

APPLIED CLINICAL TRIALS

Volume 19, Number 5 May 2010

YOUR PEER-REVIEWED GUIDE TO GLOBAL CLINICAL TRIALS MANAGEMENT appliedclinicaltrials.com

Data Integration



DIETER SPANNKNEBEL/GETTY IMAGES

John S. McIlwain

Data Standards Harmonization

A look at the past, present, and future of interoperability across the clinical research spectrum.

Some experts have argued that there will never be universal standards of interoperability within clinical research, as well as between systems that support research and patient care.

To be sure, there's a long way to go and even if such all-encompassing standards were accepted within the research community, implementing them in the broader context of patient care is an order of magnitude more difficult. Yet standards efforts are making headway and there is cause for optimism, thanks to the efforts of standards-setting groups like the Clinical Data Interchange Standards Consortium (CDISC) and Health Language Seven (HL7), as well as the momentum in medical records adoption.

Interoperability implies that disparate systems—for purposes of this discussion, principally systems used by sponsors and sites—exchange information as an integrated whole. At the investigator site, a constellation of systems impact medical research. These include laboratory, pharmacy, radiology, and even billing systems. Most notable among these is the electronic medical record (EMR). Even with its protracted history of adoption, over the past

few years EMR adoption for patient care has started to progress more rapidly at medical centers and research institutions throughout the United States. Western Europe is even further along. At the same time, there are information technologies (IT) such as electronic data capture (EDC) systems and clinical trials management systems (CTMS) helping investigators and sponsors manage clinical research data.

Despite the similarity of data collected for clinical research and patient care, communication between these two classes of systems has remained hindered for a number of reasons. These include slow adoption of technology, a lack of sophistication in some software applications, and integration priorities on the patient care side in which research is often, at best, a secondary consideration. Most significantly, communication is hindered because systems use different standards—most commonly in the United States variations of HL7 version 2 for patient care and CDISC for research—or no standards at all. As a result, data entered in one system cannot be re-used by the other, resulting in substantial amounts of work duplication.



EMR: A brief history

The distinction between the EMR and the electronic health record (EHR) is important to note, as the terms are often confused. The EMR is an individual organization's electronic medical record for patients and it is the legal record created and maintained within those individual hospitals and clinic environments. The EHR is the aggregate of electronic health-related patient information that is created and gathered across multiple care delivery organizations.¹

The idea for an EMR was born almost 50 years ago when Lawrence L. Weed, MD, first articulated the concept of computerized or electronic medical records. The system he described would automate and reorganize patient medical records to enhance the utility of patient data and benefit research, which would ultimately lead to improved patient care. Weed's work led to a collaborative effort between physicians and information technology experts, and his initiative gained ground in 1967 at the University of Vermont with the Problem Oriented Medical Information Systems (PROMIS) Laboratory group.

The group's efforts led to the development of the problem-oriented medical record or POMR, and its first use was in a medical ward of the Medical Center Hospital of Vermont in 1970. Over the next few years, drug information elements were added to the core program, allowing physicians to check for drug actions, dosages, side effects, allergies, and interactions. At the same time, diagnostic and treatment plans for over 600 common medical problems were devised. During the 1970s and 1980s, several electronic medical record systems were developed and further refined by various academic medical centers and research institutions. General advancements in computer and diagnostic applications during the 1990s allowed increasing complexity, and medical practices throughout the United States also began to adopt the EMR.²

Yet the rate of medical group and hospital adoption of the EMR has been notably slow. In his March 12, 2009, *New England Journal of Medicine* analysis, Robert Steinbrook, MD, states, "perhaps only 17% of United States physicians and 8% to 10% of United States hospitals have at least a basic electronic health record system."³ This is changing rapidly though, as many hospitals and clinics throughout the United States have made significant investments in medical records over the last three to five years. The adoption rate is likely to significantly increase over the next few years as a result of these investments, along with the American Recovery and Reinvestment Act (ARRA) funding.⁴

The rate of medical group and hospital adoption of the EMR has been notably slow.

ARRA offers a multi-year series of incentive payments to providers and hospitals for the meaningful use of certified EHR technology. The total amount of payments has been projected to be \$34 billion by the Congressional Budget Office. And according to the Certification Commission for Health Information Technology (CCHIT), 200+ EHR products, representing more than 75% of the market, were certified by mid-2009.⁵ This stimulus helps set the stage for the creation of rich clinical data sets in digital form that can be used for a variety of medical research purposes, though much work remains in defining and implementing standards to make such clinical data most useful.

A July 2002 Velos white paper identified the close functional overlap between systems for patient care and systems for research.⁶ The paper noted: "from a systems perspective, the modules required for clinical research are identical, in construct, to the systems required for patient care,"⁶ and included Table 1 by way of demonstration.

Diverse data standards

Experts have known for quite some time that it's both possible and intuitively sensible for systems that support patient care and research to be part of the integrated whole, going all the way to Dr. Weed's work in the 1960s. So why hasn't this happened? One major impediment, and the principal topic of this article, is that the two classes of systems don't use common data standards. But as aptly proposed in his October 2009 *Applied Clinical Trials* article, Wayne R. Kubick states that pursuit of those standards is beneficial, in spite of the long distance toward full semantic interoperability.⁷

Medical vs. Research Systems

Name of Patient Care Module	Name of Clinical Research Module
Treatment or Care Plan Design	Protocol Design
Patient Scheduling	Study Subject Scheduling
Electronic Medical Record	Case Report Form
Clinical Compliance System	Query Management System
HIPAA-Compliant Security and Audit Trail	HIPAA-Compliant Security and Audit Trail & 21 CFR Part 11
Data Interface Engine	Data Interface Engine
JCAHO and other Compliance Reporting	IRB and other Compliance Reporting
Budgeting and Billing	Budgeting and Billing

Source: Velos.

Table 1. The different yet very similar modules used today in patient care and clinical research.

CDISC, which targets standards for research, has been derived primarily from the perspective of FDA submission requirements. To the extent that standards are used in patient care, the most widely used standard is HL7 Version 2 (or earlier), which offers a level of flexibility and variability that does not translate well to clinical research. CDISC and HL7 have come together and formed the HL7 Regulated Clinical Research Information Management Workgroup (HL7-RCRIM) to harmonize the HL7 and CDISC standards. However, this harmonization is occurring around HL7 Version 3, also called HL7 Reference Information Model or RIM—a much more robust standard than HL7 Version 2 but one that is not in wide use today at hospitals and clinics. Unfortunately, upgrading systems from HL7 Version 2 to HL7 Version 3/RIM is not a trivial matter.

HL7 RIM uses an object model and is intended to serve as the foundation of health care interoperability. The semantic interoperability of RIM is based on standard models with bindings to standard vocabularies. Any additional system ideally understands the RIM object models and their vocabulary bindings, and thus can re-use information as needed. Some critics believe that RIM has some insurmountable obstacles and that the standard should be abandoned as an “unworkable paradigm.”⁸ However, others see RIM as representing notable and important progress toward interoperability. Either way, the practical reality is that HL7 RIM is not widely used in hospitals and clinics, so EMRs and clinical research systems remain mostly incompatible.

Steps taken

To further complete this snapshot of current medical research standards efforts, the CDISC standard in widest use is called Study Data Tabulation Model (SDTM), which is a content standard for the regulatory submission of case report form data. Logically enough, this standard is based on FDA submission requirements.

A related effort far along and based on SDTM is Clinical Data Acquisition Standards Harmonization (CDASH), described by the Consortium as “a CDISC-led collaborative initiative to develop the content standard for a minimum set of data collection fields in case report forms.” The idea is to harmonize data collection with data submission. While the CDASH standards are also not harmonized with standards generally in use today at U.S. hospitals and clinics, developing a set of content standards for case report forms is a big step in the right direction and a significant improvement over the current state of affairs.⁹

For CDISC, HL7, and other standards to work together systemically and attain what experts call semantic interoperability requires a common “language” to be used. Such a language that enables multiple health care research data standards to interoperate with one another is under development through the aptly named Biomedical Research Integrated Domain Group (BRIDG).¹⁰

BRIDG seeks to provide semantic interoperability across standards and literally bridge medical records and medical research.

BRIDG is a collaboration of CDISC, the HL7 RCRIM, the FDA, and the The National Cancer Institute (NCI), and it is one of the more exciting interoperability models to have emerged for clinical research because the collaboration seeks to harmonize major existing standards and has key stakeholders at the table. BRIDG essentially seeks to provide semantic interoperability across standards and literally “bridge” medical records and medical research by providing terminology and language standards for all aspects of the research protocol. Many are hopeful that the BRIDG Model initiative will be of great benefit in the EMR interoperability initiatives among stakeholders in clinical research. As an aside, some clinical trial systems were designed to support structured protocol models and have long been able to support models like BRIDG. The impediment is not software development but rather a consensus on standards and challenges related to use and adoption.

To more fully complete our outline of the standards-making activities, EHR-S is a standard recently published by HL7 that includes a specification of the functional requirements for regulated clinical research in an electronic health record system. EHR-S is the result of a two-year collaboration by the HL7 EHR Work Group to address a broad list of data protection, regulatory, and ethical research requirements. The HL7 EHR Work Group is comprised of pharmaceutical, biotech, clinical research technology, health care technology, and federal regulatory stakeholders from the United States and the European Union.¹¹

Concurrent to this announcement, HL7 is promoting the EHR Clinical Research Functional Profile, which defines high-level requirements that are critical for using EHR data for regulated clinical research. The profile provides a roadmap for integrating the information environment that must support both the patient care and the downstream research processes. Important functional requirements for secondary data use, such as clinical research, can be integrated into the patient care workflow and documented in the EHR.

NCI, the largest of the National Institutes of Health by funding, is a principal U.S. national research entity focused on standards for medical research. NCI has been instrumental in forwarding standards for terminology through initiatives such as the Cancer Data Standards Registry and Repository (caDSR). This “robust metadata registry is developed and maintained by the NCI Center for Bioinformatics and Information Technology, that stores NCI Common Data Elements and related attributes.”¹² Many NCI-sponsored studies use common terminology from caDSR.

One standards entity not mentioned yet is the Interna-

tional Health Terminology Standards Development Organization (IHTSO). This not-for-profit association develops and promotes the use of a comprehensive multilingual health care terminology known as Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT). The delivery of a standard clinical language for use across global information systems is the significant step promoted by SNOMED CT. The association aims to improve patient care through development of systems to record health care encounters.¹³

HL7 is coordinating efforts with SNOMED CT and other entities to define interoperable vocabularies and semantics. Yet experts recognize that this kind of initiative involves constraining and controlling the unfettered use of vocabularies, defining vocabularies, and grappling with the explosion of terms within vocabularies. Clinical research thought leaders are clinicians with therapies in continuous change due to discoveries and continuous evolution in the medical field. As researchers, they pursue a natural quest to identify new ways of classifying data. This lack of vocabulary standardization and, even more importantly, vocabulary constraints is a critical barrier to overall health care standard harmonization and true interoperability.

It remains to be seen how solutions in this area will be established and deployed. Nonetheless, progress is being made. Extensive application of NCI Common Data Elements in cancer research is one such example.¹⁴ Almost all of the above entities involved in developing standards agree that in order to reach true data interoperability, vocabularies and semantic interoperability must be defined and tightly controlled. Yet there is a lack of definition, agreement, and constraint on existing health care vocabularies within the health care industry as a whole. This prevents technologies from providing constrained (interoperable) semantics.

Impediments to wrestle

To provide a feel for the practical issues that standards developers and implementers have to wrestle with, here are a few examples. To reduce duplicate data entry, one set of data that's widely available electronically on the EMR side is demographics. Perfectly acceptable standards for demographics are generally available today for most research purposes and it's a relatively straightforward matter to extract demographic data from a medical record system into a research system or to connect the systems through Web Services.¹⁵ One would think there would be no reason to have to recapture such data in the clinical trial process for subjects already in the EMR. However, it is quite possible that a subject will have changed addresses between the time their data was captured in an EMR and the time a trial begins.

If a serious adverse event occurs to a subject in the trial, the investigator needs to inform all subjects. As such, good clinical practice is to capture or at least confirm demographic information again in the trial process. What's required in this example is: a two way interface between the

medical record system and research system (or a Web service interface), and preferably the ability of the research system user, such as the study coordinator, to also update the subject's EHR. It's common, however, that the research coordinator may not have the authorization to update the EHR, or may not even be employed by the same institution, in which case the coordinator often does not have access to the patient's EHR. It's also possible that the EHR might need a set of demographic information that differs from the information being collected in the research process.

These issues are all addressable one way or another but the point is that even seemingly simple interfaces like demographics are not quite so simple. They require processes that likely vary from study to study, not to mention department to department within a large research site, and certainly from hospital to hospital and clinic to clinic. Now consider that for any one study a site might only be enrolling a handful of patients. Is it worth having such a site- and study-specific EMR-EDC demographic interface implemented across perhaps dozens of EMRs to save data entry time for, say, a half dozen patients per site? Very likely, it is not.

There are additional unsolved issues defining and interpreting standards that continue to hamper adoption and application of research standards. Consider, for example, that for research sponsored by the NCI there is a requirement for researchers to submit summary race and ethnicity information. The specifications for such submissions are subject to some interpretation such that some sites report somewhat differently. As a result, vendors provide multiple reporting utilities to support the various interpretations and the data is not fully comparable across research sites.

Reasonable objectives and progress

With all these challenges and opportunities for improvement, what can one do today in the areas of system integration and standards application and what is a reasonable set of objectives for a few years from now?

First, while it will be sometime before system interoperability for medical research is widespread or commonplace, there are many examples of organizations applying standards and integrating source systems quite successfully. In the October 2007 issue of *Applied Clinical Trials*, I described a research consortium conducting trials electronically, using terminology standards, where large portions of the research data goes from blood draw to sponsor submission with zero manual data entry.¹⁶ To do this in a way that is economically justifiable, however, requires that the source system interface(s) at the site support many studies.

Indeed, the kind of system integrations occurring today are ones where the interfaces support dozens, hundreds, and even thousands of clinical trials. The most common kinds of clinical interfaces are for demographics and labs, both of which have well enough interface standards established in HL7 Version 2 to be useful for many research purposes.

From implementing a number of such interfaces, we find that most are still, directionally, one way only, with data coming from the EMR to the research system. However, during the last year or so this has begun to change, as organizations conducting research are also starting to look to append research data to the medical record or at least identify patients who are on clinical trials in their electronic medical record systems using a routine patient care. This is valuable for safety and billing reasons, among others.

There is also quite a bit of interface work going on at larger research sites to support administrative needs such as budgeting and billing, as well as regulatory and safety oversight. For budgeting and billing, research sites might download billing rates—sometimes called “charge masters”—from their institution’s financial system. Sites use the information to budget research costs, taking into account standard of care and nonstandard of care items.

This information can then be used to determine whether taking on a study is financially viable. Subsequently, once a study begins and billable events occur, the billing information may be passed from the research system to the financial system for invoicing and collection. Sponsor reimbursements might be recorded in a separate financial system. In this case, an electronic feed can flow back into the research system and be matched against expected research reimbursement. Discrepancies are then tracked down, generally resulting in higher reimbursement.

Most research sites have separate systems for their institutional review boards (IRBs), whose responsibility it is to ensure patient safety and ethical conduct. Data such as protocol information can be captured in a clinical trial management system for human studies. This data is also needed by the IRB for review and approval prior to study initiation. Interfaces are being put in place to avoid such work duplication and to ensure accuracy.

The vendor side

Over the last year or so, many of the leading medical record system vendors have become interested in interfacing with leading research vendors, whereas in the past, it had been hard to get the EMR vendor’s attention. The reasons for this are probably twofold:

- Their customers, many of whom wrote large checks to their EMR vendor over the last five years, have become vocal about supporting interfaces for research purposes
- The EMR vendors have presumably come to view having such interfaces as either a competitive advantage or a competitive necessity.

In any case, this development is good for medical research. The more such activity occurs, the more obvious the need for good standards will become.

Integrating the Healthcare Enterprise (IHE) is an initiative by health care professionals and industry (EMR software vendors, medical equipment vendors, etc.) that promotes the

Just interfacing research systems with EMR systems so appropriate caregivers know a patient is in a clinical trial has great value.

coordinated use of established standards. IHE’s aim is that systems communicate with one another better, are easier to implement, and enable care providers/investigators to use information more effectively. Users and developers of health care IT come together annually and these stakeholders form IHE committees following a four-step process to address interoperability in a variety of clinical domains. One of these is Quality, Research, and Public Health.

CDISC has been very active within the IHE framework, driving integration between source medical record systems and research systems. This partnership has achieved noteworthy successes through its Retrieve for Data Capture integration profile. Of all the organizations tackling clinical system interoperability, IHE is especially hands-on and has achieved tangible results in relatively short order.¹⁷

System interfaces have also been established between systems used at research sites and the systems used with research sponsors, such as the NCI. The race and ethnicity reporting described earlier is one such example. Data from hundreds, and in some cases a thousand or more trials, are routinely downloaded to the NCI directly from the research systems used at sites. Some NCI clinical trial grants require that data be downloaded to an NCI database called the Clinical Data Update System. These submission requirements are also supported electronically.¹⁸

What we can do now

First, continued integration of research administrative functions should proceed relatively unfettered. There are no serious standards issues holding up integration of financial systems, as well as other systems such as patient scheduling and regulatory systems, with research systems used at sites. With a motivated EMR vendor and customer community, there’s every reason to expect more and better system integration and multisystem workflow integration, including two-way and Web-service-based interfaces. Just interfacing research systems with EMR systems so appropriate caregivers know a patient is in a trial has great value and is not widely done today. This is also an important safety issue for hospitals and clinics who conduct medical research.

With respect to Web Services, parts of the health care industry have begun to recognize the value of Web service-based system integration and “cloud computing” for research.¹⁹ These technologies have had widespread use in other industries and consumer businesses for some time. Web service methods of system integration are likely to proliferate in medical research over the coming years—and

these will accelerate system interoperability.

The most common kinds of clinical interfaces being implemented today are for demographics and labs. Local lab data used for patient care are of limited value to industry sponsors on regulatory submission trials where central labs are required. However, policies within many research sites require that local labs be run as well and stored with the investigator's research record. HL7 lab interfaces are in wide use and relatively easy to implement. Such lab interfaces are that much more useful for investigator-initiated studies that do not involve regulatory submissions and for some Phase IV studies and registries that use lab data. Such interfaces are most useful, of course, when they're applied to multiple studies. Organizations with lab data interfaced to their EMR systems can also access lab data for subjects from the EMR through either a file-based interface or a Web service.

Even though it will be a while before patient care and research data standards are both fully harmonized and easily implemented in the patient care setting, there's still value in applying content standards like CDASH and terminology standards like the NCI Common Data Elements right now. There are vendors today that enable customers to automatically generate case report forms using the NCI Common Data Elements. In addition to being a time saver, such standard uses facilitate data analysis across clinical trials.

Conclusion

There's a good way to go before we reach the state where patient care data exchange standards and research data exchange standards are truly harmonized and widely implementable. However, a great deal of progress has been made within the last five years. We have wider agreement on common standards, and a reason to implement them, as a result of the emergence of the EMR and other source systems.

There is every reason to proceed post-haste with some system integrations. Over the past year or two, the sense of urgency for standards-based system integration has increased significantly for research and health care in general. While it will take more than a few years to achieve truly game-changing interoperability, we will soon see resolution to some of the challenges we currently face. For a reader who has evangelized standards for years and fears that retirement will precede accomplishment in this challenging arena, I wouldn't give up hope just yet.

References

1. D. Garets and M. Davis, "Electronic Medical Records vs. Electronic Health Records: Yes, There Is a Difference." *HIMSS Analytics White Paper, Healthcare Information Management Systems Society, HIMSS Analytics*, www.himssanalytics.org, January 2006.
2. K. Pinkerton, "The History of Electronic Medical Records."

<http://ezinearticles.com/?History-Of-Electronic-Medical-Records&id=254240>, July 27, 2006.

3. R. Steinbrook, "Healthcare and the American Recovery and Reinvestment Act," *The New England Journal of Medicine*, 360, 1057-1060, <http://content.nejm.org/cgi/content/full/NEJMp0900665>, March 12, 2009.
4. The American Recovery and Reinvestment Act (ARRA), <http://grants.nih.gov/recovery>.
5. Certification Commission for Health Information Technology (CCHIT), <http://www.cchit.org/about>.
6. J. McIlwain, "How to Implement Investigator-side Clinical Research Systems," White Paper, July 2002, http://www.velos.com/whitepaper_0702.shtml.
7. W.R. Kubick, "The Semantics of Health Care Interoperability," *Applied Clinical Trials*, 18 (10) (2009), <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/Technology+Viewpoint/The-Semantics-of-Health-Care-Interoperability/ArticleStandard/Article/detail/632974>.
8. B. Smith and W. Ceustersc, "HL7 RIM: An Incoherent Standard," From *Studies in Health Technology and Informatics*, 124, 133-138 (2006). Presented at Medical Informatics Europe, Maastricht, August 2006, <http://ontology.buffalo.edu/HL7/doublestandards.pdf>.
9. Clinical Data Acquisition Standards Harmonization (CDASH), <http://www.cdisc.org/cdash>.
10. Biomedical Research Integrated Domain Group (BRIDG), <https://cabig.nci.nih.gov/inventory/infrastructure/bridg>.
11. HL7 Announces EHR-S, http://www.hl7.org/documentcenter/public/pressreleases/HL7_PRESS_20091105.pdf.
12. Cancer Data Standards Registry and Repository (caDSR), <https://cabig.nci.nih.gov/concepts/caDSR>.
13. Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT), http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.
14. NCI Common Data Elements, <https://wiki.nci.nih.gov/display/CommonProjects/caCORE+Vocabulary+and+Common+Data+Elements+Tools;jsessionid=74AE00897FD857112DE15E31789BF572>.
15. Web Services, http://en.wikipedia.org/wiki/Web_services.
16. J. McIlwain, "A to Z Trial Integration," *Applied Clinical Trials Magazine*, 16 (10), <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/IT+Articles/A-to-Z-Trial-Integration/ArticleStandard/Article/detail/462200>.
17. Integrating the Healthcare Enterprise (IHE), <http://www.ihe.net>.
18. Clinical Data Update System, http://ctep.info.nih.gov/protocolDevelopment/electronic_applications/docs/cdus_inst_guidev20.pdf.
19. Cloud Computing, <http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031?page=0,2>.

John S. McIlwain, MBA, is Chairman, President, and Chief Executive Officer of Velos, 2201 Walnut Avenue, Suite 208, Fremont, CA 94538, email: jmcilwain@velos.com.

© Reprinted from Applied Clinical Trials, May 2010 AN ADVANSTAR PUBLICATION Printed in U.S.A.

