Global clinical data interchange standards are here!

Tammy Souza[1], Rebecca Kush[2] and Julie P. Evans[3]

[1] 19 Cronin Ave, Pawcatuck, CT 06379, USA
[2] 1500 Wild Basin Ledge, Austin, TX 78746, USA
[3] 1021 E Street NE, Washington, DC 20002, USA

The Clinical Data Interchange Standards Consortium (CDISC) has succeeded in developing global clinical data interchange standards that are ready for implementation. The various CDISC models are identified and explained in this article as well as how these models work together. In addition to developing the CDISC standards, CDISC is involved actively in many collaborative projects with other organizations as a result of their numerous alliances and partnerships. CDISC standards are supporting not only the pharmaceutical industry but also other initiatives and services in healthcare.

The beginning of CDISC

As computer technology became commonplace and pharmaceutical companies explored the use of electronic submissions, the industry realized that standards are needed to work most efficiently and effectively in this new technological environment. Standards were needed to implement repeatable and efficient practices for collecting and reporting data from clinical trials that, ultimately, end up in regulatory submissions.

Using standards should introduce efficiencies during the evaluation and review of regulatory submissions. Standards facilitate access and orientation to the data and analysis tools, which, in turn, requires less training, improves communication and results in faster review cycles. This was discussed at a public meeting of the US Food and Drug Administration (FDA), details of which are available at (http://www.fda.gov/oc/datacouncil/meetings/levin.pdf). Without the use of standards, the activities associated with clinical development require duplicating efforts and a lot of rework, which translates into more time and higher costs. It is in everyone's best interest to improve patient safety and reduce costs by bringing new treatments to market as quickly as possible, so the value of standardization is readily apparent.

However, in the early years of developing standards for working with clinical data, pharmaceutical companies believed that they would gain a competitive advantage by keeping their standards as proprietary information, which resulted in companies independently creating and implementing their own versions of data standards. It did not take industry long to realize the flaw in this approach. As drug-development became more dependent on sharing data with other organizations, the proliferation of these standards further complicated this already burdened process. Initially the FDA encouraged submission standards to ease the load on FDA reviewers, and improve review times. Eventually, many involved in conducting clinical trials realized that most organizations and many activities across the drug-development process (not only submission) would benefit from a common data-interchange standard for clinical data. This belief gave birth to the Clinical Data Interchange Standards Consortium (CDISC) [1].

CDISC began in 1997 as a volunteer group of 25 individuals that included representatives from the FDA, pharmaceutical companies and vendors. These volunteers came together to create the foundation and structure necessary to develop clinical data standards for use across the pharmaceutical industry. Although many volunteers also work for private, often competing, companies, their participation in CDISC is characterized as a team effort in which team members support all stakeholders in the pharmaceutical industry without promoting any individual vendor, company or organization [2]. From this modest beginning, CDISC has evolved into a not-for-profit organization with hundreds of active participants from major biopharmaceutical companies and academic institutions. Additionally, CDISC has international liaison groups in Europe, Japan and India. As illustrated in Figure 1, the entire clinical trials industry is represented by volunteer

Corresponding authors. Evans, J.P. (jevans@cdisc.org), Kush, R. (rkush@cdisc.org)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.12.012

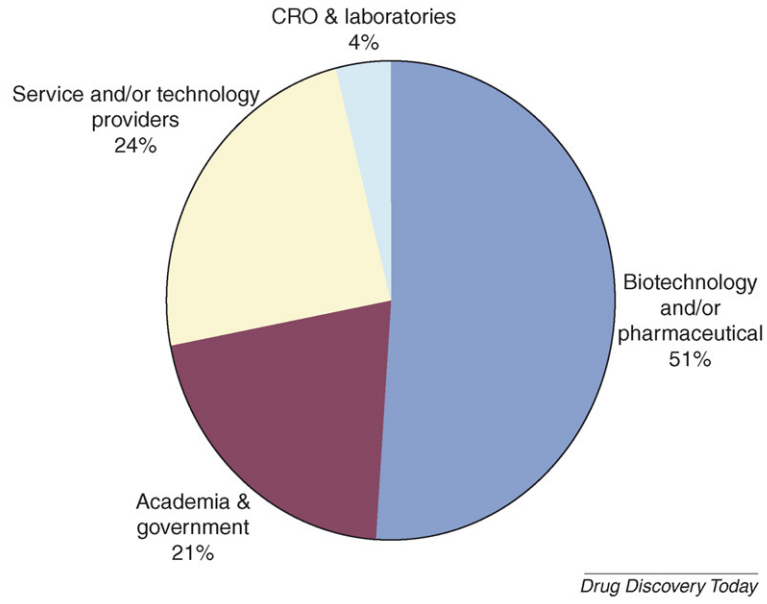

FIGURE 1

CDISC volunteers come from various segments of the Biotech and Pharma industries. This diagram shows the composition of the volunteer teams by company type.

participation on CDISC working teams and groups. The mission of CDISC is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare (http://www.cdisc.org).

Organization and structure of CDISC

The CDISC organization is led by a governing body including a President, Dr Rebecca Kush, who also founded CDISC, and the Board of Directors. Additional leadership comes from the Industry Advisory Board, which comprises representatives from corporate sponsors and benefactors. More detailed information about the CDISC organization is available at their website (http://www.cdisc.org/).

To better understand the organization and structure of CDISC it is helpful to think of the primary activities in the drug-development lifecycle. For example, the basic development lifecycle includes study design, data collection, data exchange, analysis, reporting and archiving of clinical (i.e. human) and non-clinical (i.e. animal) data. CDISC working teams focus on identifying and developing opportunities for data standards to support each of these key activities, and for both clinical and non-clinical data over the course of all study phases.

Five CDISC working teams have produced production standards; these include the Submission Data Standards (SDS) team who developed the Study Data Tabulation Model (SDTM), the Analysis Dataset Model (ADaM) team, the Operational Data Model (ODM) team, the LABORATORY team, and the SEND team who developed the Standards for the Exchange of Non-Clinical Data (now included in SDTM). The ODM team has also developed an ODM-based standard called 'define.xml' which specifies the standard for providing Case Report Tabulations Data Definitions in an XML format for submission to regulatory authorities (e.g. the FDA). The Protocol Representation (PR) group and the Terminology team are working to complete their initial charters. The working teams with production standards have revised their charters to reflect the move toward implementation and enhancements. Figure 2 maps the CDISC standards to the clinical activities they support.

CDISC standards are available freely as downloads from the CDISC website

SDS team

The focus of the SDS team centered initially on the 1999 FDA guidance document on how to submit clinical data in electronic submissions, which is available at (http://www.fda.gov/cber/gdlns/elecgen.htm). Specifically, the original goal of the SDS team was to create a data standards model associated with the 10 safety domains in regulatory submissions identified by the FDA. The SDS model (http://www.cdisc.org/standards/index.html) includes the structure and format associated with these ten safety data domains.

The structure and format of data are also known as metadata, which is sometimes referred to as the 'data about the data'. Metadata defines information about the data, including the names of the data elements, type of data such as character or numeric, and other format information, as discussed by W. Kubick (http://www.fda.gov/oc/datacouncil/meetings/kubick.pdf). Standardizing metadata is one of the central themes to the CDISC mission because this level of standardization, along with controlled terminology, is necessary to accomplish information system interoperability and data exchange. In other words, interoperable information systems should be able to exchange information and to predictably use the information that has been exchanged. Therefore, these information systems need to be based on the same semantics (i.e. meaning) [3].

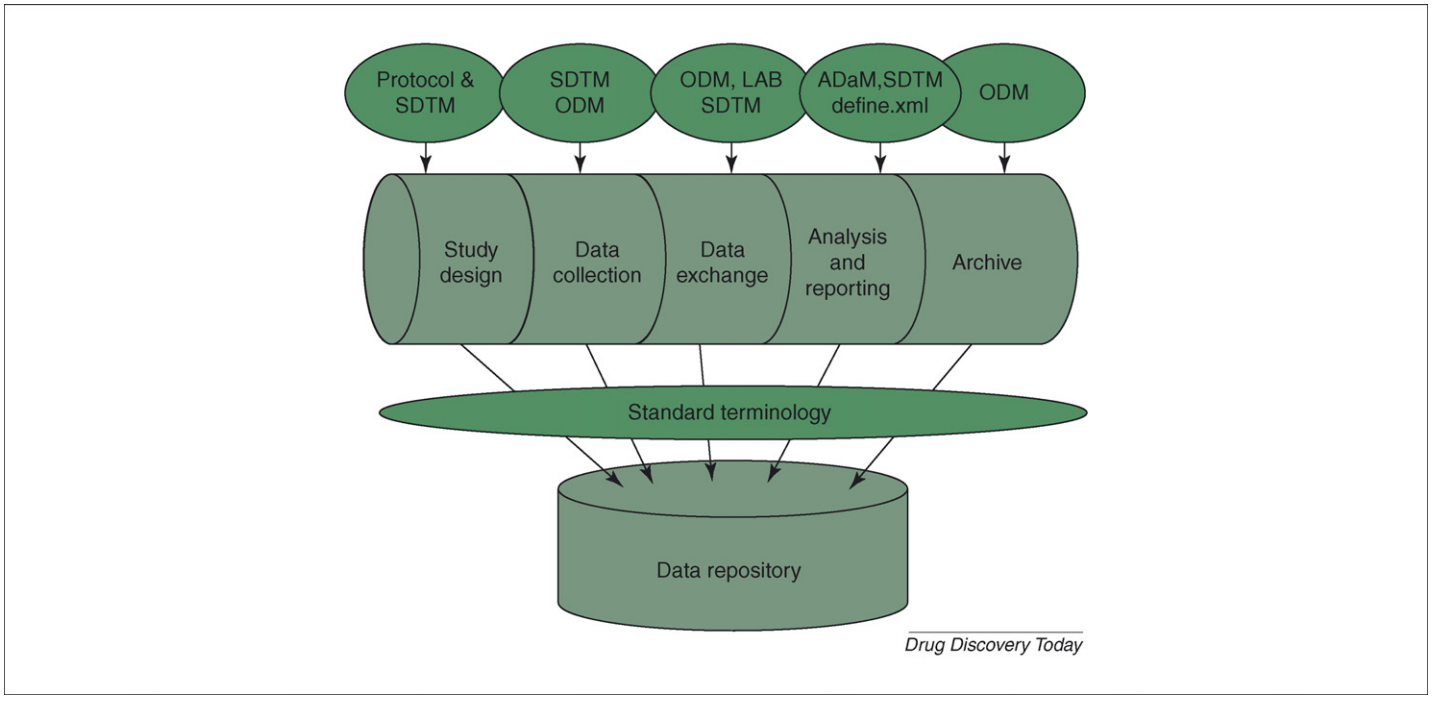

FIGURE 2

This diagram maps the CDISC standards to the clinical activities they support.

Version 1.0 of the SDS model was released in 2000 for review and comment. SDS version 2.0 was released in 2001 and piloted by nine major biopharmaceutical companies and the FDA to ensure the model would facilitate the review of regulatory submissions using standard review tools. The results from the pilot demonstrated that the SDS model needed to include more than the safety data elements to be viable for production use (http://www.lexjansen.com/pharmasug/2003/fdacompliance/fda055.pdf). In response to these results, the SDS team expanded their model to include general structures that could be used for efficacy data. This enhanced model, which became known as SDS version 3.0, was released for piloting in 2003 and the feedback incorporated into SDS version 3.1, which was renamed SDTM.

In addition to the model, an Implementation Guide was developed for using the model in human clinical trials submissions. A major milestone for CDISC was achieved in July 2004 when the FDA announced SDTM as the recommended standard for submitting clinical trial data for regulatory submissions.

ADaM

The ADaM team was formed in 2001 with the primary goal of developing metadata models and examples for analysis datasets used to generate the statistical results for regulatory submissions. ADaM specifically addresses the needs of statistical reviewers. Members of the ADaM team include statisticians, from both industry and the FDA.

The overall principle in designing the analysis datasets and related metadata are that there must be unambiguous communication of the content, source and quality of the datasets that support the statistical analyses performed in a clinical study. This is summarized in a report by D. Christiansen at (http://www.amstat.org/meetings/fdaworkshop/presentations/2005/P07_Christiansen_CDISC.ppt).

The ADaM model builds on the SDS (SDTM) standards. In addition to defining metadata for the analysis datasets, ADaM also includes documentation about the analytic results, such as statistical methods, transformations, assumptions and derivations. This detailed information acts as a 'roadmap' that eases regulatory review. Eventually, regulatory submission reviewers will be able to replicate most analyses, tables, graphs and listings with minimal transformations, and more easily evaluate and subset the data that are used to generate any analysis, table, graph and listing, which will minimize the need for complex programming (http://www.lexjansen.com/pharmasug/2003/fdacompliance/fda014.pdf).

ODM

The original goal of the ODM team was to develop a data model to support the acquisition and exchange of operational data. Operational data consists of clinical data about subjects that is collected from several sources during the course of the clinical trial, before analysis and submission. The ODM also supports the archiving of electronic data (at sites and sponsors) when the trial is complete. This is described in the article by D. Iberson-Hurst (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=103363). With the original goal complete, the ODM team expanded its mission to support the adoption and implementation of the model by stakeholders in the biopharmaceutical market sector and the US regulatory authorities.

The ODM team includes representatives from organizations and stakeholders that contribute to the clinical trial and data-flow process, such as electronic data capture (EDC) vendors and information technology experts.

Highlights of the ODM are that it includes an audit trail, utilizes XML technology, has machine-readable metadata, and contains all the information needed to import clinical data into a database.

The ODM enables a sponsor company and contract research organization (CRO) to exchange data regardless even if they maintain data in different clinical database systems, as described by S. Hume (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=92051). In addition, the ODM provides a study archive function that gives sponsors a hardware- and software-independent method to store study data.

The significance of using XML technology is that XML is a vendor-neutral, platform-independent computer programming language that can be processed by several third-party tools. Historically, the pharmaceutical industry has favoured SAS (a vendor-specific product) for data exchange. The transition to the international use of XML for electronic submissions began with the electronic Common Technical Document (e-CTD). The e-CTD is a specification of electronic standards for regulatory submissions that was developed by the International Conference on Harmonization (ICH), an international regulatory authority. In response to the global transition to XML, SAS has developed tools that support XML and work with ODM (http://www.lexjansen.com/pharmasug/2004/sasinstitute/sas3.pdf).

ODM metadata are likened to an annotated case report form (CRF) in XML format. The ODM XML file includes, in XML terminology, both 'content', meaning the actual values of data and associated metadata, plus 'mark-up', which is the XML tags that define the meaning of the values, discussed in the CDISC White Paper (http://www.cdisc.org/pdf/CDISCReadWriteODM27.pdf). In simpler terms, the ODM XML file includes information such as study visits, types of data (e.g. adverse event, physical exam and efficacy), data attributes (e.g. numeric, character, codelist and length), and the actual data value. The ODM is the XML version of the CRF and contains all the necessary information to transport and archive CRF data (http://www.nthanalytics.com/doc/MTodd%20CDISC%20Presentation2%20Jan%2024%202006.pdf).

Laboratory data model: LAB

The LAB team was formed as a result of the first CDISC customer-requirements survey (http://www.cdisc.org/models/odm/v1.1/customerrequirementsprojectresults_3.pdf), which indicated that the exchange of clinical laboratory data was the highest priority for interchange standards in the pharmaceutical industry. Laboratories that support laboratory data for clinical trials currently use many different interfaces for exchanging data; often, larger laboratories support hundreds of different interfaces for their clients (including multiple different interfaces within a client company). Having a laboratory data standard greatly reduces the time and effort required to maintain different interfaces by decreasing the set-up time between laboratories and their clients, and reducing the amount of programming needed to support laboratory data, discussed at (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=176908).

The LAB team released the first version of the LAB model in 2002. The model included the standard representation of laboratory data generated during the conduct of clinical trials and can be used for the interchange of laboratory data from clinical laboratories to sponsors, vendors and CROs. The LAB model supports laboratory data that is associated with routine testing including chemistry, hematology and urinalysis. The LAB model is flexible because it can be implemented in ASCII, SAS, XML and HL7 V3 message [accredited through the American National Standards Institute (ANSI)], and estimates indicate that it reduces costs by 30–50% (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=192753).

The PR group

The PR group began in 2002 with an undertaking to identify standard data elements (including a subset of those in SDTM) and semantics for standard content of a clinical trial protocol. Standardizing the protocol in this fashion helps to facilitate study design, regulatory compliance, project management, trial conduct and data interchange among consumers and systems, see the webcast at (http://www.bettermanagement.com/seminars/seminar.aspx?l=12162).

The PR Group has published the Standard Protocol Elements (for Regulated Clinical Trials) Version 2.0, which is currently under review. This proposed standard defines the key elements of clinical protocols and is based on the ICH guidance for good clinical practice (http://www.fda.gov/cder/regulatory/ersr/ectd.htm) with emphasis on sections ICH E6, E3 and E9. A machine-readable model for the PR standards is the ultimate goal of this team. This model includes the Trial Design from SDTM.

Terminology team

The CDISC Terminology team was formalized in 2005 and is comprised of representatives from government, academia and pharmaceutical companies. The goal of this team is to identify and/or develop standard controlled terminology and codelists to support the CDISC production standards, beginning with the SDTM. Work is conducted with two principals in mind: (1) ensure a harmonized approach to terminology development; and (2) ensure terminology recommendations suit international needs for global organizations and projects.

The Terminology team published the first set of 32 codelist recommendations, known as SDTM Package-1, that addresses the needs of the SDS/SDTM data model and a forthcoming set of codelist recommendations will be issued as SDTM Package-2 (http://www.cdisc.org/publications/interchange2005/session2/CDISCControlledTerm-BronKisler.pdf). CDISC has a working relationship with the National Cancer Institute (NCI) Enterprise Vocabulary Services to provide the delivery services and environment for production terminology. Through this partnership, CDISC can leverage existing services, expertise and infrastructure that is designed specifically for the development, harmonization, delivery and support, and maintenance of controlled terminology.

Standards for the exchange of non-clinical data (SEND) consortium

SEND began in 2002 and became an official CDISC team in May 2004. SEND focuses on data collected from preclinical toxicology studies. The underlying conceptual model behind the SEND standard is the SDTM (http://www.fda.gov/cder/present/DIA2005/Papoian_SEND.ppt). SEND includes guidance for the organization, structure and format of preclinical data submitted to the FDA. SEND v.2.3 also includes additional data standards guidance for single-dose, repeat-dose and carcinogenicity studies.

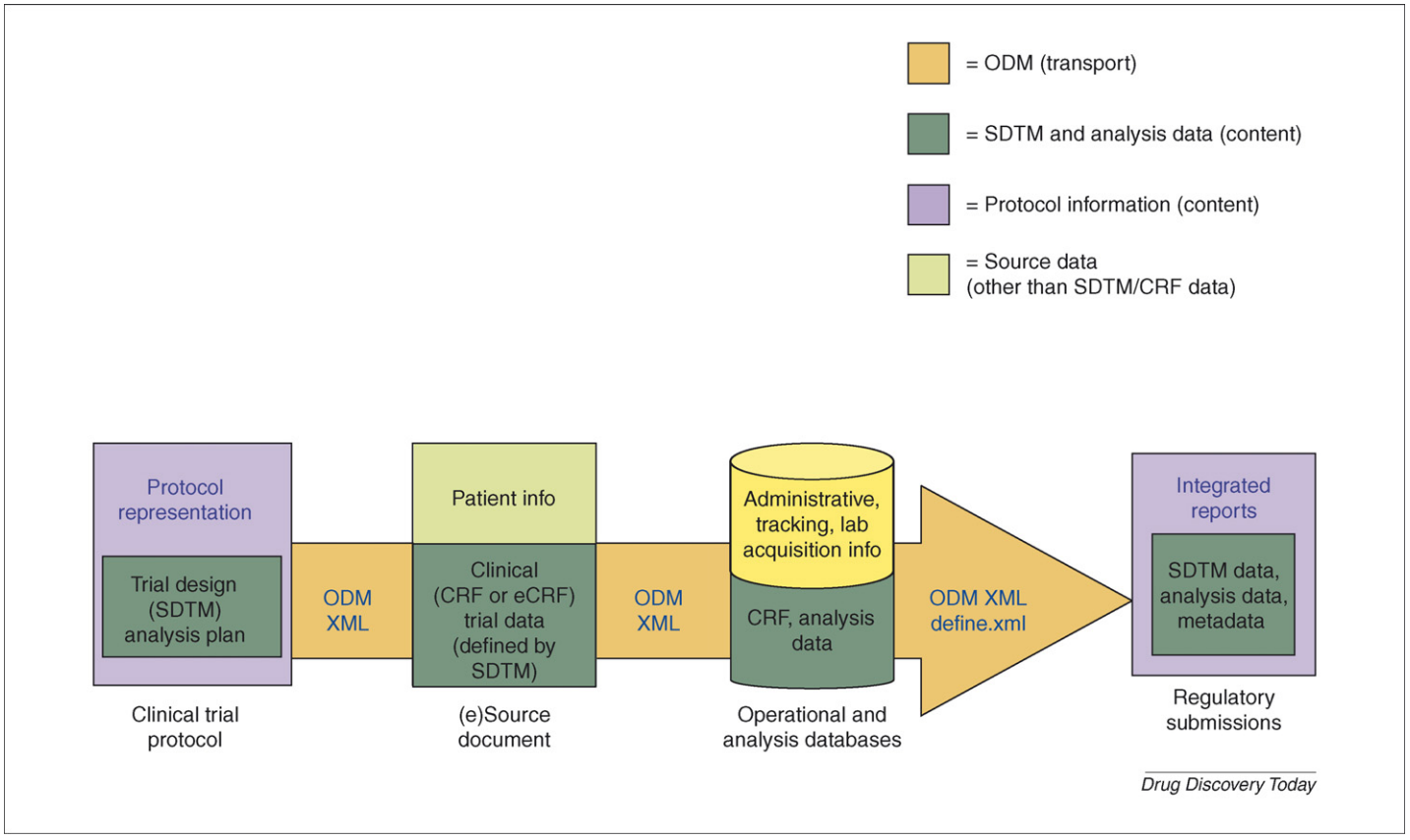

FIGURE 3

The end-to-end data flow across the clinical research process using CDISC standards. The content data is defined by the Protocol, SDTM (including SEND), LAB and ADaM standards. The ODM provides the transport standard that allows the content data to flow seamlessly through the clinical trial process, from source to analysis and reporting to regulatory submission and archive.

Putting it all together

Figure 3 depicts the end-to-end data flow across the clinical research process, using CDISC standards. The content data are defined by the Protocol, SDTM (including SEND), LABORATORY and ADaM standards, and the ODM provides the transport standard that allows the content data to flow seamlessly through the clinical trial process – from source to analysis and reporting to regulatory submission and archive, as described in (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=334571).

Capitalizing on the use of metadata by integrating it into the data domain and analysis dataset creation process can reduce the time and effort, and improve the quality in electronic submissions (http://www2.sas.com/proceedings/sugi29/108-29.pdf).

The global perspective

The mission of CDISC is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. In support of the global aspect of the standards, CDISC groups in Japan and Europe hold annual 'interchanges', or conferences, with presentations, workshops and tutorials of interest to CDISC implementers. CDISC has also met with interested parties in India, see Schedules of International Conferences on the CDISC international website (http://www.cdisc.org/international/index.html). Companies outside the USA wish to know about CDISC so they can implement the standards for their clinical trial processes and do business in the USA. In addition to the annual interchanges, the European CDISC community is also hosting 'CDISC Days' to bring local CDISC users and potential users together, see 'Acktuelles' section of CDISC Europe website (http://www.ecdisc.org/).

Implementing CDISC

After years of development, the CDISC standards are stable and ready to use in production environments. Pharmaceutical companies no longer question whether to implement CDISC standards, but rather how and when they should be implemented. The complexity of clinical data coupled with the numerous technologies involved can makes the implementation of new data standards challenging (http://www.lexjansen.com/pharmasug/2006/handsonworkshops/hw06.pdf). Several pharmaceutical companies are in the process of implementing some components of the CDISC standards, and some have already successfully implemented most of the components of CDISC. According to a survey in 2004, nearly 50% of North American pharmaceutical companies reported utilizing CDISC standards (http://www.cdisc.org/pdf/Interchange2005/02-EuroInterchangeRDKUpdate27April05fin.pdf). An update of this survey is in progress.

There is a general debate over whether to implement CDISC standards at the beginning or end of a clinical project. Choosing to implement at the end requires mapping existing data to the CDISC standards. Companies who decide to map existing data have further decisions to make regarding the timing and starting point for which data/datasets will be mapped. The options, which include mapping clinical trial data before creating analysis datasets, after the analysis datasets are created, and in parallel with the creation of the datasets, are described at (http://www.lexjansen.com/phuse/2005/cd/cd11.pdf).

Although implementing CDISC at the end of a project is a viable business decision, especially for legacy data, there are several considerations to keep in mind. This approach does not upgrade systems to make them compliant with the CDISC standards, the process must be replicated for each project, which means additional costs, and multiple versions of metadata exist (http://www.nthanalytics.com/doc/MTodd%20CDISC%20Presentation2%20Jan%2024%202006.pdf).

Benefits of implementing the CDISC standards include:

(i) Companies with rapid growth based on increasing numbers of acquisitions and/or many compounds in development might find that their old methods for conducting business are no longer practical and scalable. Therefore, implementing CDISC at the onset of a new clinical project will reap benefits throughout the life of the project (http://www.cdisc.org/pdf/StreamliningThirdPartyDataCapturewithODM_MarkWheeldon.pdf).

(ii) The benefits of implementing CDISC at the beginning of a project include having SDTM structures immediately available, streamlined reporting and analysis programs, and there is only one set of metadata. The closer to the beginning of the clinical study process standards are implemented, the greater the return on investment.

(iii) Other benefits include using the ODM standard to automate the production of electronic CRFs and the operational database setup with EDC vendors. One company reports that using ODM reduces study set-up time threefold (http://www.cdisc.org/pdf/StreamliningThirdPartyDataCapturewithODM_MarkWheeldon.pdf) and another states that using CDISC standards improves the quality of data from CROs [4].

Alliances and collaborations

One of the founding principles of CDISC is to collaborate with other organizations, thus, avoiding duplicating efforts and/or developing competing standards. Therefore it is no surprise that CDISC has embraced several successful partnerships. Since the onset of CDISC, the organization has worked closely with the FDA (http://www.fda.gov/), which recommends the SDTM model as the standard for regulatory submissions of clinical trial data. FDA also refers to SDTM in its guidance documentation for implementing the electronic Common Technical Document (http://www.fda.gov/cder/regulatory/ersr/ectd.htm). A recent Federal Register notice (FR Doc 69-73119) recommends that electronic submissions based on standards for study data become regulation (http://ciir.cs.umass.edu/cgi-bin/ua/web_fetch_doc? dataset=ua&db=agendaSpring2003&query=and&doc_id=787).

Other regulatory agencies do not require the submission of data in marketing applications but have shown interest in CDISC. The Protocol Representation Standard is of interest to the European Medicines Evaluation Agency and the Japanese Ministry of Health, Labor and Welfare supports work by the University of Tokyo to explore ways to digitalize clinical trials, which have identified the CDISC standards as a key opportunity in this process, see the CDISC e-Newsletter (http://www.cdisc.org/newsletter/article.asp?issue=200507&n=5).

Another alliance began in March 2001 when CDISC signed an associate charter agreement with Health Level Seven (HL7). HL7 is one of several Standards Developing Organizations that are accredited by ANSI and operate in the healthcare arena (http://www.hl7.org/). Working with HL7 has allowed CDISC to capitalize on its 15 years of experience in developing standards and to begin to harmonize the CDISC standards with those of healthcare, see A Multidisciplinary Approach to Data Standards for Clinical Development –Progress Update, Applied Clinical Trials (http://www.cdisc.org/pdf/ACT5856e.pdf).

CDISC also collaborates with the NCI (http://www.cancer.gov/) and the National Institutes of Health (NIH) (http://www.nih.gov/), who are working with CDISC on the PR group and Terminology team to standardize medical terminology between academia and industry.

In 2004, CDISC initiated a domain analysis model called the Biomedical Research Integrated Domain Group (BRIDG) model. A domain analysis model is software engineering tool that documents information requirements in a structured, clear way so that domain and business experts can verify the requirements are complete. The development of BRIDG became a collaborative project with HL7, NIH, FDA and others. The vision of BRIDG is to create a domain analysis model, using the HL7 development framework in the context of the HL7 Reference Information Model in a way that harmonizes clinical research standards between clinical research and healthcare, see the Next Generation Pharmaceutical website (http://www.ngpharma.com/pastissue/article.asp?art=25518&issue=143). In other words, the goal of BRIDG is to align the CDISC standard (and HL7 standards related to clinical research) in one, overarching standards-based model that bridges the standards between healthcare and clinical research.

In 2006, FDA announced a collaborative group to support the development of data collection standards as part of the Human Subject Protection and Bioresearch Monitoring Initiative, and Critical Path Initiative. CDISC is leading this collaborative group that includes organizations such as the FDA, NIH, American Medical Informatics Association (AMIA), Association of Contract Research Organizations, and Pharmaceutical Research Manufacturing Association.

CDISC also has a memorandum of understanding with the Critical Path Institute and organizational memberships have been exchanged with HL7, AMIA and Healthcare Information and Management Systems Society.

Next steps and overcoming challenges

The CDISC Roadmap (Figure 4) shows the future of CDISC with a single CDISC standard comprising a set of fully integrated, consistent models. Further integration of the existing CDISC

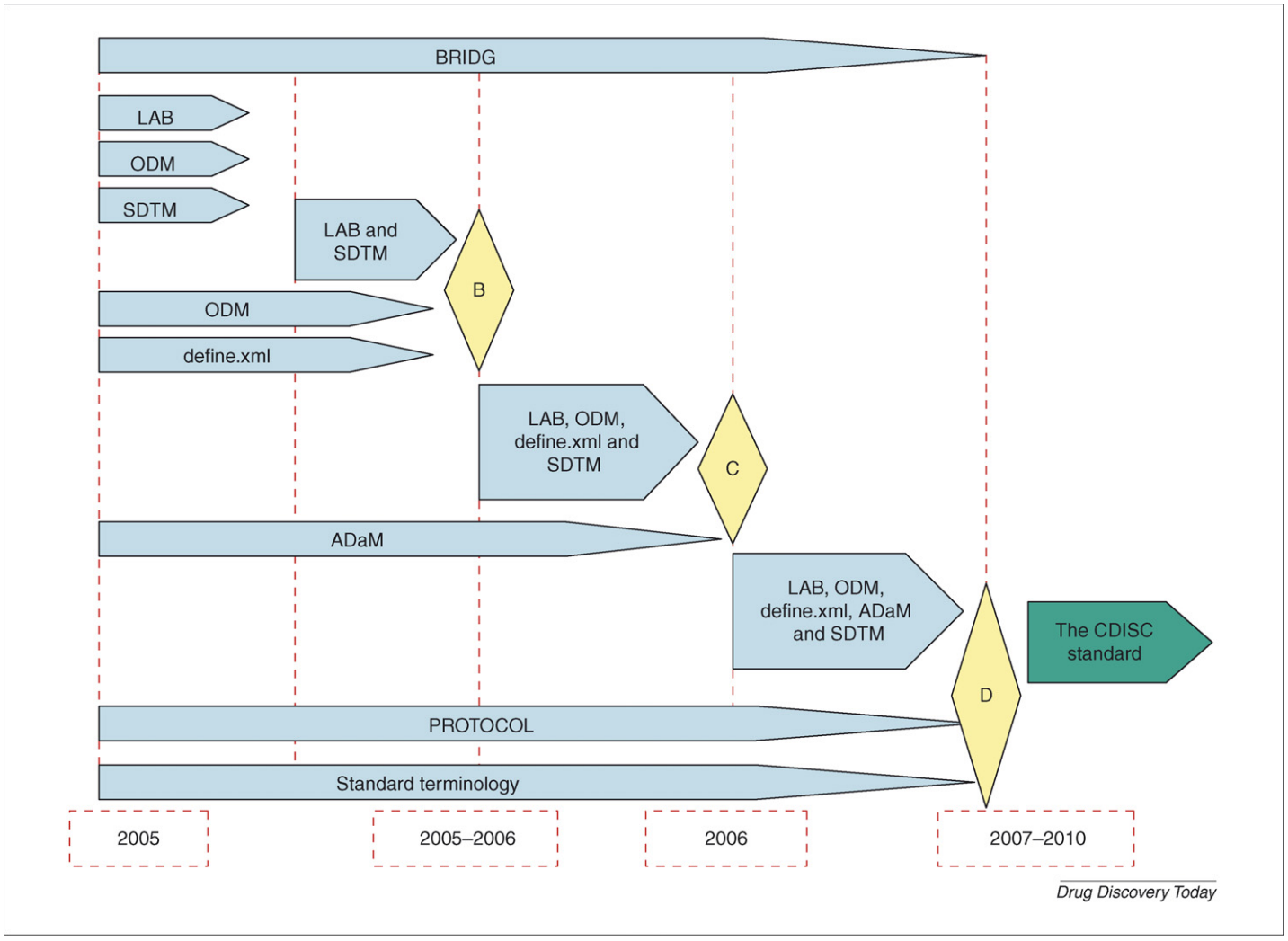

FIGURE 4

The CDISC roadmap shows the future of CDISC with a single CDISC standard that comprises a set of fully integrated models. Further integration of the existing CDISC models is underway to achieve this goal following the timeline proposed. The labeled arrows represent the CDISC standards. The standards are integrated using the BRIDG model and are supported by standard terminology.

models is needed and efforts are underway to achieve this goal, following the proposed timeline. The factors that have been defined by CDISC to gauge whether it has achieved its goals are that: (i) all submissions to the FDA are made using the CDISC standard; (ii) the set of CDISC models are in use across the full life-cycle of clinical trials; and (iii) the CDISC standard is adopted globally.

Although the use of CDISC is growing, there are challenges to overcome. A broader implementation of CDISC standards by data management systems and data collection technology providers is needed to more fully realize the benefits of CDISC. Since the beginning of CDISC, the benefits associated with the regulatory submission process have been well understood; however, it is important to also realize the benefits of CDISC standards for all stakeholders in clinical research, including those working in project management and monitoring roles. These groups have traditionally been less involved in CDISC, but using CDISC-compliant tools will help to make this shift and support the implementation of CDISC standards earlier in the process.

Another challenge for pharmaceutical companies is the concern that CDISC standards are either not stable or might change frequently. To address this concern CDISC has committed to maintaining production standards and making any enhanced versions backward compatible (http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=334571).

An opportunity for improvement involves replacing the outdated technologies used to submit data to the FDA with current technology, such as XML. Through CDISC standards are now established firmly; therefore, technologies can capitalize fully on the value of data standards. This is especially true in terms of further integration with healthcare services.

Concluding remarks

Since its inception, CDISC has worked with industry and the FDA to define data interchange standards and models that support clinical drug development. The STDM, ODM and ADaM models are in production and ready for use. These standards reduce the time, energy, resources and costs associated with the clinical trial and regulatory submission processes. Early implementation of

CDISC standards is crucial to streamlining the overall clinical research process and to improving data quality.

CDISC standards support not only the pharmaceutical industry but also other initiatives and services in healthcare. The partnerships that CDISC has formed with NIH, NCI, and the BRIDG stakeholders will help promote the global use of CDISC standards and, ultimately, contribute to the assurance of safe, effective treatments for patients.

References

1 Kush, R. (2003) *The world of standards for clinical research,* Business Briefing: Pharmatech (5th edn), pp. 100–102

2 Kush, R. (2002) A multidisciplinary approach to data standards for clinical development –progress update. *Applied Clinical Trials* 11, 35–44

3 IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries (610-1990)

4 Kush, R. and Hardison, D. (2004) How necessary are data standards. *Script Magazine*, May 29–32