

## HEALTH DATA

# Liberating Health Data for Clinical Research Applications

Jessica J. Nadler and Gregory J. Downing\*

Published 10 February 2010; Volume 2 Issue 18 18cm6

**Global investments in health information technology (HIT) in the form of electronic health record systems represent important new strategies to improve the quality and efficiency of health care. These potential benefits of HIT will not be achieved without direct attention to the use of health care information for clinical research purposes. To support clinical research effectively, the systems put in place in the next few years must implement appropriate interoperability, information-sharing policies, and infrastructure that can liberate health data from clinical records.**

## INTRODUCTION

Clinical research is a process of hypothesis testing, skilled observation, information collection, and analysis to inform evidence-based decision-making in health care practices. On one hand, the health care enterprise produces an abundance of data that has the potential to fuel discovery; on the other hand, many obstacles can deter the patient and researcher from its use (1). The concept of a “learning health care system” promotes the generation and application of the best evidence for collaborative health care choices of patient and provider; drives the process of discovery as a natural outgrowth of patient care; and ensures innovation, quality, safety, and value in health care (2). Many aspects of medical and consumer health practices today are not supported by formal guidelines and evidence-based recommendations. On a promising note, however, opportunities for progress are now available.

Advances in communication technologies are often viewed as the lifeblood of innovation and economic growth (3). Global technology investments in electronic information infrastructure will add important capabilities for health care and, ultimately, clinical research. For instance, the Australian government is building a National Broadband Network to bring network capabilities to 90% of Australian homes and businesses. In the United States, the American Recovery and Reinvestment Act (ARRA) authorizes the Federal Communications Commission to create a Na-

tional Broadband Plan, that through the investment of \$7.2 billion “shall seek to ensure that all people of the United States have access to broadband capability” (4). Investments in these and other infrastructure projects are tapping new information resources through advances in transmission speed, multimodal digital integration of data sources, enhanced portability of and access to data through wireless mobile technologies, and enhanced storage and computational capabilities, all of which are rapidly expanding the possibilities for innovative clinical research and technology development. Collectively, these investments are likely to lower the costs of data acquisition and management substantially.

Fiscal, regulatory, and reimbursement policies established through ARRA and the Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH) are designed to improve efficiency and effectiveness in medical practice. These policies may have substantial indirect effects on clinical research by creating information resources that support many types of inquiry not currently feasible. For instance, the anticipated large-scale adoption of electronic health record (EHR) systems, supported by HITECH, presents the opportunity to make health care data accessible through the implementation of common vocabularies, standardized messaging, structured data architecture, and information exchange networks, streamlining the use of these data. Clinical decision support (CDS) is included among the parameters of “meaningful use” that establish regulatory incentives for use of EHR systems in health care (5). All of these capabilities are critically important for the

development of a learning health care system. However, in the absence of new approaches, designs, infrastructure, and resources, the opportunity presented by EHR systems will be missed.

Data captured during the course of clinical care (Fig. 1) can enable clinical research on many topics, including comparative effectiveness, health care delivery practices, biosurveillance and pharmacovigilance, individualized treatment options, and patient stratification by molecular characterization. Knowledge gained from these studies can be delivered back to health care providers through CDS in the EHR they use during patient encounters. This feedback loop, from patient care to clinical research and back to patient care, is enabled by interoperable information systems that can exchange data between clinical research and clinical care.

## AN INFORMATION FRAMEWORK SUPPORTING THE LEARNING HEALTH CARE SYSTEM

In some respects, health care is becoming a knowledge-based service sector. Commonly, the context for this perspective is the use of EHRs as the data resource for accounting practices such as determination of eligibility for services, claims processing, and avoidance of fraudulent use of goods and services. Beyond these financial applications, though, the value of health care data to fuel the knowledge economy is increasingly recognized. Clinical care data are needed to support clinicians and patients in making medical decisions, help organizations and administrations develop guidelines and evidence-based practice, and, in some cases, help payers determine outcomes-based reimbursement policies. Although promising, these diverse applications of clinical care information often provide challenges in the use of these data.

The discovery of knowledge through clinical research and its delivery through clinical decision support are interdependent and must be connected by an electronic infrastructure to facilitate a learning health care system. Currently, data for clinical research and clinical care records are relatively separate for historical and cultural reasons. To connect these systems, the transition to health care based on digital information must incorporate three components: interoperability, policies, and infrastructure.

Office of the Secretary, Hubert H. Humphrey Building, Suite 445F.5, Department of Health and Human Services, Washington, DC 20201, USA.

\*Corresponding author. E-mail: gregory.downing@hhs.gov

Interoperability is defined as “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” (6). It ensures that the meanings of words or data points are consistent in all systems that exchange the data, allowing data to be entered into a system once and subsequently used by other systems without reentry. Currently, interoperability specifications from the Health Information Technology Standards Panel are available for data in EHR systems. Interoperability is on the horizon for molecular information as well. Soon, it will also be possible to connect detailed health outcomes data with information about the cost of care. Ultimately, interoperability will allow the integration of data, including geographic and environmental information, and will enable the investigation of how multiple factors impinge on human health.

Infrastructure is the framework around which information resources are developed. It includes hardware and software resources in addition to tools for innovative data representation and visualization. Networks, such as the National Health Information Network and the Public Health Information Network, can be the highways that carry health information and connect components of this infrastructure (7). Governance and policies for secure gateways to these networks are critical for the confidence necessary for individuals and organizations to share information (8). Trained personnel—with knowledge in both clinical research and informatics, who will innovate process improvements in clinical research—are a final critical infrastructure component.

### INTEROPERABILITY TRANSLATES INTO DATA MOBILITY TO ENABLE RESEARCH

Technical parameters that enable information technology (IT) systems developers and users to seek common applications of various IT systems represent a first step toward advancing the use of EHRs for clinical research. An interoperability specification (IS) was developed to support regulated clinical trials, prospective randomized controlled



**Fig. 1. Information source.** Data from clinical care can contribute to clinical research, which in turn can improve patient care. This feedback loop will work best if data can be easily exchanged.

trials, interventional trials, observational and epidemiological studies, and comparative effectiveness research (9). The IS covers the exchange of a core data set of information from an EHR into clinical research systems. Major benefits are to ensure the safety of research participants, improve data quality by reducing the transcription and reentry of data, and decrease the administrative burden of research for clinicians. Additionally, longitudinal information present in an EHR describes the progression of disease states over time in a way that a finite clinical research protocol may not. The EHR Clinical Research IS specifies terminologies for data elements, ranging from demographic information to medical history to concomitant medications. Although this capability does not define every data element for every study, the ability to exchange these general data elements will streamline clinical research processes.

The IS also articulates the flow of information for three general scenarios: (i) submission to regulatory, public health, and other agencies; (ii) exchange of information from EHRs to registries; and (iii) exchange of information from EHRs in a distributed research network. These scenarios encompass the information transactions necessary for many forms of clinical research.

International involvement in interoperability efforts is critical because clinical research is a global endeavor, carried out by investigators collaborating across countries and continents. A common language for the exchange of information worldwide will facilitate these collaborations and enable clin-

ical research funding to focus on studies, rather than translating and reformatting data. One global effort focused on the development of a common language for clinical information is the International Health Terminology Standards Development Organisation, which aims to promote the accurate and effective exchange of clinical and related health information throughout the world (10).

### POLICIES FOR HEALTH INFORMATION USE

The separation of clinical care and clinical research data has been driven in part by the need for compliance with regulatory parameters. The ability to use

structured data from EHR systems for clinical research may facilitate compliance with harmonized regulatory frameworks both in the United States and abroad. Indeed, there are efforts among regulatory agencies internationally, such as the U.S. Food and Drug Administration and the European Medicines Agency, to move toward greater consistency in regulation through organizations such as the Global Harmonization Task Force for medical devices and the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Until requirements are identical among regulatory agencies around the world, the use of structured data for clinical trials will greatly facilitate the processes through which a new therapy is brought to markets globally. Clinical research must also comply with human-subjects oversight requirements from around the world. The U.S. Department of Health and Human Services Office for Human Research Protections has released an International Compilation of Human Research Protections for 2010 that enumerates the ~1100 laws, regulations, and guidelines that govern human-subjects research in 96 countries, as well as standards from a number of international and regional organizations (11). Compliance with subject registration and consent, review board approvals, and patient safety monitoring at study sites in different countries can be assisted through electronic management of these types of information.

The integration of information from clinical care and clinical research is also the topic of several funding opportunities in Europe. The 2009 European Union Seventh Framework Programme supports the exchange of information between clinical care and clinical research systems. Similarly, the upcoming 2010 European Innovative Medicines Initiative will fund the use of EHR data to support medical research (12).

The Oncology Information Exchange, a collaboration between the UK National Cancer Research Institute and the U.S. National Cancer Institute (NCI), is a portal to access a central register of cancer-related information that allows the investigation of data across multiple sources and types (13). Halfway around the world, BioGrid Australia's mission is to allow "life science research teams to access and share genetic and clinical research data across multiple organisations in an ethically approved and secure way" (14). This platform implements a common language to enable greater use of clinical information. Collectively, these collaborative contributions to infrastructure may soon enable the first international exchanges of fully electronic clinical research data.

### ENABLING INNOVATION FOR CLINICAL INFORMATION APPLICATIONS

Genomic and molecular medicine need highly detailed, accurate, and reproducible patient care information. These fields require the integration of clinical data, such as medical history, interventions, and demographics, with molecular data, such as genotype, gene expression levels, and protein characteristics (15). An understanding of how to use this information is developed through clinical research and can be accessed through CDS.

To combine clinical and molecular data meaningfully, the information in a clinical record needs to be of adequate biological precision. Coding designed for administrative purposes lacks the specificity to define a clinical phenotype for clinical research, which requires data such as time of onset or disease stage. Structured information, using standards outlined in the IS above, combined with natural language processing, provides a level of precision that can support clinical research. This type of approach is being taken by two projects in the United States: the Electronic Medical Records and Genomics Network (known as eMERGE) and In-

formatics for Integrating Biology and the Bedside (known as i2b2) (16).

Genetic and genomic information is an additional element that can be exchanged with the core data set to enhance EHR support of clinical research. This ability depends on controlled vocabulary to describe genes and their expression, polymorphisms, mutations, and DNA methylation states. Individual genetic variation influences a patient's response to clinical interventions and the environment, including pollution or diet. The realization of molecular medicine requires translational research to build an evidence base for these interventions. Currently, data generated by genetic tests may not be entered into the clinical record; the only information allowed is interpretations of a test result based on knowledge at the time of the test.

As detailed in Khoury *et al.*, much of the research in genomics is based in the lab (17). Translating these findings into health care requires an emphasis on the later phases of translational research, namely T2 research to assess the utility of a genomic finding for clinical care, T3 research to move evidence-based guidelines into practice, and T4 research to evaluate outcomes of an intervention based on genomic information in a health delivery system. A current barrier to genomic medicine is the lack of long-term outcomes data on the clinical validity of genetic diagnostics. The T2 through T4 stages of translational research will be streamlined through the use of EHR data (18).

Molecular medicine is based on the idea that therapies address biological mechanisms of pathogenesis. In the future, diagnostic tests will use genetic information to identify patients who will respond to therapy. These diagnostics will be developed alongside therapeutics; their codevelopment presents challenges for the regulation of these pairings (19, 20). Needless to say, as drugs and diagnostics based on molecular information become available for clinical care, the decision-making process for ordering them becomes more complex. Systems that integrate knowledge from multiple sources and display this information in a usable format to health care providers will be critical for the appropriate use of these new interventions. An EHR using the interoperability, policies, and infrastructure described here can provide access to this type of clinical decision support (21, 22).

The NCI's Cancer Bioinformatics Grid (caBIG) is a prototype for a system that supports the learning health care system advancing toward molecular medicine (23). caBIG enables both the use of information gathered in the course of clinical care to support the clinical research enterprise and the validation of research results in clinical care settings. It integrates information from a number of different sources and facilitates analysis in the area of cancer care. The caBIG model is extensible to the health care system in general.

### CONCLUSION

The ability to use clinical information from an EHR for research will facilitate the process by which basic discoveries are translated into health care practice. This capability is made possible through structured data, with appropriate flexibility to accommodate scientific advances. A common language for the exchange of data will streamline clinical research processes. The result of implementing these data standards and architecture will be the greater data consistency and availability that are necessary for the adoption of molecular medicine. Molecular information, connected with clinical phenotype information from the core data set, will provide a basis to stratify patients and disease processes, providing specificity in diagnosis and therapy for both clinical research and care purposes.

Furthermore, these data standards and architecture will create an infrastructure supporting integrative platforms to bring together data types from diverse sources that affect the health of people around the world. These include financial data, such as claims and reimbursement, and environmental information, such as measures of global warming. Looking to the future, investments in broadband connectivity will enable the exchange of bandwidth-intensive clinical data, such as images and genomics information. The application of Web 2.0 technologies that support social networking collaboration and participation gives patients the capability for direct involvement in clinical research, as exemplified by PatientsLikeMe and the Army of Women (24, 25). And finally, next-generation data visualization will enable more people to productively derive knowledge from data.

The steps discussed here provide a framework for innovation to develop and deliver new knowledge and technologies in

health care by opening access to previously inaccessible data. Liberating health information allows multidisciplinary groups to bring new perspectives to bear on complex problems, enabling open innovation in research. Realization of this potential, however, requires the implementation of appropriate interoperability, policies, and infrastructure that support both clinical care and clinical research.

## REFERENCES AND NOTES

1. L. M. Etheredge, A rapid-learning health system. *Health Aff. (Millwood)* **26**, w107–w118 (2007).
2. L. Olsen, D. Aisner, J. M. McGinnis, *The Learning Health-care System: Workshop Summary*, Institute of Medicine, Roundtable on Evidence-Based Medicine (National Academies Press, Washington, DC, 2007).
3. R. D. Atkinson, S. M. Andes, in *The Atlantic Century: Benchmarking EU & U.S. Innovation and Competitiveness* (The Information Technology and Innovation Foundation, Washington, DC, 2009), pp. 17–28.
4. American Recovery and Reinvestment Act (U.S. Congress, 2009).
5. D. Blumenthal, Stimulating the adoption of health information technology. *N. Engl. J. Med.* **360**, 1477–1479 (2009).
6. International Communications Satellite Systems Conference Committee, *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*, 610 (Institute of Electrical and Electronics Engineers, New York, 1990).
7. A. J. McMurry, C. A. Gilbert, B. Y. Reis, H. C. Chueh, I. S. Kohane, K. D. Mandl, A self-scaling, distributed information architecture for public health, research, and clinical care. *J. Am. Med. Inform. Assoc.* **14**, 527–533 (2007).
8. M. D. Greenberg, M. S. Ridgely, R. J. Hillestad, Crossed wires: How yesterday's privacy rules might undercut tomorrow's nationwide health information network. *Health Aff. (Millwood)* **28**, 450–452 (2009).
9. Healthcare Information Technology Standards Panel (HITSP), *HITSP Clinical Research Interoperability Specification* (HITSP, New York, 2010).
10. International Health Terminology Standards Development Organisation (<http://www.ihtsdo.org/about-ihtsdo>).
11. Office for Human Research Protections (OHRP), *International Compilation of Human Research Protections* (OHRP, U.S. Department of Health and Human Services, Rockville, MD, 2010).
12. *Innovative Medicines Initiative 2009* (available at [http://imi.europa.eu/docs/imi-ju-2009-call-topics-14092009\\_en.pdf](http://imi.europa.eu/docs/imi-ju-2009-call-topics-14092009_en.pdf)).
13. National Cancer Research Institute Oncology Information Exchange (<http://www.ncri-onix.org.uk/portal/#S1>).
14. BioGrid Australia (<http://www.biogrid.org.au/>).
15. G. S. Ginsburg, H. F. Willard, Genomic and personalized medicine: Foundations and applications. *Transl. Res.* **154**, 277–287 (2009).
16. S. Murphy, S. Churchill, L. Bry, H. Chueh, S. Weiss, R. Lazarus, Q. Zeng, A. Dubey, V. Gainer, M. Mendis, J. Glaser, I. Kohane, Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res.* **19**, 1675–1681 (2009).
17. M. J. Khoury, M. Gwinn, P. W. Yoon, N. Dowling, C. A. Moore, L. Bradley, The continuum of translation research in genomic medicine: How can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* **9**, 665–674 (2007).
18. F. H. Wehbe, S. H. Brown, P. P. Massion, C. S. Gadd, D. R. Masys, C. F. Aliferis, A novel information retrieval model for high-throughput molecular medicine modalities. *Cancer Inform.* **8**, 1–17 (2009).
19. L. Hinman, B. Spear, Z. Tsuchihashi, J. Kelly, P. Bross, F. Goodsaid, F. Kalush, Drug-diagnostic codevelopment strategies: FDA and industry dialog at the 4th FDA/DIA/PhRMA/PWG/BIO Pharmacogenomics Workshop. *Pharmacogenomics* **10**, 127–136 (2009).
20. S. E. Taube, G. M. Clark, J. E. Dancesy, L. M. McShane, C. C. Sigman, S. I. Gutman, A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl. Cancer Inst.* **101**, 1453–1463 (2009).
21. G. J. Downing, S. N. Boyle, K. M. Brinner, J. A. Osheroff, Information management to enable personalized medicine: Stakeholder roles in building clinical decision support. *BMC Med. Inform. Decis. Mak.* **9**, 44 (2009).
22. A. Wright, D. F. Sittig, SANDS: An architecture for clinical decision support in a National Health Information Network. *AMIA Annu. Symp. Proc.* 816–820 (2007).
23. K. H. Buetow, J. Niederhuber, Infrastructure for a learning health care system: CaBIG. *Health Aff. (Millwood)* **28**, 923–924, author reply 924–925 (2009).
24. PatientsLikeMe (<http://www.patientslikeme.com/>).
25. Army of Women (<http://www.armyofwomen.org/>).
26. Competing interests: J. Nadler is an employee of Deloitte Consulting.

10.1126/scitranslmed.3000764

**Citation:** J. J. Nadler, G. J. Downing, Liberating health data for clinical research applications. *Sci. Transl. Med.* **2**, 18cm6 (2010).

# Science Translational Medicine

## Liberating Health Data for Clinical Research Applications

Jessica J. Nadler and Gregory J. Downing

*Sci Transl Med* **2**, 18cm618cm6.  
DOI: 10.1126/scitranslmed.3000764

<b>ARTICLE TOOLS</b>	<a href="http://stm.sciencemag.org/content/2/18/18cm6">http://stm.sciencemag.org/content/2/18/18cm6</a>
<b>RELATED CONTENT</b>	<a href="http://stm.sciencemag.org/content/scitransmed/5/179/179lr1.full">http://stm.sciencemag.org/content/scitransmed/5/179/179lr1.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/5/179/179le1.full">http://stm.sciencemag.org/content/scitransmed/5/179/179le1.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/4/165/165cm15.full">http://stm.sciencemag.org/content/scitransmed/4/165/165cm15.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/4/162/162fs41.full">http://stm.sciencemag.org/content/scitransmed/4/162/162fs41.full</a>
<b>REFERENCES</b>	This article cites 12 articles, 3 of which you can access for free <a href="http://stm.sciencemag.org/content/2/18/18cm6#BIBL">http://stm.sciencemag.org/content/2/18/18cm6#BIBL</a>
<b>PERMISSIONS</b>	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

Use of this article is subject to the [Terms of Service](#)

---

*Science Translational Medicine* (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS.

Copyright © 2010, American Association for the Advancement of Science