One of the largest components of clinical trial data is laboratory data. Developing a standard lab data format is critical to achieving CDISC's mission of creating standard data models that support the end-to-end data flow of clinical trials, from the data sources into an operational database and through to analysis and submission.

Standards for lab data interchange exist (e.g., ACDM, ASTM, and HL7), but they are seldom used within the biopharmaceutical industry because they have structures and rules which are not easily applicable to clinical trial laboratory data, limiting their usefulness to central laboratories, CROs, or biopharmaceutical companies.

To further complicate matters, standards tend to be developed on a per-study basis, resulting in a plethora of standards within the industry. Large laboratories are then required to support several hundred standards, increasing costs as well as complicating quality control, data interchange, and data verification issues.

Covance — a worldwide drug development services company — provides a wide variety of central laboratory testing and cardiac safety services to support the clinical trial process. Its central lab service, located in Indianapolis, routinely handles between 1200 and 1400 relatively unique data transmission formats.

Currently, the average cost to develop and validate a new custom format for a specific customer runs about $10,000 per format. This means that between $800K and $1 million are being spent just to develop data transmission formats for new studies.

Seeking a Cost-Effective Model for Submissions

The biggest single issue for Covance's Central Lab Services was to develop a cost-effective model to move data from its central lab to the sponsor, and then to the FDA.

Clearly, Phil Pochon, director of the lab services data management group, had a vested interest in standardizing formats for lab data. Covance joined CDISC back in 2000, with Pochon as a founding member of the Lab data model team. [The team also included representatives from pharmaceutical companies, contract research organizations (CROs), and technology application developers.]

Pochon now serves as chair of the Lab team, and according to Julie Evans, the technical services director for CDISC, "Phil is the single biggest contributor to CDISC's Lab standard."

The CDISC Lab team developed time estimates of doing business with the current lack of standardization. These time estimates ranged from 1.5 to 12 months to set up new studies with new data requirements.

The team then defined exactly what the industry means by 'clinical laboratory data' in order to build a 'superset' of data items that fully describes a clinical trial to the satisfaction of all the stakeholders involved in it. This superset of fields constitutes the content of the model.

Next, the structural definitions of those data items were defined in terms of data type, length, default values, standards of representation, code lists, and whether or not the data items should be optional or required. Wherever appropriate, existing standards were employed.

In November 2002, the team published the first version of the Lab Model (CDISC Lab Model V1.0.0), and Pochon's data management group worked to implement the model as soon as it was published.

Pochon's standards implementation team at Covance includes Senior Data Manager Ben Welch, as well as Data Manager Matthew Kleumper.

According to Welch, Covance was unique in getting in on the ground floor with the early implementation of the CDISC lab standard. "This was a unique opportunity for us, giving us the chance to experiment with the standards format and attempt to map the most common data items first," said Welch.

"... getting in on the ground floor with the early implementation ... was a unique opportunity for us"

"The initial pilots went well in the early implementation phases," said Kleumper. "We didn't have a lot of customers requesting the new data formats, so we were able to test the data maps and translations with a small subset of less complex data translations."

During this early implementation stage, Pochon's team developed their own "home grown" data-extract-transform tool using Java to create a map of the format, link to fields in the database, and then specify translation details.

"As a result of this application development effort, we can now set up a new study for our smaller, less complex translation jobs in a couple of hours," said Kleumper. "Larger, more customized jobs may take 20 hours."

Welch agrees. "The smaller companies have little data legacy investment, so it makes sense and is quicker to send them the 'vanilla' CDISC standard for data submissions."

This is particularly true with the vocabularies. "Our larger customers are still getting a customized version of their data formats because they have more complex translation tables with their own field specifications," said Pochon. "The vocabulary issues are on the submissions end and will eventually be pushed back to the lab. The FDA will need to institute the code format, and that will force the sponsors to start implementing a standard (CDISC) vocabulary."

"And we expect better vocabulary compliance once the vocabulary for the CDISC SDTM is locked down," added Pochon.

To ease the impact of the Lab standard implementation for their customers, the Covance team offered different flavors of implementation for different needs. They defined levels of conformance to help determine the appropriate implementation methodology. Level 1 is defined as structural conformance, and means that all of the fields are in the proper place. A level 2 ranking means that some vocabulary conformance exists on the very basic fields. And level 3 means that the customer is using fields that reference large international standards dictionaries (e.g., LOINC).

"Our first implementation was 2 years ago," said Welch. "We're now sending data in CDISC, on a production basis, to 20+ clients, and have piloted the standard with 10 other customers who are moving to production soon."

Despite their success, the team recognizes that more needs to be done.

"The standard could use an additional text description, catch-all field," said Kleumper. "Some of the data we have just doesn't fit. For instance, textual range fields which end up capturing more text than numerical values. We find that most of our customers use at least one catch-all text field and it's difficult to map random free text to a specific standardized field."

Learning from Experience and Sharing the Lessons

Still, this team is excited by the progress being made.

"Based on our experience of having a [home grown] tool to automate the mapping and translation process, I'd strongly suggest looking at off-the-shelf solutions, as opposed to using hours of programmer time to map and translate data," said Pochon.

"There are vendors in the marketplace selling software that will give you the full map of CDISC formats. The full CDISC structure is already built as a map so experiment with some of the extract format tools," continued Pochon. "We were successful because we developed a tool to automate the process. As an architect, I would definitely recommend a review of the latest mapping software."

The team also suggests that companies consider what it is they need in terms of data translation, and focus on those fields, rather than digest the whole Lab standard at one time.

"The Lab format is very large [92 fields], for very specific reasons. But companies usually read only 25-45 fields. They don't need to load the entire format into their main data management database" continued Pochon. "Don't take on more than you need. The format includes a lot of optional data so focus only on the data you need and that will help to speed up and simplify your implementation process."

"Standardization saves both time and money, and is essential to ensuring the integrity and value of the data being collected during clinical trial research."

Another implementation tip is to budget for a pilot to certify results. Choose a study, and pilot the Lab format by sending data both with CDISC and via the normal channels. Compare the data translation from both methods and determine if the customer is able to deal with the CDISC standard.

From this point on, the Covance team is moving all new studies to use the CDISC Lab format.

"The financial and personnel resources saved as a result of moving to this standard and automating the translation process have been tremendous," concludes Pochon. "Standardization saves both time and money, and is essential to ensuring the integrity and value of the data being collected during clinical trial research."

Case study based on material kindly supplied by Phil Pochon
Director, Data Management, Covance
Ben Welch, Senior Data Manager, Covance
Matthew Kleumper, Data Manager, Covance
June 2006.