitt EM, Kimura Y, Beukelman T, et al. Consensus treatment plans for new-onset systemic juvenile idiopathic arthritis. *Arthritis Care Res* 2012;64:1001–10.
- Huber AM, Giannini EH, Bowyer SL, et al. Protocols for the initial treatment of moderately severe juvenile dermatomyositis: results of a Children's Arthritis and Rheumatology Research Alliance Consensus Conference. *Arthritis Care Res* 2010;62:219–25.
- Li SC, Torok KS, Pope E, et al. Development of consensus treatment plans for juvenile localized scleroderma: a roadmap toward comparative effectiveness studies in juvenile localized scleroderma. *Arthritis Care Res* 2012;64:1175–85.
- Mina R, von Scheven E, Ardoin SP, et al. Consensus treatment plans for induction therapy of newly diagnosed proliferative lupus nephritis in juvenile systemic lupus erythematosus. *Arthritis Care Res* 2012;64:375–83.
- Crandall W, Kappelman MD, Colletti RB, et al. ImproveCareNow: the development of a pediatric inflammatory bowel disease improvement network. *Inflamm Bowel Dis* 2011;17:450–7.
- Uzuner O. Second i2b2 workshop on natural language processing challenges for clinical records. *AMIA Annu Symp Proc* 2008:1252–3.