

Lessons Learned from SEND and Data Standardization Implementations at the FDA and Sponsors

Shree Nath and Kristi Johnson

PointCross Life Sciences, Inc., 1291 E. Hillsdale Blvd., Suite 304, Foster City, CA 94404, USA

Contact: shree@pointcross.com

1 – Introduction

The FDA’s February 2014 draft guidance for standardized electronic data that must accompany future regulatory IND, NDA and BLA drug applications calls attention to an accompanying Study Data Technical Conformance Guide which provides a useful governing technical framework for sponsors and CROs to prepare submission ready datasets.

This Guide includes technical specifications for planning and delivering standardized study data (e.g., study data standardization plan, study data reviewer’s guide), exchange file formats (e.g., XPT, XML, PDF files), nonclinical data submission formats (the CDISC® SEND data model for nonclinical data), controlled terminology references, and data fitness considerations designed to support effective, efficient and consistent regulatory reviews (e.g., data validation checks and study data traceability).

In this presentation, we present case studies and lessons learned from ongoing implementations of informatics tools, workflows and supporting business processes for the management of standardized nonclinical electronic study data at the FDA and at four global sponsor companies.

We also consider the technical and business process impacts of the FDA’s recently issued draft guidance on these implementations.

2 – Implementation of the NIMS Program at the FDA

The FDA has implemented informatics tools and related processes for the receipt, loading, storage, visualization, analysis and review of standardized nonclinical study data that accompany regulatory drug submission applications by sponsors.

This effort commenced in 2011 with the setup of the informatics infrastructure for general toxicology and carcinogenicity studies; data receipt, transfer and loading workflows; and on preliminary training for reviewers to access and use the Agency’s Nonclinical Information Management System (NIMS).

Since 2013, the following four areas are also being addressed:

- (a) Implementation of new data validation rules for standards conformance and quality checks to support meaningful reviewer analysis. These rules will be used to check sponsor submitted datasets. They are published on the Agency’s standards website and are expected to be evolve over time (<http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>)
- (b) Incorporation of NIMS into the regulatory review process and identification of new reviewer requirements for reporting, analysis, and visualization
- (c) Standardization of submitted non-standardized general toxicology and carcinogenicity study data (in PDF and non-standardized electronic formats) for use in NIMS, and
- (d) Standardization of data to enable extensions of the NIMS data model for safety pharmacology and reproductive toxicology studies that are on the SEND roadmap.

3 – Standards Implementations at Sponsor Companies

The sponsors with whom we have been implementing SEND solutions range from one mid-size sponsor to three large ones. In general, all of the implementations will generate SEND ready datasets along with system generated define files – common to these deployments are the load, transform, harmonize, store, view and submit functions. The solutions are at various points in their life cycle from initiation to go-live in terms of being able to generate submission ready datasets. All four sponsors have a combination of internal data collection systems (2-4 each) as well as CRO generated datasets. Initial analyses conducted at the outset of our client engagements indicate that SEND readiness gaps fall into one or more of the following areas: collection, e-data, extraction, terminology and formatting. Across all sponsors, we have observed challenges with the quality, completeness and state of readiness for study data to be represented in SEND.

For example, with a phase 1 implementation at one sponsor, over 20% of the data needed to be manually populated, 29% can be directly used and 36% remains to be addressed to achieve complete readiness. For another sponsor, 59% of the data can be directly used and 27% can be used with transformation; in part recent modernization of their source systems has accelerated goals towards achieving SEND readiness. For a third sponsor, assessment of one of their CROs indicated that significant data transformation (approximately 40%) would be required implying specification of a large set of business rules and therefore increased complexity. Such analyses can help define focus areas for sponsor-CRO data exchange sooner in the deployment process.

Note that these assessments do not account for the quality of the actual data, and to the extent to which these are properly populated in the originating data sources. It has become evident that triaging of representative studies early on in the process can assist in uncovering these gaps to draw attention to challenges in these areas.

Additionally, gap analyses evaluation may call for changes in business processes, source systems and extraction interfaces, and terminology management; in some cases, these work streams are undertaken in parallel with a SEND implementation. In other cases, sponsors make a conscious choice to manually populate some parts of a SEND data package (e.g., protocols) to achieve an initially compliant state and use the learnings to direct future process and system changes in an iterative manner. Other sponsors attempt to make these changes as they proceed towards implementation increasing the scope, complexity and dependencies within the context of a much larger program.

Since SEND permits significant degrees of freedom in how study datasets are modeled, there can be variability in implementations across sponsors and their CROs. The approach of sponsors towards terminology management also varies ranging from strict compliance with SEND to a broader strategy of standardizing vocabularies far beyond SEND.

One of four sponsors is building the full lifecycle of study data flows from definition through submission into their overall operations as part of a strategic SEND driven business process change. Another sponsor is integrating SEND data flow with their existing solution for R&D visualization and analytics built on the same technology foundation, and intends to use a hosted solution for their SEND solution. All four sponsors will validate the solutions in accordance with their respective QA processes and controls.

Involvement of business, IT/informatics and quality personnel varies across sponsors from very extensive engagements with multiple business process, source system upgrade and systems integration workstreams, to a much more targeted involvement driven by the phase of implementation with varying degrees of inter-dependencies and decision making complexity.

4 – Observations and Conclusions

Based on our experience with the FDA and with sponsors, it is evident that there is a fundamental difference in the emphasis of their respective SEND based implementations. The Agency is interested in accelerating the flow of standardized e-data to bring efficiency to the review process and to advance regulatory sciences. The Agency’s repository of data in NIMS include both sponsor submitted SEND datasets as well as data standardization of previously submitted studies. Exploitation of data using viewing and analytics tools that work within and across studies is critical to these goals, as is the use of such data for reviewer reporting. Likewise, as the final destination for sponsor submitted datasets, the Agency is preparing for the future by investing in tools and processes that not only address immediate targets to handle general toxicology and carcinogenicity studies, but other study types in the near to longer term. Learnings from their data standardization initiatives are being used to improve data validation, and viewing capabilities of NIMS. Finally, the Agency is also piloting alternative submission formats (Excel, XML) besides SAS transport files which can benefit sponsors and CROs alike.

Sponsors are currently focused on processes and tools to receive, transform, store and create submission ready SEND datasets. Their decisions and implementations are presently driven by the nature and variability of their data sources, understanding of SEND, determining how to best prepare and generate SEND datasets that will successfully load into NIMS, and putting in place quality assurance and data governance controls. These are key foundation steps in any implementation. As the industry’s experience evolves, and with the February 2014 publication of the Agency’s draft guidance for e-data submissions, it is clear that sponsor implementations will extend into study planning and preparation stages, and expand the needs of their processes and tools to support submission requirements. It is only then that they will likely step into the realm of how they themselves can routinely consume and use standardized data in research and development activities. One of our sponsors started with such a goal back in 2009, and their SEND implementation has benefited significantly from experience with extracting and standardizing data and terminologies, and using this resource for R&D. Collectively, the technology, tools, and templates that we have been building over the past few years are proving to be useful in helping both sponsors and the Agency to meet their respective goals.

5 – Future Directions

The challenges faced by sponsors during their implementation have led to a number of new learnings and highlight key areas that other organizations embarking on their own SEND journey could benefit from in order to make choices that reflect their own specific business situation. Organizations like PhUSE (e.g., the Interorganizational SEND team within the Nonclinical Working Group) and CDISC provide vehicles for exchanging practical lessons learned and to address emerging areas including planning and preparing for a study, determining how to best check and manage data packages for submission conformance, and identifying upcoming challenges particularly with regards to handling complex study types including reproductive toxicology.

As sponsors and regulators move closer to the goal of streamlining and industrializing data flow and access, new opportunities to better harvest standardized data and terminologies will likely open up. For sponsors, these may range from reducing the largely manual processes involved in preparing PDF study reports to potentially using a common data resource for data analysis, reporting and regulatory e-data submissions. It is evident that the use of standardized e-data has the potential to significantly reduce interaction costs and inefficiencies between CROs, sponsors and regulators. Likewise, new opportunities to leverage the data resource for R&D and other purposes will open up in the future, and will create demand for new tools and ways to exploit them and support broader goals such as translational analytics.

6 – Acknowledgements

We wish to acknowledge the FDA, the sponsors we work with and their staff for contributing to the insights and knowledge that we have gained over the past several years.

We also want to acknowledge the technical, programmatic and data standardization staff at PointCross Life Sciences who have been involved over the past many years in delivering our products and services to the industry.