Envisioning a Learning Health Care System: The Electronic Primary Care Research Network, A Case Study

Brendan C. Delaney, MD, FRCP, MRCGP[1]

Kevin A. Peterson, MD, MPH[2]

Stuart Speedie, PhD, MED[3]

Adel Taweel, PhD[1,4]

Theodoros N. Arvanitis, PhD[5]

F. D. Richard Hobbs, FRCGP[6]

[1]Department of Primary Care and Public Health Sciences, King's College, London, UK

[2]Department of Family Medicine and Community Health, University of Minnesota Medical School, Saint Paul, Minnesota

[3]Department of Health Informatics, The University of Minnesota, Minneapolis, Minnesota

[4]Department of Informatics, King's College, London, United Kingdom

[5]School of Electronic, Electrical and Computer Engineering, The University of Birmingham, Birmingham, United Kingdom

[6]Department of Primary Care Clinical Sciences, School of Health and Population Sciences, The University of Birmingham, Birmingham, United Kingdom

CORRESPONDING AUTHOR

Brendan C. Delaney, MD, FRCP, MRCGP
Division of Health and
Social Care Research
Kings College London
7th Floor, Capital House
42 Weston St
London SE1 3QD
Brendan.Delaney@kcl.ac.uk

ABSTRACT

PURPOSE The learning health care system refers to the cycle of turning health care data into knowledge, translating that knowledge into practice, and creating new data by means of advanced information technology. The electronic Primary Care Research Network (ePCRN) was a project, funded by the US National Institutes of Health, with the aim to facilitate clinical research using primary care electronic health records (EHRs).

METHODS We identified the requirements necessary to deliver clinical studies via a distributed electronic network linked to EHRs. After we explored a variety of informatics solutions, we constructed a functional prototype of the software. We then explored the barriers to adoption of the prototype software within US practice-based research networks.

RESULTS We developed a system to assist in the identification of eligible cohorts from EHR data. To preserve privacy, counts and flagging were performed remotely, and no data were transferred out of the EHR. A lack of batch export facilities from EHR systems and ambiguities in the coding of clinical data, such as blood pressure, have so far prevented a full-scale deployment. We created an international consortium and a model for sharing further ePCRN development across a variety of ongoing projects in the United States and Europe.

CONCLUSIONS A means of accessing health care data for research is not sufficient in itself to deliver a learning health care system. EHR systems need to use sophisticated tools to capture and preserve rich clinical context in coded data, and business models need to be developed that incentivize all stakeholders from clinicians to vendors to participate in the system.

INTRODUCTION

Clinical practice should be based on a solid foundation of evidence. A health care system that collects data from routine care for research and facilitates the use of evidence to improve care has been defined as a learning health care system.[1] As such, learning health care systems require software systems capable of managing information-intensive clinical work flows.[2] Considerable work has gone into aspects of this digital infrastructure during the past 20 years. Pay-for-performance initiatives combined with electronic health record (EHR) templates to aid coding have been successful in promoting the adoption of treatment guidelines.[3] Computerized prompts and reminder systems have been shown to improve patient safety, enhance preventive activities, and increase adherence to disease management pathways, including prescribing.[4,5] Many of these features are promoted by the US meaningful use criteria for reimbursement of EHR systems.[6] In terms of generating greater research evidence, single EHR systems supporting large academic health sciences centers have been quite

successful in providing support for research (eg, Kaiser Permanente, Mayo Clinic, Geisinger Health System). The development of systems to facilitate research within community practice settings has been absent, however.

Systems to facilitate research within community practice settings is important. Primary care can recruit research subjects that are difficult to find in specialist centers, for example, people at risk of disease, patients with a new diagnosis, or older patients with complex comorbidities.[7]

As primary care is usually based in small community practices, the problems of recognition, recruitment, and follow-up of study subjects, maintaining data quality, and standardization of interventions become important challenges for the successful completion of these studies.[8] Among the principal tasks in any interventional research study is the recruitment of a sufficient number of subjects to constitute a statistically adequate sample size. A considerable effort has been made by researchers and funders in many countries to organize groups of primary care practices into practice-based research networks, with defined organizational structures that primary-care–based studies support.[7,9]

What is missing is the ability to use the EHR and workflow mechanisms to promote research within these networks.[10] EHRs can be used to identify potentially suitable subjects for research or to capture outcome data, or via prompts, they can remind staff to approach potentially eligible subjects.[11] It should also be possible to link separate sites for larger-scale research projects.[12] The development of standards for exchanging clinical information, which also facilitate integration and aggregation of such data, is proceeding rapidly both in Europe and the United States, along with robust Internet technical standards and Web services.[10] A solution to the problems of security, identity, and data discovery would offer the potential to create large networks of integrated systems to support research.[13] The aim of the electronic Primary Care Research Network (ePCRN) project was to facilitate clinical research (especially randomized controlled trials) using primary care EHRs. In this article, we describe the requirements identified, the solutions and informatics approaches adopted, and the lessons learned (Figure 1).

ePCRN

In 2005 the US National Institutes of Health funded the ePCRN project as one of 12 pilot projects for the National Institutes of Health (NIH) Roadmap on developing clinical research.[14] The ePCRN developed a pilot virtual infrastructure for the US Federation of Practice-based Research Networks. Because practice-based research networks consist of small independent practices, scattered geographically, and with no obligation to participate in research, a combination of informatics approaches and grid software engineering was used to incorporate necessary privacy and approvals steps, enable the identification of suitable subjects from electronic health record data, and establish a standards-based approach in managing electronic remote data capture.[15] The project was built collaboratively around open standards and open-source software. The research intended to pilot methods that could be taken up internationally rather than create a specific deployment and concentrated on the following key requirements. In each case we describe the requirement, the solution adopted, and our informatics approach. (A glossary of terms is available as a Supplemental Appendix, at http://www.annfammed.org/content/10/1/54/suppl/DC1.)
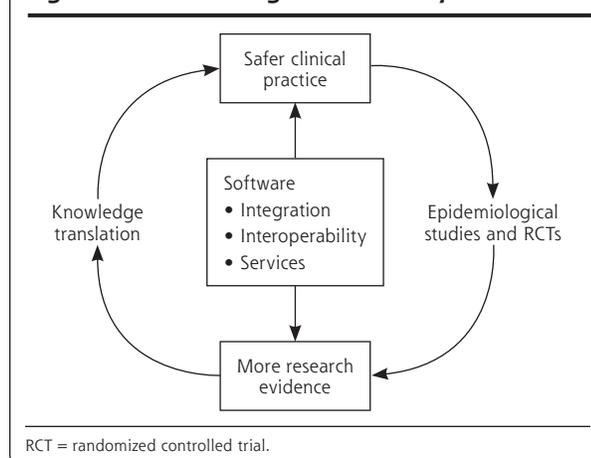
Requirement 1: Identification of Subjects From Clinical Data

Researchers are not allowed to access practice data to identify subjects for research studies. In the United States the Health Insurance Portability and Accountability Act Privacy Rule prevents such access, and in the European Union, interpretation of the Data Protection Directive in many member states has a similar effect.[5]

Solution Adopted

The ePCRN deploys a Web-based workbench (the ePCRN Research Workbench) to enable researchers to specify eligibility criteria and create a distributed query of EHR data. Searches run against registered repositories of clinical data, which can be either an individual EHR or a local data aggregation of EHR data maintained by a local health care organization with a number of practices. Both are known as ePCRN gateways. The queries return only counts of potentially eligible subjects at each site and aggregated by network, protecting the

Figure 1. The learning health care system.

RCT = randomized controlled trial.

privacy of the source data. With appropriate permissions the researchers are able to flag, remotely, subjects identified as eligible for later contact by the patient's clinician.

Informatics Approach

The problem is that different EHR systems use different coding systems (vocabularies) to represent clinical concepts. We adopted the approach of a controlled terminology, where a variety of terminologies (SNOMED [Systematized Nomenclature of Medicine], ICD [*International Classification of Diseases*], and LOINC [Logical Observation Identifiers Names and Codes], RxNorm) are mapped to a common concept, represented by the National Library of Medicine's Unified Medical Language System (UMLS) (http://www.nlm. nih.gov/research/umls/). A Web "service" access to this controlled terminology is integrated into the research workbench. In addition, queries need a standard representation so that the target databases can provide a consistent response. Accordingly, a standard model developed for the ePCRN project (the Primary Care Research Object Model v1.0, PCROM)[16] was translated into a specific query for each target clinical data model. Using a defined data model gives the system wide flexibility in deployment, as the target databases do not need to have a uniform, standard data structure to work with the system so long as it can be mapped into PCROM. Importantly, users are able to browse terminologies, using a simple interface, to find which clinical terms are applicable for their study and explore the effects of different study eligibility criteria on the number of eligible subjects (Figure 2).

Requirement 2: Appropriate Security and Privacy Controls

To allow access to clinical sites, it is essential to have robust security capable of preventing unauthorized access both for users and potential external parties. Patient-level data cannot be extracted from clinical records, however, because the system is not designed to allow that; only counts are provided to researchers.

Solution Adopted

Participating sites are registered with the system, and certificates at both ends of the search authorize only queries originating from and terminating in the ePCRN servers. Access and authentication of users on the ePCRN portal is controlled via secure Web services. The software is not distributed but provided as a service to authenticated users through the portal. The security certificates, site registrations, and certificates are all generated as deployments of the gateway software, such that each deployment can be separately managed. This approach allows the project to benefit

from the open-source software distribution model while providing adequate security.

Informatics Approach

ePCRN uses 2 grid tools, Globus Toolkit 4 (University of Chicago, Chicago, Illinois) and OGSA-DAI (University of Edinburgh, Edinburgh, UK), for the purpose of access to gateways and data transmission (Supplemental Figure 1, available at http://www. annfammed.org/content/10/1/54/suppl/DC1).
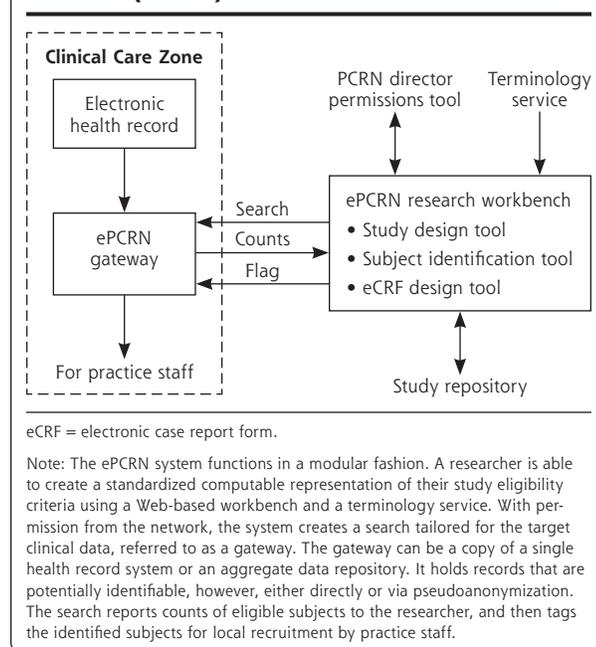
Requirement 3: Collection of Clinical Study Data

The principal problem with electronic remote data collection is that, although Web-based systems allow for direct data entry from remote sites, the data collected may be no more standardized than that on a paper-based case report form. An intelligent system with in-built validation, for example, by within-range rules and flow rules to enable skipping subsections, is necessary. If data are to be passed from one system to another, however, more robust control of meaning is required.

Solution

For data to be interpreted in the same way by a different system, implicit knowledge about the data or metadata needs to be transmitted with that data. ePCRN makes use of this metadata to create a machine-readable document that represents the study in sufficient detail for its execution. This document is used by the system to deploy a variety of forms and linked data-

---

Figure 2. The electronic Primary Care Research Network (ePCRN).

eCRF = electronic case report form.

Note: The ePCRN system functions in a modular fashion. A researcher is able to create a standardized computable representation of their study eligibility criteria using a Web-based workbench and a terminology service. With permission from the network, the system creates a search tailored for the target clinical data, referred to as a gateway. The gateway can be a copy of a single health record system or an aggregate data repository. It holds records that are potentially identifiable, however, either directly or via pseudoanonymization. The search reports counts of eligible subjects to the researcher, and then tags the identified subjects for local recruitment by practice staff.

Table 1. Lessons Learned From the ePCRN (Electronic Primary Care Research Network) Project

Problem	Potential Solution
Extracting coded data from an EHR loses information and may introduce inaccuracies, as different EHR systems do not record some clinical data items consistently, eg, blood pressure. Inaccuracies usually arise where the main terminology adopted within the EHR is a disease-based classification, such as *ICD*, that only partially meets the system's requirement for clinical terminology	Promote the uptake of standard clinical concept representations, such as archetypes based on the CEN/ISO 13606 standard or HL7 detailed clinical models. Rather than a simple term, a more detailed description of the clinical context and value of the clinical data are maintained
Data extraction standards based on a snapshot transfer of EHR data on a given date (such as CCD/CCR) can be unwieldy when data needs to be extracted on a repeated basis, and bulk export facilities are rarely provided by EHR vendors	Promote mandatory adoption of HL7 information exchange standards as an international standard for the exchange of data between and with EHR systems. Include the necessary terminology services to map between different vocabulary systems where necessary
Clinicians and health care providers are rarely incentivized to maintain good data quality on the basis of research use alone	Provide good clinical reasons for data quality and detailed record keeping, such as audit or decision support
Legal and ethical constraints in many countries limit linkage of data and its use for research without consent	Promote international consensus as to how and when data can be linked without consent, and develop systems for managing consent to extraction or study participation across systems. Include adoption of privacy-enhancing technologies, such as filters and third-party linkage systems
Researchers largely unaware of potential benefits offered by electronic systems to support research, and do not therefore create demand for wider deployment	Conduct well-publicized pilot deployments and evaluations

CCD = continuity of care document; CCR = Continuity of Care Record; CEN = European Committee for Standardization; EHR = electronic health record; HL7 = Health Level Seven International; *ICD = International Classification of Diseases*; ISO = International Organization for Standardization.

bases, complete with flow rules, validation, and display preferences. Similar to the subject identification function, the workbench was extended to enable informatics-naive users to manage studies, form collections, and data elements through to project deployment.*

Informatics Approach

Specific data elements represent instances of more general classes. For example, blood pressure can be defined as a standard data element with values for systolic and diastolic pressure in milligrams of mercury at rest and by a protocol-defined method. The International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179 standard defines how repositories of such metadata standards can be established and maintained. ISO/IEC 11179 is a generic standard and requires a domain model to use it. A domain model is a representation of data types and their relationships within a particular field of activity, often expressed using a formal modeling approach. The PCROM defines concepts and associated metadata specific to primary care research and provides the relationships between them.[16] To ensure that primary care is not kept in a silo, unable to link to other domains, the PCROM was also mapped to a less-specific model of clinical research, Biomedical Research Integrated Domain Group (BRIDG) 2.0, produced via a collaboration between the National Cancer Institute caBIG project and the Clinical Data Interchange Standards Consortium (CDISC).[17] Via BRIDG, PCROM is able

to map to the widely used standard for clinical data, the HL7 Reference Information Model (Health Level Seven International, Ann Arbor, Michigan).

Lessons Learned From the Project

The project was funded to provide a proof of concept and explore the feasibility of the proposed solutions to the research requirements. In the United States, the project aimed at a small number of demonstrations in specific practice-based research networks, and planned to use the American Society for Testing and Materials (ASTM) Continuity of Care Record (CCR) standard[18] as a standard representation of information to be extracted from EHRs where these existed and to add data from billing records and laboratory findings to create local repositories of clinical data, held within the clinical domain and called a gateway. Although a number of EHR vendors in the United States had agreed to provide the ability to export patient data in the CCR format, they had limited ability to export records with the correct metadata tags to satisfy the CCR standard. The main reason was that many EHR systems do not use a concept-of-data type, simply placing all observation codes in 1 or 2 fields. In addition, market forces caused concern about exporting an entire database, resulting in the potential for dilution of vendor-specific solutions. Table 1 summarizes the principal lessons from the project, and below we discuss the implications.

Simply providing a means of accessing health care data for research is inadequate as a driver of adoption. The principal problems of data access appear to be due to cultural rather than technical issues. Although it is feasible to create technical solutions to clinical data

*It should be noted that PCROM v1.0 provides the framework by which a blood pressure data element could be defined, rather than reaching down to that level of specification itself.

access across multiple, independent primary care practices for research purposes, it has not happened to any large scale anywhere in the world for the same reason that gateways have failed to be widely deployed in the United States. Because the business model for EHR vendors does not include research facilitation, the extent to which an EHR addresses population-based issues, such as quality improvement and patient safety, depends primarily on such national initiatives as the Quality and Outcomes Framework (QOF) in the United Kingdom and meaningful use in the United States.[18] In particular, the QOF is an important example of what can be done when targets are linked to payments and facilitated by technological development. In the United Kingdom and the United States, contractual and reimbursement initiatives are in place to promote the uptake of EHR systems with specific functionalities, but these do not in our view go far enough to solve the current market failure whereby EHRs are sold as a single solution in a highly fragmented market, rather than being able to work together to provide different functions.

Simple mapping of vocabularies is inadequate in coupling the research and clinical domains. The entire premise of the learning health care system rests on the ability to exchange data between clinical and research systems (system interoperability). Terminology services that map one terminology to another are a necessary but not sufficient approach to this problem. The development of standards in data models is an international problem that requires an international solution. At present, this work is being carried out from the research perspective under the auspices of CDISC, a consortium originally set up to ensure interoperability of data for the commercial clinical trials community, HL7, supported by EHR vendors, and caBIG, until recently supported by the NIH. The needs of the learning health care system require a greater integration of research and clinical data standards, already evident with the mapping of BRIDG v3 to the HL7 Reference Information Model and a much deeper delving into domain concepts in contrast to simple data items. Support for this work, as well as the maintenance of terminology services, needs to be built into the business models that support the meaningful use of EHRs.

EHR systems need to use more sophisticated tools to capture and preserve rich clinical context in coded data. To reuse EHR data directly as research data (or e-source), the recording of clinical data needs to be controlled as precisely as the collection of research data. Most EHRs allow for searching for simple text matches among available codes, and some for traversing a hierarchy of codes where these are part of a classification (such as *ICD-10 [International Classification of Diseases, Tenth Revision]*), but the array of preferred terms and synonyms presented can be confusing. The usual means of dealing

with this diversity of terminology is to assemble chosen codes in templates for use within particular care pathways. An alternative strategy is to use archetypes rather than using the terminology alone. An archetype is a small standardized representation of a piece of clinical data, defined by the European Committee for Standardization/International Organization for Standardization (CEN/ISO) 13606 standard, that allows for the clinical context in which data was recorded to also be stored and maintained, independently of the EHR system, the database format or the terminology used.[19] For example, a blood pressure archetype would record systolic and diastolic pressure correctly, note whether the patient was lying or standing, on or off treatment, and the location of the reading (ambulatory, home, or clinic). Archetypes may also contain coded terminologies for a chosen concept.[20] The widespread use of archetypes in EHR systems may present a solution to the problem of lack of detail in terminologies and provide a means of spreading innovation throughout EHR systems.

Recognizing that further funding would likely be from a variety of funders and addressing specific national or regional priorities, the ePCRN project has created a formal consortium with the aim of continuing the international collaboration around the following:

1. Development and maintenance of the PCROM model to ensure continuing linkage to BRIDG and to establish the model as a standard for primary care research. This effort is now taking place via the European Commission Seventh Framework Programme TRANSFoRm project (http://www.transformproject.eu).

2. Completion of the ePCRN clinical trials workbench through to a controlled demonstration. The single biggest issue for ePCRN was getting appropriately coded data from EHR systems. Ongoing work emphasizes the provision of controlled vocabularies through Web services to facilitate this process.

3. The ability to scale a production-based solution present new challenges with versioning and maintenance. The consortium will provide a mechanism for negotiating and standardizing Web services within the proposed model.

4. Development and maintenance of the ePCRN software as an open-source consortium and to facilitate deployment of portals and gateways based on the system. The key requirement is the active involvement of either EHR vendors or established third-party solutions for data access. Developing a robust business model for achieving this in any network setting is a priority.

The creation of learning health care systems can be achieved only through major international collaboration and multidisciplinary collaboration. A consistent vision of system function and oversight is necessary for establishment of a single shared architecture. Steps that

Table 2. Steps That May Support a Learning Health Care System

Knowledge translation

Develop meaningful use of the EHR to support diagnostic and therapeutic support based on evidence

Develop a computable representation of research evidence and make that available to EHR systems as a Web service

Develop a means of providing diagnostic or therapeutic prompts within an EHR that works across a variety of EHR systems

Turning data into knowledge

Develop a standardized means of using archetypes based on CEN/ISO 13606 within EHR systems

Maintain up-to-date terminology services that can be used by both clinical and research systems

Promote the wide reuse and archiving of defined research data elements via CEN/ISO 11179

Start to link well-defined clinical concepts with well-defined research concepts by mapping archetypes to defined data elements

Create domain specific data models, linking together via BRIDG to enable sharing of archetypes and data elements between domains

Create open-source middleware capable of brokering data exchange and clinical research activities (recruitment, follow-up, etc) between clinical systems

Extend clinical research standards to enable sharing computable representations of clinical study protocols between systems

BRIDG = Biomedical Research Integrated Domain Group; EHR = electronic health record; CEN = European Committee for Standardization; ISO = International Organization for Standardization.

References

1. McGinnis JM. Evidence-based medicine: engineering the Learning Healthcare System. *Stud Health Technol Inform.* 2010;153:145-157.

2. i2010 - A European Information Society for growth and employment [Web site]. 2009. http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm.

3. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet.* 2008;372(9640):728-736.

4. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med.* 2003;163(12):1409-1416.

5. Adams T, Budden M, Hoare C, Sanderson H. Lessons from the central Hampshire electronic health record pilot project: issues of data protection and consent. *BMJ.* 2004;328(7444):871-874.

6. Department of Health and Human Services. Centers for Medicare & Medicaid Services: Medicare and Medicaid Programs; Electronic Health Record Incentive Program, GAO-10-992R, Aug 12, 2010.

7. Sullivan F, Butler C, Cupples M, Kinmonth AL. Primary care research networks in the United Kingdom. *BMJ.* 2007;334(7603):1093-1094.

8. Delaney B. Primary care research in the postmodern world. *Fam Pract.* 2004;21(2):123-124.

9. Pavlic DR. European general practice research network and its mission for strong research in primary health care. *Eur J Gen Pract.* 2007;13(1):46-47.

10. Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf Med.* 2009;48(1):45-54.

11. Bodenheimer T, Grumbach K. Electronic technology: a spark to revitalize primary care? *JAMA.* 2003;290(2):259-264.

12. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ.* 2009;338:b81.

13. Sinnott R, Ajayi O, Stell A, Young A. Towards a virtual anonymisation grid for unified access to remote clinical data. *Stud Health Technol Inform.* 2008;138:90-101.

14. Peterson KA, Fontaine P, Speedie S. The Electronic Primary Care Research Network (ePCRN): a new era in practice-based research. *J Am Board Fam Med.* 2006;19(1):93-97.

15. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. *J Am Med Inform Assoc.* 2010;17(2):169-177.

16. Speedie SM, Taweel A, Sim I, Arvanitis TN, Delaney B, Peterson KA. The Primary Care Research Object Model (PCROM): a computable information model for practice-based primary care research. *J Am Med Inform Assoc.* 2008;15(5):661-670.

17. Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. *J Am Med Inform Assoc.* 2008;15(2):130-137.

18. Waegemann CP, Peters R, Sullivan TE, Tessier C, Waldren SE. Meaningful use: continuity of care with the CCR and CCD. *Telemed J E Health.* 2010;16(3):266-273.

19. Garde S, Knaup P, Schuler T, Hovenga E. Can open EHR archetypes empower multi-centre clinical research? *Stud Health Technol Inform.* 2005;116:971-976.

20. Leslie H. International developments in openEHR archetypes and templates. *HIM J.* 2008;37(1):38-39.

could be supported by funders and industry in supporting the development of a learning health care system are outlined in Table 2. Many of these steps rely on a convergence of aims on the part of academic research, the EHR and health care industry, research organizations, both public and private, and all those responsible for setting requirements for choice and reimbursement of EHR systems. Complex technical solutions can be developed only by an incremental, collaborative, and open-source effort, where innovation in one domain can influence and be reused by others. The experience of the ePCRN project has been one of technical innovation constrained by market limitations. Since 2005, much essential work has been done in both the United States and the European Union, including the development and adoption of meaningful use criteria for reimbursement of EHR systems in the United States, the i2010 strategy of the European Union, and the recent signing of a US-EU declaration on eHealth. The creation of an open market for eHealth applications, which can build and contribute to existing EHRs, would be both good science and sound economic sense.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/content/10/1/54.