Paper DS16

SDTM Automation: Implementation, Rollout and Lessons Learnt

Minda Quiroz, Alcon, Fort Worth, USA
Bruce Chambers, Veramed, Twickenham, UK

ABSTRACT

CDISC standards have traditionally been applied to pharmaceutical trials, but the industry is now recognising their value in Medical Device studies. Alcon aimed to become "standards-ready" for its medical device portfolio while ensuring CDISC compliance for its pharmaceutical programmes.

By implementing a unified standards framework and automation tools, Alcon achieved measurable improvements in data quality, and cross-study consistency and timelines. This approach enabled scalable adoption across therapeutic areas and study types, while also reducing manual effort and rework.

This presentation will detail Alcon's journey from initial standards development to enterprise-wide rollout, highlighting key challenges, lessons learnt and future ambitions. Attendees will gain practical insights into how automation and governance can transform SDTM production across diverse clinical programmes.

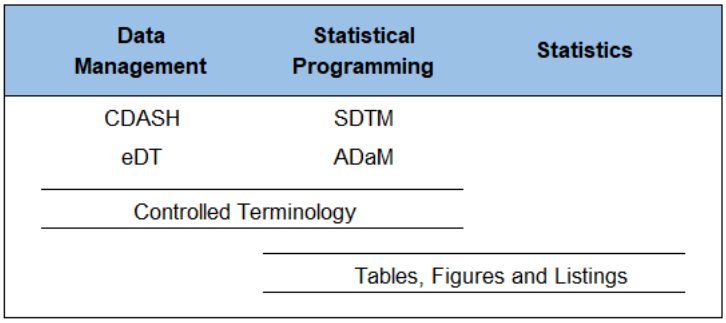
INTRODUCTION

Most companies associated with PHUSE are now at least a decade or more past CDISC day 0 but for many medical device (MD) companies, fully compliant CDISC implementations are new. While a future FDA mandate is anticipated, many MD companies are still waiting for a definite requirement date before taking action because of the significant resources needed to embark on a fully compliant CDISC implementation.

As an original medical device company, Alcon's re-entry into Pharma provided a good opportunity to start implementing CDISC for our MD programmes too. This paper is intended to share our rationale for implementing CDISC for MD and the strategy to develop our Pharma data standards, processes and tools with MD in mind, enabling us to optimally leverage our Pharma implementation to build compliance for MD too. This paper will also discuss how we are transitioning the organization to adopt CDISC standards for MD while maintaining a delicate balance between acquiring organizational expertise and staying efficient with project execution.

VALUE PROPOSITION FOR CDISC IN MD

Alcon has a distributed data standards team. Statistical Programming is responsible for generating SDTM, ADaM and TFLs. Statistical programmers at Alcon work across our entire Pharma and MD programmes.



Performing analysis and reporting for Pharma in CDISC while MD remains legacy means using two different sets of processes and tools which can cause confusion even for seasoned programmers, entail steeper learning curve for new programmers, and necessitate higher maintenance efforts. Thus, from the beginning, there was a desire to streamline operations using one process company-wide that is based on CDISC. The biggest appeal, however, is the prospect of building harmonized clinical data which will facilitate end-to-end automation and get the organization ready for additional use cases such as generation of synthetic data and the use of AI. If done right, our operations for MD will also hugely benefit from the data standards library, existing automation tools and governance that will be established for Pharma. The reuse of these standards artifacts from tabulation to analysis and reporting will bring improved data quality and increased efficiency.

CLINICAL DATA STANDARDIZATION PROJECT

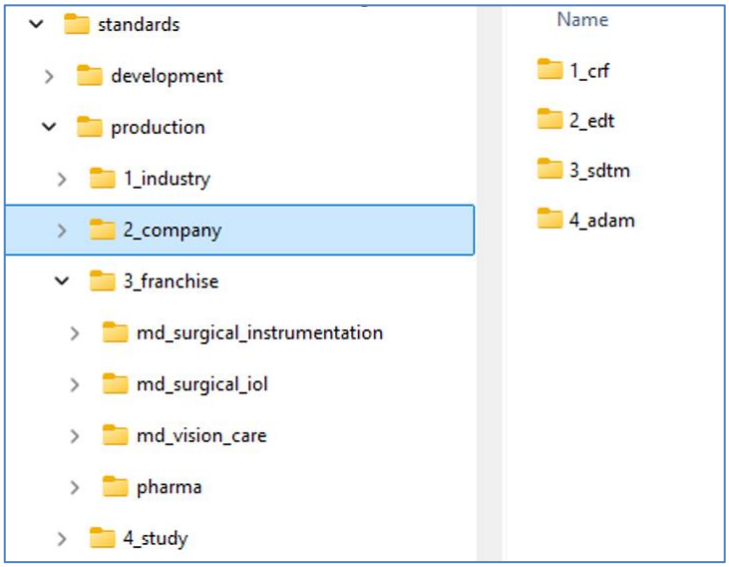
Alcon implemented a Clinical Data Standardization project (CDS) to get the organization ready for Pharma re-entry while considering future CDISC requirement in MD. A project team was formed with representatives from statistical programming, data management, statistics, IT business partners and CDISC consultants.

Goals of the CDS project:

- Build standards library
- Gain internal CDISC expertise
- Develop automation tools and processes including governance of standards
- Get ready for Pharma Submission
- Lay the foundations for CDISC-based MD studies

THE STANDARDS LIBRARY

In consultation with the CDS project team, CDISC consultants built an initial library of SDTM and ADaM specifications with mappings from source datasets and machine-readable instructions for creation of SDTM datasets.



The file structure of the library follows a straightforward pattern of nested folders which are separated out over four broad sections. Each group can exist in 'development' or 'production'; all new entries to the library should be placed first into the 'development' area until they are ready to move to the 'production' area.

- **Industry** folder stores published CDISC Models and Implementation Guides (IG) used by the company along with any associated External Dictionaries and Codelists
- **Company** folder stores artifacts that apply across all Alcon franchises. They should be used for any study that collects company level clinical data.
- **Franchise** folder stores artifacts that apply across a specific franchise only. Some clinical data may also apply to a specific franchise only thus the need for a franchise level artifact.
- **Study** folder stores any artifacts that apply to a single study only such as specialized versions of company and franchise CRFs for a specific study.

Based on study source dataset names, a starting draft define specifications and SDTM pipeline specifications can be produced where a source dataset name has library mappings, with study level specifications able to override library ones if needed.

THE SDTM PIPELINE

The SDTM pipeline is a SAS-based application to transform raw data collected in an electronic data capture system (EDC) as well as external data into SDTM following a metadata-driven process. It is made up of several components which perform discrete operations and are chained together to form a pipeline.

15 mapping functions are provided (more planned) that cover most mapping scenarios. For other scenarios, code can be injected where it is simple (added lines within the same SAS datastep). More complex (multi-step) code injections can be accommodated as bespoke additions – but few needed to date.

As each SDTM conversion will have subtle differences, provision has been made to allow for study-specific activities to occur at certain stages within the pipeline. As the pipeline is metadata-driven, the actual mappings created and executed are specific to each conversion, however, the macros which drive conversion remain static throughout. While code and standards are managed centrally, overrides to both are possible – but with monitoring enabled by

Standards and Tools Managers – to enable ongoing review of where overrides/deviations occur to understand why and feed this back into code and standards adoption review cycles.

Additional functionality can be added to the pipeline as required, ensuring that it is able to change as our requirements do. Notably since the first release, the addition of built-in runtime quality checks has been the most significant evolution.

- Progress report showing any unmapped source items with a % mapped overall score for the study
- Flag records that do not meet SDTM unique key requirements
- Identify non-printing special characters
- Compare datasets vs. specifications (actual vs planned)
- Compare CT present with superset
- Identify CT synonym use and auto-decode to submission values

JUST IN-TIME CDISC TRAININGS

CDISC consultants conducted training on CDISC Fundamentals including NCI Controlled Terminologies (CT), P21 Conformance Checks and Submission Package construction. The development of internal expertise was very intentional, conducted in batches, for specific purposes that aligned with project activities and roll-out of standards.

- Batch 1 – CDS project team who actively participated in the initial use of the standards ("Dry Run 1"), the fine tuning of the standards library, process definition and tools development. After the fundamentals, the timing of succeeding training courses coincided with related activities in Dry Run 1 which provided ample hands-on exercises for the project team.
- Batch 2 – Programmers and statisticians for the first Pharma and MD studies
- Batch 3 – All Biometrics associates (programmers, statisticians, data managers) in the US and Japan
- Batch 4 – All Biometrics associates in other countries

By the time Batch 4 was conducted, the 2-year CDS project had been successfully completed. The members of the project team who have gained the necessary expertise in different CDISC topics successfully conducted the Batch 4 training courses with minimal support from a CDISC consultant.

FIRST STUDIES

CDISC consultants mentored and partnered with project team for first studies. After the training of the project team and the initial library of SDTM and ADaM mapping specifications were available, Dry Run 1 was conducted using source data from a completed MD study to determine the completeness and accuracy of the mappings from source to SDTM to ADaM, and to develop processes for the reuse of these standards. Aside from SDTM and ADaM datasets, selected TFLs were also generated and compared with the outputs from the completed study.

Dry Run 1 was a long and tedious exercise which surfaced the bottlenecks and many opportunities to improve our initial implementation. The use of an MD study, where the team has higher expertise, allowed us to navigate the project better, make more objective decisions and resolve issues faster.

Dry Run 1 Issues	Enhancements
Manual harvesting of specifications from the library	Developed programs that automatically harvest specifications from the library to create combined study specifications for all applicable domains. These specifications include mappings from source datasets and machine-readable instructions for creation of SDTM datasets.
Separate specifications for P21 define.xml causing redundant manual work	Enriched the mapping specifications to also include all information needed for P21 define.xml Developed programs that automatically create near complete P21 define input files from the enriched specifications – needing few manual/final updates
The SDTM and ADaM mapping specifications had very different structures and attributes requiring different steps for similar downstream activities	Harmonized the SDTM and ADaM mapping specifications as much as possible
Change of mindset from study-specific to consuming standards	Evolved processes and reports to enable optional items to be excluded where not used and enable additions

After Dry Run 2, which used another completed MD study, Dry Run 3 was performed using data from a completed outsourced Pharma study. This required an iterative process of evaluating company versus franchise level standards and harmonizing company artifacts to cover both Pharma and MD. There are cases when Pharma requires a

specialized version of a company CRF for the collection of extra data specific only to Pharma. The same is true for MD. These CRFs and related artifacts were moved to franchise level to avoid the effort to maintain extra artifacts at company level without a lot of reuse.

Sample guidelines to determine which domains become part of company standards:

- Company level if CRF is the same across all franchises (e.g., Medical History, Concomitant Medication)
- Company level if CRF is the same except for a couple of extra variables in MD or Pharma which can be easily dropped using the SDTM pipeline
- Franchise level if there is significant variation (e.g., AE001 for MD, AE002 for PH)

After successful dry runs, Alcon quickly took on library maintenance and capability build with less external support, becoming mostly self-sufficient quickly. The first 5 studies were the biggest challenge, from that point metadata and processes are well dovetailed but more optimization still ahead.

KEYS TO SUCCESS

Many factors contributed to the successful CDISC implementation, but the biggest lever is the combination of the consultants' CDISC expertise and Alcon's strong process orientation.

ONE-SIZE-FITS-ALL STUDIES

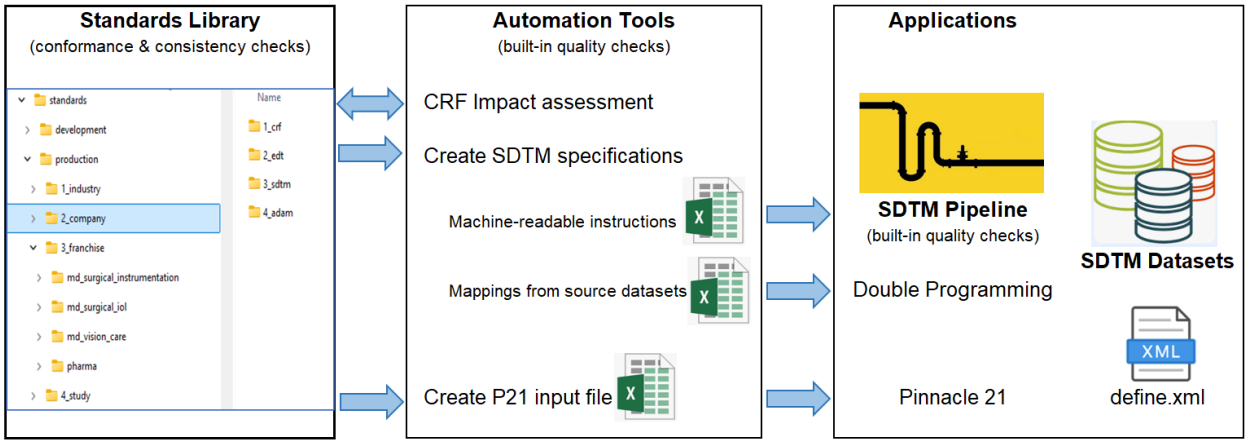
The CDS project team took intentional steps to design processes and tools that work across most studies.

- Gained understanding of the steps and developed expertise needed to create SDTM package
- Optimized the sequence of activities and streamlined tasks
- Automated some steps to get consistent results and improve efficiency
- Made the process repeatable by establishing clear guidelines, documenting step-by-step instructions in Users Guide and promoting the use of checklists. When programming teams feel supported, they become more open to new ways of doing things.
- Standards mindset is enterprise level. Office hour sessions, CDISC training sessions and many opportunities for small group and/or one-on-one mentoring have played a big part in CDISC enablement.

AUTOMATION TOOLS

The standards library and SDTM pipeline are supplemented by tools that automate the repetitive steps in the creation of SDTM datasets.

The SDTM Framework



Standards Governance

- CRF Impact assessment - For every new study, Trial programmers run a report against the Study Design Specifications to identify which company and franchise standards must be used, and what study-specific specifications should be created.
- Create SDTM specifications – This program automatically harvests specifications from the library to create combined study specifications for all applicable domains which are enriched with metadata e.g., adding CT values from a central lookup or flag if CT is missing from the library.
    - The machine-readable instructions are used by the SDTM pipeline to create SDTM datasets.
    - The mappings from source datasets are used by independent programmers to validate the SDTM datasets created by the SDTM pipeline.
- Create P21 input file – This program automatically creates P21 define input file from SDTM specifications.
- As part of organizational learning, reporting capability has been added to understand the degree to which standards are reused as a measure of organizational success for the standards journey. More than 90% of company and franchise level standards are reused for Pharma studies. For MD, where 75-80% are reused,

our goal is to get closer to 90% by creating more franchise level standards and continuing to enhance tools to automatically handle small variations in assessments.

When assessments are not highly standardized, agreeing to new standards is time consuming. Thus, we prioritized franchises with highly standardized CRFs, analysis and reporting. The plan is to add therapeutic area level to standards library for other franchises.

SDTM Standards Library contributions for latest studies (by franchise)

Level	Pharma	MD Vision Care	MD IOL	MD Instrumentation
Company	45%	35%	35%	Not started
Franchise	45%	40%	45%	
Total	90%	75%	80%	

STANDARDS MANAGEMENT AND GOVERNANCE

Standards Managers play a crucial role in sustaining the effective reuse of standards in the next projects and studies. They provide prompt guidance and support to programming teams.

- Standards implementation needs active management at study setup time. Standards Manager closely partner with study team to ensure correct implementation. This is a very different approach from legacy processes where varied approaches do not have long-term implications.
- Standards Managers maximize the use of standards by monitoring deviations from the library and ensuring that deviations are justified exceptions. They provide fast turn-around time when changes are needed in the library for ongoing studies. Otherwise, programming teams tend to create workarounds that introduce unnecessary duplication or misalignment with standards library.
- With the continuing expansion of the standards library, it has become even more important to check the library metadata for integrity, conformance as well as consistency. Standards Managers ensure that the library is complete, accurate and high quality. They work closely with their counterparts in Data Management to anticipate new assessments that may need new standards artifacts. They determine what is required by regulatory vs recommended and jointly decide to adjust CRFs/CTs or keep because of site considerations.

They practice continuous improvement, conduct lessons learnt sessions and feed improvements into the library. They maintain a tracker for fixes, enhancements, and wish lists with an aim for process refinement and optimization. This feedback loop is very important because standard studies mean easier integrations later.

INTERNAL CDISC EXPERTISE

The CDS project team members were not fully allocated to the project, so standards work takes lower priority than project and study work. During CDISC implementation, Pharma-related activities were prioritized. The team worked on MD as bandwidth permits.

There were many new things to learn but not enough bandwidth. The project team tapped other programmers to become extended team members and helped them develop expertise in specific CDISC areas through training, mentoring and partnering in the first studies. This divide-and-conquer approach built a layered support pyramid for the growing users of the standards library.

Selection of staff for the first studies and SME roles shows that some programmers are well suited to be early adopters while some find metadata driven automation challenging. The key is to understand and leverage each other's strengths, and pivot as needed.

ROADMAP

The feedback from first NDA was positive, a credit to the work of the last 3 years and confirmation that a single set of processes and tools for creating CDISC deliverables can work for both Pharma and MD. As we continue to extend the library for MD, a plan is also in place to further automate and explore ways to gain more efficiency.

- Enable dataflow from protocol and clinical trial management system to facilitate the automatic creation of SDTM Trial Domains and BIMO package
- Work towards generating SDTM+ by FPFV for selected MD studies for ongoing study data review by wider study team members and roles
- Create an ADaM pipeline which will facilitate the automatic generation of ADaM datasets using machine-readable instructions for highly standardized domains
- Improve management of standards with the use of workflows, versioning, audit trail, rollback and other capabilities
- Explore how AI can use the standards library for relevant use cases (e.g., map new CRF to SDTM)

CONCLUSION

MD has a high variety of data types and formats. With new innovative study designs, decentralized clinical trials and real-world data, the number, size and complexity of data collected in clinical trials continue to increase. Transitioning to CDISC standards can be planned holistically and implemented incrementally, starting with highly reused data then growing the library with every new study.

It is important to embrace the iterative nature of standards development and implementation. Unlike legacy processes that could reach enduring steady states, CDISC standards work will never 'complete' in the same way. There is always a new IG or new CT version to prepare for, and a to-do list that is always full of entries that will help teams work smarter, not harder.

Aim to streamline but do not cut essential steps. Establishing quality and complete specifications that show end-to-end traceability takes significant efforts but saves a lot of time downstream when creating submission packages and reusing standards artifacts for next studies.

A unified standards framework consisting of a well-maintained standards library plus the SDTM pipeline with built-in quality checks, combined with automation tools and strong governance, enabled Alcon to achieve high quality submission deliverables with less manual effort and rework. Starting the transition to CDISC before mandated requirements allowed the formulation of strategy that considers speed, simplicity and long-term efficiency. One set of well-planned applications, processes and tools that work for both Pharma and MD can be achieved.

CONTACT INFORMATION

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

Author Name: Bruce Chambers
Company: Veramed
Address: 5th Floor Regal House, 70 London Road, Twickenham, London, TW1 3QS, UK
Work Phone: +44 20 3696 7240
Email: bruce.chambers@veramed.com
Website: https://veramed.com