PharmaSUG 2016 - Paper DS06

"It is a standard, so it is simple, right?": Misconceptions and Organizational Challenges of Implementing CDISC at a CRO

Susan H.M. Boquist, PAREXEL, Billerica, MA
Adam J. Sicard, PAREXEL, Durham, NC

ABSTRACT

Implementing CDISC standards into the clinical trial process presents several challenges ranging from understanding the complexity of the task to whom should be doing which part of it but can yield many positive effects. There are many possible approaches, all with advantages and disadvantages.

After experimenting with several different options, we have adopted a two team model of database build and "end to end" SAS. While the approach is relatively new, by beginning to include CDISC standards into both teams, we are starting to see anecdotal evidence of positive effects in terms of reduced timelines, increased quality of outputs and employee engagement.

This presentation will review the path we have followed in the quest to achieve full CDISC implementation and additional information on possible misconceptions, suggestions for streamlining processes, employee recruitment and training, resourcing models and sponsor participation in hopes that others may benefit from our mistakes and successes.

INTRODUCTION

The CDISC mission statement explains that:

"CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. **The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.** CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website."

The short form is that CDISC provides a standardized way to submit data to the FDA. While the standards encompass multiple facets of clinical trials, this paper will focus on methods of implementing the Study Data Tabulation Model (SDTM) and the Analysis Dataset Model (ADaM) into the work stream of tabulating and analyzing clinical trial case report form data from the perspective of a Contract Research Organization (CRO).

UNDERSTANDING THE TASK OF IMPLEMENTING CDISC STANDARDS

Most companies are aware that CDISC standards should be followed to submit data to the FDA but the level of understanding varies dramatically from knowing very little to having very definite ideas about how the standards should be used. Some companies have their own versions of the standards as well. Along with that comes a wide range of misconceptions and misunderstandings. Part of the job of a CRO is to help correct those issues and guide the study team in the right direction.

"IT IS A STANDARD SO IT IS SIMPLE, RIGHT?"

We hear this all the time. Those that are unfamiliar with SDTM hear the word "Standard" and assume it must be straightforward to implement. Just read and apply and you are done. Though it is a standard there are multiple grey areas that are open to interpretation based on study design and analysis intentions. A CRO has the additional challenge of working with multiple organizations that may have their own unique interpretations of the standard. Meanwhile the official standards continue to evolve, which adds to the complexity of the data mapping task. The team can be unsure about which standard to follow and whether it is okay to adopt provisional versions. It needs to be a team effort to make any decisions on how to adopt the standard, with the study sponsor having the ultimate authority.

WHEN CAN YOU START THINKING ABOUT IT?

It is rare now but very late in the process, some study teams have suddenly mentioned that they are not sure but they think they might need "SDTM", followed by questions about what is it exactly. Not an ideal scenario but better than not thinking about it at all. Building the clinical database with SDTM in mind makes the mapping process much easier. The Statistical Analysis Plan (SAP) can also be formatted in a way to make creating SDTM and ADaM easier. It really is never too early in the study to start thinking about SDTM and ADaM.

WHO OWNS IT?

There tends to be a lot of confusion about who owns the SDTM data sets. Clinical database developers have a lot of control over the structure of the raw data that gets read into SDTM and in some cases they could create the SDTM data sets. Data management cleans the raw data but they are also involved in the database design. Statistics uses the SDTM data sets to create ADaM and have clearly defined needs from the SAP as to what needs to be included. The ADaM data sets unquestionably are owned by statistics but development of the specifications could be done by either a biostatistician or a programmer. When there are issues with the analysis outputs that are data driven, is the problem in ADaM, SDTM or the raw data? Is it a programming issue or a data value issue? Often times any study issue is initially perceived as an SDTM problem, until the actual root cause is discovered. Without clear ownership, there can be delays in solving issues.

MYTH #1: THERE IS AN SDTM BUTTON

In spite of some statements to the contrary, there is no magic tool out there with a button that can be pressed to map raw data into SDTM. It is an evolving standard and therefore changing at regular intervals which lately have been getting shorter and shorter. This makes developing and maintaining a tool difficult. There are tools that will perform simple mappings for basic domains but to do a complete mapping base SAS programming is required. A tool cannot take into account all of the nuances and needs of every study, which vary depending on the focus and intent of the analysis. There is no substitution for human experience and intuition in this complex process.

MYTH #2: THERE IS AN ADAM BUTTON

Similarly, there is no magic tool out there with a button that can be pressed to map SDTM data into ADaM. It is a flexible standard, so it cannot be covered by a single tool. There are tools that can help with initial mapping specifications but not beyond that. ADaM often contains multiple complex derivations that have to be programmed in base SAS. There is also no getting around the fact that a human interface is required to interpret the SAP and translate it into specifications for creating ADaM data sets from the SDTM data sets.

POTENTIAL ORGANIZATIONAL CONSIDERATIONS FOR CDISC IMPLEMENTATION

To create SDTM and ADaM, the programmers need to be organized into teams. Over the years, we have tried a few different approaches:

SEPARATE DATA MANAGEMENT, SDTM MAPPING AND STATISTICAL PROGRAMMING TEAMS

At first glance, this looks like a great solution, since SDTM mapping requires very specific skills and there can be advantages to having a team that focuses only on SDTM mapping. But that also means that they are disconnected from the database build and analysis teams. They may not have any input into how the database is built and they have no insight into where the data is going for analysis. With teams in siloes it also makes resourcing and employee development a challenge. If the analysis team is overbooked and the SDTM team is not, they can't help each other out. Employees either do SDTM or they don't, which may not be of interest to everyone.

SEPARATE CLINICAL PROGRAMMING AND STATISTICAL PROGRAMMING TEAMS

Database build, data review listings, vendor data integration, and SDTM mapping fall under clinical programming. This helps with the SDTM mappers being more involved in the database build but then they are still disconnected from the analysis team. With some EDC systems, two teams would still be needed to do the database build and do the SAS programming part. Putting just ADaM and TLFs under the statistical programming team still keeps that team disconnected from data management and there could also be issues with communication and timing for deliverables.

ALL ONE TEAM, EVERYONE DOES EVERYTHING

All programmers work on their studies from database build, straight through to analysis and submission. This method gives programmers ultimate control over their projects and they can guide the study from the very beginning, keeping the end point analyses in mind. While it allows for more efficient data cleaning and mapping, it is difficult for programmers to be proficient in all of the steps of the process, as well as to develop in multiple EDC platforms. There is also the issue of overlapping work streams during deliverables, making resourcing and project management a nightmare.

OUR CURRENT TEAM STRUCTURE

After trying the above approaches and getting a feel for the positive and negative aspects of each, we are using the following scenario, which is working out well:

CLINICAL PROGRAMMING BUILDS DATABASES

The clinical programming team builds and maintains the databases, writes on-line edit checks and performs the data extracts. There are specialists in each EDC platform, with some able to work on multiple platforms. They receive training so that they can implement CDASH and SDTM-like formats during the database build to help downstream mapping activities. The team is also responsible for scheduling extracts and working with vendors as needed to maintain consistent working relationships.

STATISTICAL PROGRAMMING IS "END TO END" SAS

The statistical programming team is responsible for everything having to do with SAS. They receive the EDC extracts and any other data generated for the study, such as IVRS, lab, PK, ECG, diary, etc. and read it in to SAS to create everything from data review listings through to TLFs. They are also responsible for reviewing the SAP and mock shells, which are created by the biostatistician. They can then keep the analysis end points in mind while they are programming the data review listings and creating the specifications for SDTM and ADaM, which can also make ADaM easier to program. As a result, the leads have complete overview of the project and take more ownership of it. With oversight of all SAS activities, the lead has an easier time pinpointing whether the source of project issues are related to programming or actual data values. They can catch issues of missing data points much earlier in the process as well. Resourcing is much easier since programmers have a wider range of skills to allow them to fill in any gaps. And in the end, the whole process is more efficient.

HOW WE MAKE IT WORK

In statistical programming, we recruit a mixture of talents within the team and emphasize the opportunities to expand their skills and knowledge, which is very attractive to desirable candidates. We pair up "ex-stats" with "ex-DM" peers so they can cross train each other and have found that being faced with something new, they each suggest ways to make tasks more efficient and easier. We also pair up experienced "end to end" programmers with junior programmers to receive well-rounded mentorship. Every programmer is provided in-house CDISC training to get a consistent base knowledge and we have a CDISC steering committee with multiple SME's to provide oversight and support across departments.

Clinical Data Analysts get training in SDTM so they can keep it in mind during the database design phase. Since they also contribute to the reviewer's guide for submission, they get some training in define.xml. We have anecdotal evidence that having CDISC knowledge helps them design better databases, more efficiently. Using controlled terminology for code lists helps mapping programming, data cleaning and allows for reuse of forms in database builds for additional efficiencies.

Biostatisticians also get training in SDTM and ADaM. They review both sets of specifications during development, as well as the output data sets. By getting them involved earlier in the process, we get more robust SDTM and ADaM for the analysis.

And lastly, we try to promote using official CDISC versions with our study teams, which allows for multiple efficiencies in both training and development. The first time a programmer works with SDTM and ADaM, they are influenced by the version that they are working with. If it is a non-standard version, they have to be retrained to follow the official version later on. It seems to be easier to start with the official version first and then try to work with a different version, if necessary.

NEGATIVE EFFECTS

Some programmers simply don't want to do the parts that they don't know or are simply resistant to working on other aspects of projects. Some new leads may feel overwhelmed when they realize all that they are responsible for. This can make team dynamics, personnel management and resourcing challenging. Some new hires coming in will need extra ramp up time to learn about the part(s) that they don't know. For some reason, confusion about who does the data review listings is an issue with some of the other departments and organizations, since it is still thought of as a data management "thing". If a particular project does not have the statistical analysis component in the contract but includes data transfers and listings, there can be some opinions that statistical programmers do not need to be assigned to the study. Line managers also need to be especially aware of individual skill sets when resourcing and pairing up the right combination of people to cover the work.

POSITIVE EFFECTS

Overall, employees are impressed with receiving full training and have given a lot of positive feedback. Even very experienced SDTM mappers who questioned whether they needed training, afterwards often mention that they learned things that they had missed or never knew. The "ex-stats" programmers have a new respect for "ex-DM" programming tasks, which they had previously thought of as being pretty straight-forward. The best comment was from one "ex-stats" programmer proclaiming after spending several weeks on data review listings, "Programming from raw data sets is hard!" Sponsors and project leaders like having one SAS programming lead to go to for everything. It makes it much easier for them to figure out whom to talk to. Ownership of the SDTM data sets is clear. With programmers having a better understanding of where the data comes from and where it is going, the entire process goes quicker and easier. This leads to increased quality of outputs and reduction in issues of collecting the wrong data or data missing entirely from the data collection instrument. With controlled terminology being introduced during the database build the programming is easier too. And finally, employee retention is improved with fewer issues in the work stream and being able to learn multiple skills. Not being stuck in a silo also results in fewer issues of boredom and burn-out.

CONCLUSION

With the FDA mandate that all studies starting after December 17, 2016 have to be submitted in CDISC format and strong encouragement for studies started before that, CDISC implementation for new studies and conversion for older studies is going to become increasingly more important. By adopting standards early on, putting some thought into organizing the teams involved and providing training and support, the task of collecting and analyzing data can become more streamlined with higher quality outputs. This can lead to happier teams and more satisfied customers.

Don't forget to:

"CDISC it!"

ACKNOWLEDGMENTS

We would like to acknowledge our awesome programming teams and sponsors, without whom we would not be here.

RECOMMENDED READING

- CDISC Data Standards Submission Team. "Study Data Tabulation Model Version 1.4" CDISC web site. November 26, 2013. Available at: www.cdisc.org
- CDISC Data Standards Submission Team. "Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.2." CDISC web site. November 26, 2013. Available at: www.cdisc.org
- CDISC Analysis Data Model Team. "Analysis Data Model (ADaM) Implementation Guide V1.0." CDISC web site. December 17, 2009. Available at: www.cdisc.org

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Susan H.M. Boquist
Enterprise: PAREXEL
Address: 1 Federal Street
City, State ZIP: Billerica, MA 01821
Work Phone: (978) 313-1673
E-mail: susan.boquist@parexel.com
Web: www.parexel.com

Name: Adam J. Sicard
Enterprise: PAREXEL
Address: 2520 Meridian Parkway, Suite 200
City, State ZIP: Durham, NC 27713
Work Phone: (919) 294-5463
E-mail: adam.sicard@parexel.com
Web: www.parexel.com