

CDISC Standards from the Start

Dave Ibersen-Hurst and AJ de Montjoie at CDISC discuss standards development over a bottle of wine

CDISC is the leading standards development organisation for the medical research domain, throughout the world. With over 240 member organisations, from global pharmaceutical companies to small consultancies, CDISC encourages all of its members to become involved with the development of the standards. CDISC has seen increasing involvement outside its traditional audiences in the US, Europe and Japan, with a now established coordinating group in China and others being developed in India and Korea, as well as reaching out to new areas such as Australia, Brazil and Russia. With such interest comes the additional responsibility of ensuring these new communities are informed of the latest developments at CDISC. So what is new? How is CDISC preparing for the next decade? How is the organisation taking the experiences of the past 10 years into the future? We begin the journey back in 2005, reflecting on the challenges of combining technical content with experience and understanding.

WINE AND METADATA

In 2005, CDISC held their annual European conference, the CDISC Interchange, in Paris. Over lunch one day, Dave was asked to explain the define.xml standard. There were, as we were in France, two bottles of wine on the table, one red and one white. Dave picked one up and asked the inquisitor to think about the bottle of wine. The content is the important item, the wine itself. The choice of red or white is usually an easy one dictated by circumstances: the food, the time of day or the occasion. But, given a selection of several white or red wines, further information is needed for one to make a considered choice. So the wine is the important content but, to open the right bottle, we need the information written on the label to establish the facts about that important content: the vineyard, the age of the wine, the grapes, perhaps the country of origin, and so on. Dave told the inquisitor to think of the label as the metadata for the

wine, providing the important information needed to understand the contents of the bottle, the wine. He asked them to think of the bottle merely as a container to transport the wine and, of course, the wine as the all-important content. Returning to define.xml, define is the equivalent of the wine's label for a package of SDTM/ADaM data (the important content) held within a SAS transport file (the bottle, the container). Define tells us, and more importantly, it allows a computer to be told, what the contents of a given package of data are. Of course, that package of data could be part of a submission to the FDA.

Now imagine entering a room full of racks of wine, all of which are different. Without the labels on the bottles, the task of finding the desired bottle becomes difficult, if not impossible.

The metadata on the label, or whatever mechanism is used to record the contents of the racks and of each

bottle, becomes all important; without it we will waste an immense amount of time and rely on people's memory of where bottles were placed. This is hardly a good idea – imagine all the wasted wine!

Metadata for clinical research data is no less important, and it is →

The new SHARE repository will provide for easier access to the existing CDISC standards and any subsequent development to those standards. In particular, it will allow organisations to access the expansion of the Study Data Tabulation Model (SDTM) and the Clinical Data Acquisition Standards Harmonisation (CDASH) standards as new therapeutic domains are developed. By having this managed centrally and available electronically, it will decrease costs for those using the standards.

something that will become central to CDISC's work in 2010. Our data is our important content (the wine), but we constantly need to understand it; without understanding we cannot use it effectively, or worse, we may use it incorrectly. In the same way that we need high quality data, we also need high quality metadata.

In terms of the CDISC standards, these metadata are provided in the manner of variable and associated terminology definitions. At present, these metadata are detailed within documents that are delivered to users in the form of a PDF document. The metadata contained therein needs to be transposed into a machine-readable form, if it is to be employed within an electronic system. This is potentially being performed many times across many organisations involved in clinical research; it is not efficient and is prone to error. What CDISC wishes to do in 2010 is to build an electronic repository of metadata that can be easily and readily consumed by organisations and their computer applications. The CDISC Shared Health and Clinical Research Electronic library (SHARE) is the project tasked with delivering this electronic metadata repository to the clinical research community. In collaboration with the National Cancer Institute in the US, who will provide the technology base for the project, CDISC hopes that the first incarnation will be available to users early in 2011.

A SHARED VISION

The new SHARE repository will provide for easier access to the existing CDISC standards and any subsequent development to those standards. In particular, it will allow organisations to access the expansion of the Study Data Tabulation Model (SDTM) and the Clinical Data Acquisition Standards Harmonisation (CDASH) standards as new therapeutic domains are

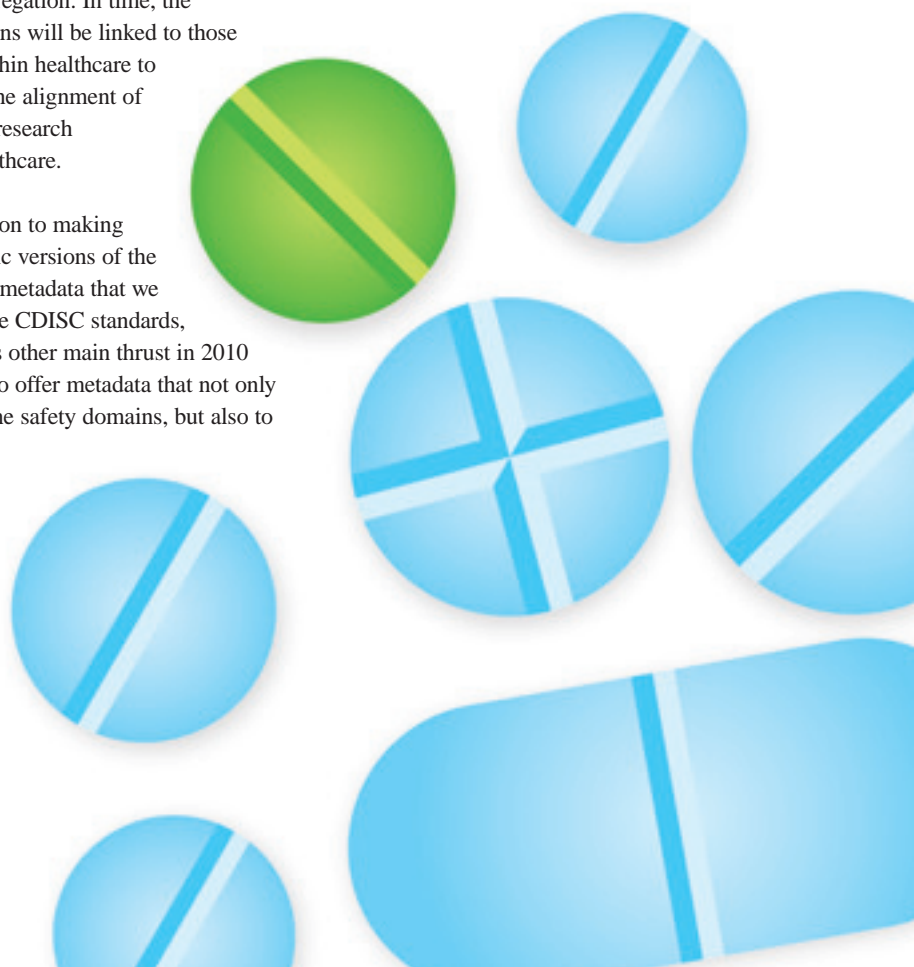
developed. By having this managed centrally and available electronically, it will decrease costs for those using the standards. The repository will also provide for a consistent approach to standards definition. This will, in turn, increase the quality of the standards by allowing for unambiguous definitions to be defined once, but used many times in a consistent fashion across all of the standards as needed. One thing that CDISC wishes to achieve with its standards is moving from an implicit world, where metadata definitions and rules of use are open to interpretation, to an explicit one, where definitions and use are very clearly determined.

SHARE will give users easier access – and applications machine-readable access – to the one 'gold' data standard that will facilitate greater data re-use and aggregation. In time, the definitions will be linked to those used within healthcare to permit the alignment of clinical research and healthcare.

In addition to making electronic versions of the existing metadata that we see in the CDISC standards, CDISC's other main thrust in 2010 will be to offer metadata that not only covers the safety domains, but also to

begin to provide metadata – and thus standards – that cover the efficacy data, increasing the breadth of data encompassed by the CDISC standards. Again, these metadata would be delivered using the SHARE metadata repository. One major feature of the efficacy data development is that it will be tackled on a therapeutic area basis in conjunction with those active in relevant areas and with the backing of the FDA. The areas we are currently working on are cardiovascular disease, diabetes, Parkinson's disease, Alzheimer's disease, polycystic kidney disease and tuberculosis. These will be augmented by additional areas that will be announced in due course.

In order to bring these standards together, we need to ensure we have a framework within which we can align everything, and that framework is the Biomedical Research Integrated Domain Group model, or →



BRIDG. Imagine trying to complete a complex jigsaw puzzle without anything other than the pieces themselves. We could probably complete it, but it would take time. However, with access to the picture on the front of the jigsaw box, everything becomes a lot clearer. BRIDG is our picture; it allows us to position each piece of information and relate it to all the others. So not only will we have each piece of metadata within the SHARE repository, we will also have the BRIDG model that will allow us to relate each piece of metadata to all of the others and relate those small groups of items that need to be explicitly linked.

2010 will be the year in which the major components of CDISC are brought together as one: a consistent and electronically deliverable set of standards. In late 2009, we saw the publication of Release 3.0 of BRIDG and, in January, we saw the Protocol Representation Model Version 1.0 that has been derived from the BRIDG model and delivered as a separate standalone standard.

WINE BY THE BARREL, GLASS OR BOTTLE?

Returning to the analogy of the wine bottle, we have discussed the label, the metadata, and we have discussed the wine, our precious content. What about the bottle? One thing CDISC has been working on over the last few years, and one of the main drivers behind the BRIDG model, is to separate the content from the means by which we move that content. In the same way as you can order wine by the bottle, you can also order it in bulk in a barrel, box or some other container. The content remains the same, irrespective of transport mechanism. Once we have achieved that separation through BRIDG and SHARE, we are far better placed to allow technology changes to take place without fear of that technology change rippling through and affecting the content.

Even with wine, we have seen developments in bottle technology. While the glass bottle remains the same, we have seen a move from the traditional cork, in use for over 300 years, to a plastic cork and even to screw tops on wine bottles. The technology has advanced without affecting the content, some wine aficionados may disagree, but we did see a quote which revealed: "The dirty secret is that most

About the authors



Dave Ibersen-Hurst is VP, Technical Strategy, for CDISC, where he provides technical direction and vision. He has been working within the pharmaceutical industry as both a consultant and trusted advisor for the past 10 years. Dave co-leads the CDISC Technical Advisory Committee (TAC), is a member of the Joint Initiative Council (JIC) a joint ISO, CEN, HL7, CDISC and IHTSDO standards initiative and was a co-lead of the CDISC electronic Source Data Interchange (eSDI) group that worked with the FDA. Email: dibersonhurst@cdisc.org



AJ de Montjoie is Director of Media and Communications at CDISC where she promotes and expounds the benefits of CDISC standards to a global audience through diverse communications channels. AJ has now been working for CDISC for three years and has led the development of new digital communications methods and the redesign of the CDISC website. AJ has previously worked in education as a teacher and manager for 12 years, and prior to that as an engineer for BBC Television. Email: ademontjoie@cdisc.org

winemakers I talk to say they'd love to switch to screwcaps but fear buyers will think them cheap!"

CDISC will continue to develop and enhance XML-based exchange formats for CDISC content, such as CDASH (the existing ODM standard), for protocol where new XML will be developed, for trial design where prototype XML is already in existence and for data tabulations to be exchanged between systems where we are considering the best options for industry on how to remove the constraining shackles of the SAS transport file mechanism.

BE PART OF THE FUTURE OF STANDARDS

One of CDISC's key messages for 2010 is that companies need to look at using CDISC standards at the start of the research process. We are aware that the focus for many years has been on the Submission Data Standards, as they contain data required by regulatory authorities in some countries. However, CDISC now has a complete suite of standards that cover the entire medical research process with the recently released Protocol Representation Model (including study design) and the well-established CDASH for case report forms through to SDTM and ADaM. In addition, the Operational Data Model facilitates the archive and exchange of both metadata and data. As demonstrated above, the BRIDG model is a fundamental part of future standards development as it ensures all CDISC standards are harmonised.

There are increasing demands for CDISC to work with healthcare and to ensure a smooth path for the data from healthcare to research, and this is being addressed by the CDISC Healthcare Link. This has seen results in recent years with the development of the Integrating the Healthcare Enterprise (IHE), and Retrieve Form for Data capture profile (RFD). This work uses the CDISC ODM standard and CDASH for case report forms. To gain valuable insights into the standards (their benefits and areas that need attention), CDISC needs companies to be using them and giving us feedback. Of course, the incredible benefits of standards, including interoperability between departments and partner organisations, cost and time savings, are not to be ignored.

CDISC also collaborates widely with other organisations, including the European Medicines Agency, ISO, Health Level 7, the FDA and may other authorities. CDISC represents its members' views around many global tables and can take an independent view of proceedings in order to ensure that the standards meet the needs of all the major stakeholders.

CDISC has a very busy 2010 ahead: there is increasing demand from the regulators for CDISC to produce more standards and to produce them faster. Clearly it is important that the CDISC community (technical teams, members, user groups and a vast network of volunteers) are part of these future developments.