

# Implementing a Standardized Data Model – A Real Life Experience

Curtis Wolf & Paul Gilbert, DataCeutics, Inc, Pottstown, PA

Patrick Genyn, Janssen Pharmaceuticals, Titusville, NJ

## ABSTRACT

As organizations strive for gains in efficiencies in processing and using clinical information, the need for a common set of terminology and data standards becomes vitally important. As partnering and use of outside organizations become more common, this need takes on an even broader challenge. One such company was the setting for a project we will be discussing in this presentation.

The intent of the project was to facilitate the sharing of data across multiple platforms and systems thereby enhancing and creating efficiencies in the data management, review and reporting process and enabling sharing of resources and unique developments across the organization. This model is being used as the basis for the designing care report forms (CRFs), clinical databases and analysis files and consequently for Case Report Tabulations (CRTs) for submission to the FDA and other regulatory agencies, as well as the data source for a variety of other reporting tools.

The actual data format generated in the Standard Data Model format was not confined to one type but could be represented in SAS Data files, ASCII text files, Oracle tables, or XML as a future implementation. The Standard Data Model uses components of the CDISC submission data model.

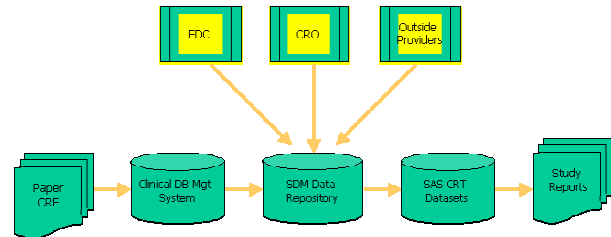
## INTRODUCTION

To gain an understanding of the scope and complexity of this project, it would be beneficial to understand the reasons this daunting task was undertaken. The setting for this project was an organization which had decentralized IT systems and a variety of clinical database models. Management saw the opportunity to standardize the clinical data models, thus enabling the sharing of resources and unique developments across the combined organizations and with external partners. At the same time, the opportunity presented itself to move toward using an emerging industry data standard model known the Submission Data Model (SDM) (1) that was being developed by the Clinical Data Interchange Standards Consortium (CDISC).

The resulting Standard Data Model is not a one-size fits all solution. The Standard Data Model is meant to be flexible to allow for system-specific implementations that would enable more efficient processing within a certain segment of the clinical data process stream. However, by maintaining as consistent as possible variable names, variable labels and coding standards, sharing of knowledge and data will be enhanced across the entire organization.

This paper discusses the conduct of the Standard Data Model standardized model development project. Its focus is the development process that was employed, the standards used to guide the development and the deliverables of the project. Below is a graphic representation of the clinical data flow and how Standard Data Model fits into that flow.

## Standardized Data Model



## COMPONENTS OF THE PROCESS

For any project to be successful, having an organized approach is essential. Our project plan addressed many components they included:

- Assemble a cross-functional team
- Educate the team on standards
- Establish a meeting schedule
- Scope the project and design standards
- Model the standard structures
- Review and approval
- Obtain feedback from user community
- Finalize the Standard Data Model
- Conduct the proof-of-concept studies
- Formulate the implementation plan
- Develop the training material
- Develop a Standard Data Model maintenance plan

These steps are discussed in detail below.

**ASSEMBLE A CROSS-FUNCTIONAL TEAM** – This was a key component of the whole process. Selecting a knowledgeable, communicative group of individuals representing the key areas across the entire global organization involved in the clinical data process. Members were encouraged to take leadership roles within their groups and act as liaisons between the Standard Data Model committee and their respective clinical data departments.

At the same time, other initiatives were underway to harmonize the two organizations in areas such as CRF design, analysis data standards, and data collection systems and methods. With this parallel approach, it was challenging to keep in step with each other. The Standard Data Model committee was given the lead role for the most part, with other initiative committees instructed to follow the Standard Data Model lead.

**EDUCATE THE TEAM ON STANDARDS** – Several standards were used as the basis for the development. It was essential that each team member gain an understanding of the standards. This was accomplished by delivering formatted presentations on each of the standards. The standards used include the CDISC

SDM (1,2), FDA electronic submission documents (3,4,5) and internal data standards (6,7,8).

**ESTABLISH A MEETING SCHEDULE** – The task handed to the Standard Data Model team was quite time consuming and required substantial commitments from each member. A regular weekly meeting schedule was established and adhered to as closely as possible in order to keep the process moving forward and meet management's aggressive timeline. It was important to be flexible allowing for conference calls, lunch meetings, and variable meeting locations. This flexibility helped keep the project going when the staff was faced with competing priorities and tough timelines.

**SCOPE THE PROJECT AND DESIGN STANDARDS** – The team needed to agree on the scope of the project, which included several points.

- Identifying the deliverables. The team agreed to include documentation for an overview, data model, data coding, presentations and training materials.
- Establishing both short and long term timelines.
- Selecting the standards to guide the model. The team agreed to use the CDISC SDM and augment it with current internal standards.
- Selecting the method and scope for developing the model. The team selected a top-down approach where we determined domains then agreed to identify domain naming conventions, keys, variable naming conventions, and finally modeling the individual domains.
- Selecting a process for draft, review and approval. The team selected a multi-step process where: a consultant created the draft model, a small "Core" team met to review the model, the full team reviewed the model, the first draft was finalized and then sent to the users for review, the Core team reviewed all users comments and made changes appropriately, and finally a version was created.

The next steps discuss the modeling process.

**MODEL THE STANDARD DOMAINS** – This first step in modeling was to identify the keys, naming conventions, and data conventions that were to be used when developing the individual domain models. The second step of the modeling was to design individual domain definitions that would meet the clinical information needs of the entire organization and also to move closer to the CDISC Standard Metadata Model and FDA guidelines. The modeling process required many compromises and lengthy discussions to come to mutual agreement. It was impossible to discuss domain definitions without discussing and gaining an understanding of each organizations data practices and methodologies as well. Additionally, with international team members involved, there were complications due to unique clinical definitions and cultural peculiarities.

**REVIEW AND APPROVAL** – All steps of the process went through a team review and approval process. One of the rules we established early on, was that revisiting of agreed upon changes and/or standards, would be highly discouraged. Also, downstream work brought up issues, which require changes to the high-level standards. Meeting minutes were helpful and the distribution of updated documents was maintained by utilizing a document repository system. Each committee member was responsible for reviewing the draft domains prior to final release of the initial version of the Standard Data Model.

**OBTAIN FEEDBACK FROM USER COMMUNITY** – Once the initial draft version of the Standard Data Model was created, each Core team member met with their respective departments and obtained feedback. All of the feedback was then consolidated into

one document arranged by domain and was reviewed by the Core team. Responses were given for each issue and a decision as to whether a change to the Standard Data Model was warranted. The near-final draft was sent out the full user community for review. This was preceded by formal presentations to the user community, given by the team members.

**FINALIZE THE STANDARD DATA MODEL** – All user feedback, team issues, and other input from various sources was evaluated and a final version for initial release was created. The Core team reviewed all input and discussed each point or issue raised by the users.

**CONDUCT THE PROOF-OF-CONCEPT STUDIES** – Several representative studies were selected and used to perform proof-of-concept (poc) data conversions into the Standard Data Model model. The poc studies included two phase 3 randomized double blind studies. Issues were documented throughout the conduct of the poc studies and reviewed by the hammer team. The Standard Data Model was subsequently modified if warranted.

**FORMULATE THE IMPLEMENTATION PLAN** – This was key to the success of the project. The best data model in the world would fail without the proper rollout and strategic implementation plan. All affected systems and procedures were identified and input from key personnel was obtained in order to formulate a comprehensive implementation plan. This plan underwent several modifications as we more accurately assessed the impact of converting existing studies versus beginning new studies with the new Standard Data Model.

**DEVELOP THE TRAINING MATERIALS** – In order to train the various user groups who would be implementing and working with the Standard Data Model, materials were developed to address the various aspects of the Standard Data Model. A Training/User manual was developed which expanded upon the domain definitions and included real life implementation examples. Sample data and corresponding metadata were hyperlinked into the training manual to eliminate the need to flip back and forth between various documents. In addition to the training manual, annotated versions of the new standardized CRF's as well as CRF's from the proof-of-concept studies were included as a reference for users to understand the typical mapping from CRF to the Standard Data Model variables. The CRF's were hyperlinked to the training manual.

Conversion guidelines were also developed to assist programmers with the task of upgrading any studies that are to be converted to the new Standard Data Model standards. The guidelines were developed as an outgrowth of the proof-of-concept studies.

**DEVELOP A STANDARD DATA MODEL MAINTENANCE PLAN** - A plan was developed to facilitate the long-term maintenance of the Standard Data Model whereby suggestions and/or requests for modifications are accumulated by a Standard Data Model administrator. Periodically, those requests will be reviewed by a team and any approved additions will be made and a updated version released. In most cases, changes to the model will consist of additions of new domains and variables. Deletion of variables and changes in variable type are not desirable changes.

Additionally, as new versions of the CDISC/SDM are released, a careful evaluation of the nature and impact of the changes will be conducted. A subsequent decision will be made as to whether to continue with the current version for new trials in the same drug development program or to migrate to the new standard. Maintaining multiple versions will be a management challenge.

## CDISC ISSUES

As previously stated, one of the overall objectives of the project was to move towards the emerging industry standard known as

CDISC. However, there were obviously going to be deviations from CDISC based on specific organizational preferences and practices. You might ask the question, 'Why were deviations from CDISC allowed at all?'. There are two primary reasons for CDISC deviations. First, the Standard Data Model is focused on source data only, therefore derived data is not included. Second, the CDISC standard is based upon SAS as the data storage medium. The Standard Data Model is a platform neutral data structure and so is not restricted to SAS methods or standards. The areas of deviation are as follows:

- Derived data and variables that are copied onto each domain were not stored in the Standard Data Model. This is added in a downstream step when the data is used in reporting.
- Many optional, study-specific variables were defined in the Standard Data Model that could be added on an as-needed basis at the study design stage.
- In some domains, character date fields were defined in the Standard Data Model to accommodate the possibility of missing date components.
- Several core (required) variables defined in CDISC were not collected on the case report forms (crfs) and therefore were not defined in the Standard Data Model.
- Standard data collected on the new crfs but not defined in CDISC were added to the Standard Data Model
- As CDISC evolved, naming standards were not always clearly maintained causing conflicts which we tried to avoid. Hopefully, these inconsistencies within CDISC will be addressed in later versions.

## STANDARD DATA MODEL DOCUMENTATION

As part of the project scope mentioned previously, documentation of the project was defined. The project documents included the following.

**Standard Data Model Overview** – A comprehensive document was created to explain the reasoning behind the Standard Data Model. All sources of information such as CDISC or other FDA guidance documents were identified. Also, detailed descriptions of each domain, naming conventions, keys, and rules for adding additional domains and/or variables to the Standard Data Model were included.

**Standard Data Model Definitions** – the metadata for each domain was defined including variable names, data types, keys, detailed descriptions, and optional variables. The metadata contained the following components in each domain definition: Variable Name, Variable Label, Key Order, Variable Type, Variable Length, Decodes/Formats, Origin, and Comments.

**Standard Data Model Code Lists** – As part of the harmonization process, the code lists were standardized. This had to be coordinated with the design of the Standard Data Model and new crfs.

**Training Materials** – The overview document was combined with graphical process diagrams and sample data from actual studies to create a useful training manual. This was hyper-linked to the full set of data created by the proof-of-concept studies for further perusal by trainees.

**Standard Data Model Update Process** - A set of maintenance procedures for the long-term maintenance of the Standard Data Model were developed and documented and will be followed when changes are required for the model.

## IMPLEMENTATION PLAN

As the Standard Data Model design phase neared completion, discussions were initiated to define a phased implementation

plan. The steps involved in determining the implementation of the Standard Data Model were as follows:

**Assess short and long term impacts on existing systems and procedures.** Throughout the design process, additional issues were identified that would impact delivery and implementation of the Standard Data Model. As part of the overall project, new systems for various pieces of the clinical data flow were being determined. All of these decisions had to be coordinated and assessed in light of the Standard Data Model design. The team members had many concerns initially about the impact on their particular area. Naturally, people were prone to want to protect their own systems and procedures that had been developed over many years. Most of the fears were alleviated though as people were included in the decision making process and compromises were made to accommodate nuances within each group.

**Develop a conversion strategy.** One major issue was, what studies should be converted and what was the most logical and least disruptive method for phasing in the Standard Data Model. The decision was made to convert all new drug development programs as well as development programs moving into the integration stage (ISS/ISE) to the new standards. For all other cases, a clear business benefit will need to be presented as to why the new data model should be applied.

**Keeping the long-range goal in mind.** Over and over the team members needed reassurance that the overall goal was to make life better for them, not more difficult. Of course, this is easy to say for any project. But the benefits of a harmonized organization all speaking the same language and utilizing industry standard nomenclature eventually won over the most ardent opponents.

## CONCLUSIONS

In conclusion, the Standard Data Model was a monumental task that could only have succeeded by the committed input and sacrifices on many individuals. Management also played a key role in keeping the committee on track by occasionally addressing the committee on the importance of the task and the long-range benefits. Some of the obvious benefits of implementing the Standard Data Model were:

- A spirit of cooperation and team work was fostered across the entire organization as a result of building cross-functional and cross-organizational teams to develop the Standard Data Model.
- By ensuring that the Standard Data Model was CDISC compliant, we moved the entire organization towards an industry standard and aligned with other regulatory requirements.
- The Standard Data Model facilitated the sharing of data across multiple platforms, organizations, and with external partners. This reduces the time and cost involved when transferring data.
- Having a common set of terms and data storage standards allowed efficient use of resources by fostering sharing of personnel across the organization.
- The conversion of data between internal systems was streamlined by having near identical data structures and terms.
- The process of dealing with external partners was simplified by supplying a clear and standard data specification structure for data delivery.
- Overall, the drug development cycle will be shortened by the gains in efficiency in the overall clinical IT process.

Other benefits will only be seen over the course of several projects but based on similar projects conducted at other clients, we are confident that it was all worth it.

## REFERENCES

- 1) **CDISC Submission Metadata Model , Version 1.1**, April 30, 2000 and the **Individual Submission Metadata Model** Data files June 16, 2001.
- 2) **CDISC Submission Metadata Model , Version 2.0**, November 26, 2001
- 3) **Providing Regulatory Submissions in Electronic Format – FDA CDER, General Considerations**, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), January 1999.
- 4) **Providing Regulatory Submissions in Electronic Format – NDAs**, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), January 1999.
- 5) **Guidance for Industry Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format**—Biologics Marketing Applications [Biologics License Application (BLA), Product License Application (PLA)/Establishment License Application (ELA) and New Drug Applications (NDA)], U.S. Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research (CBER), November 1999 REVISED
- 6) **Internal documentation on Tables and Rules for Database Structures**
- 7) **Internal documentation on Global Data Standards for Clinical Trials**
- 8) **Internal documentation on Standardized Case Report Forms**

**Curtis Wolf** is a Sr. Project Manager at DataCeutics Inc., a leader in clinical information technology. His seventeen years of experience with clinical information systems include development of clinical database and reporting software, development of standardized clinical data model requirements, regulatory submission-driven programming project management, and applications support of for SAS/PH-Clinical and SAS/IntrNet.

**Paul Gilbert** is President of DataCeutics Inc., a leader in clinical information technology. His twenty years' experience includes clinical data management, implementing and maintaining clinical systems, designing and developing SAS-based biostatistics reporting systems, managing SAS based NDA programming support, supporting electronic regulatory, submissions, SAS/PH-Clinical implementation and support, SAS/IntrNet-based systems, and clinical systems integration.

