PharmaSUG 2020 - Paper SI-170

Moving A Hybrid Organization Towards CDISC Standardization

Kobie O'Brian, Sara Shoemaker, Robert Kleemann, and Kate Ostbye
SCHARP Fred Hutchinson, Seattle, WA

ABSTRACT

This paper discusses the experience of implementing standardization of data collection and data set development of submission-ready data sets at a unique organization at the intersection of Academia and Industry Partners. SCHARP (Statistical Center for HIV/AIDS Research and Prevention) at Fred Hutchinson is an academic center with a nonprofit business model. It is in a unique position requiring a balance of standard regulatory reporting requirements as well as specific sponsor needs with stakeholders including the National Institutes of Health (NIH), academic centers, nonprofit foundations, and pharmaceutical manufacturers. This requires a tailored approach using Clinical Data Interchange Standards Consortium (CDISC) standards for collecting, submitting, and analyzing data across the organization. Governance of the different CDISC implementation strategies for organization-wide data collection, storage, and analysis is discussed as well.

INTRODUCTION

The movement towards CDISC standardization for data collection, storage, analysis, and operations has been encouraged for organizations that conduct clinical trials. In particular, the FDA now requires all tabulation data submitted for drug approval to be in Study Data Tabulation Model (SDTM) format. SCHARP, as a world-class statistical and data management center, has both created a data collection global library (GLIB) aligned by Clinical Data Acquisition Standards Harmonization (CDASH), as well as incorporated important unique case report forms to meet individual sponsor requirements. SCHARP has also adopted an operational variation of SDTM known as SDTM +/- to: help prepare clinical trial study data sets at SCHARP prior to submission to regulatory authorities; for ease and consistency of internal and external reporting; for the development of Analysis Data Model (ADaM) data sets; as well as for data sharing best practices. SCHARP had successfully implemented SDTM+/- by using an internally programmed derived data set platform called Delphi to transform electronically collected data into operational SDTM data sets and SDTM domains. These are some of the ways that will be described in greater details in this paper where SCHARP has moved our dynamic statistical data center towards best practices for data collection, curation, organization, management, analysis, and regulatory reporting and submission. Figure 1 shows high level view of SCHARP's Data Standards Model.

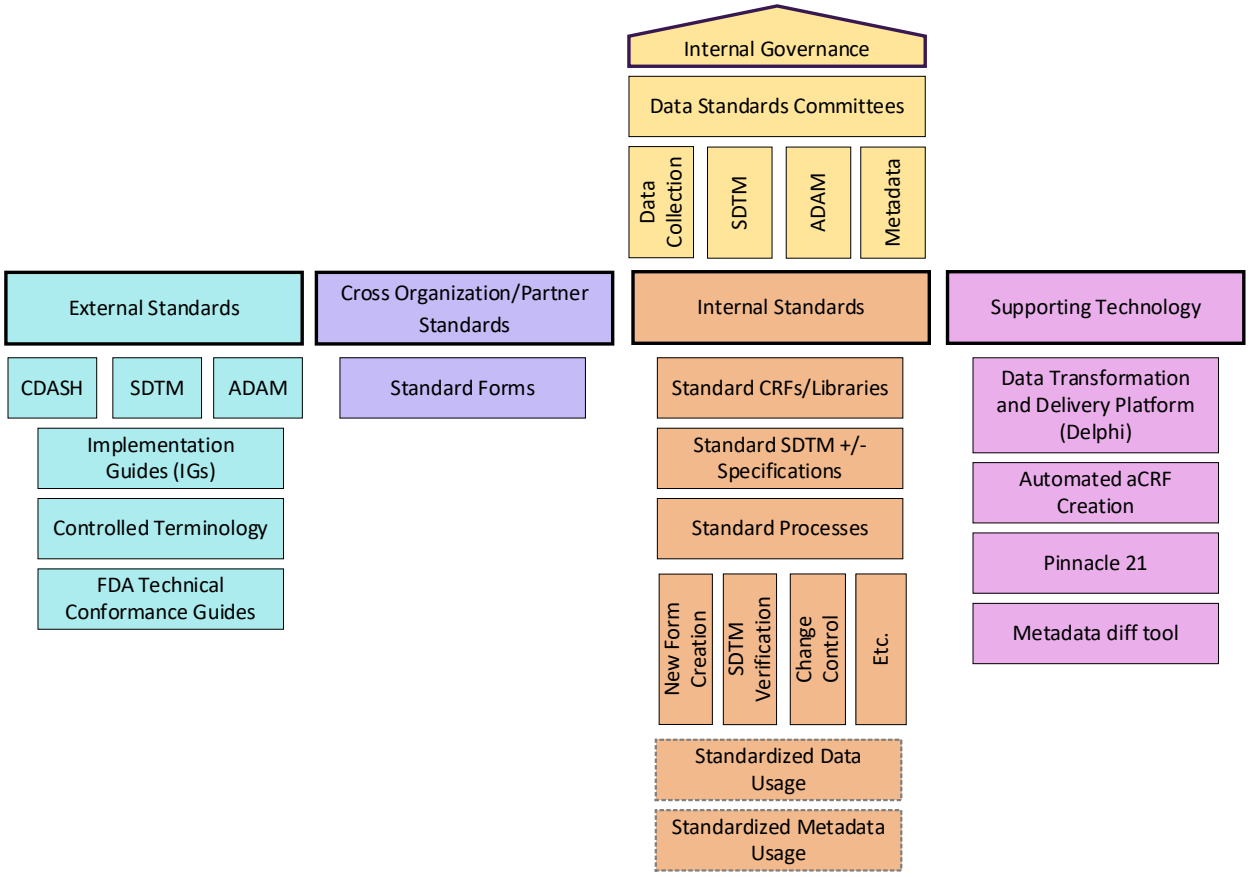

Figure 1. SCHARP Data Standards Infrastructure.

ABOUT SCHARP

Established in 1992 to support global HIV prevention research at Fred Hutchinson in Seattle, SCHARP operates according to all national and international regulatory standards. SCHARP holds many teams that are highly skilled in programming, biostatistics, IT data systems, quality assurance, project management and management of clinical and lab data that is collected from multiple sources and sites worldwide. SCHARP supports clinical researchers and administrators with high quality clinical and laboratory data management and statistical services of preclinical through phase 4 clinical trials. SCHARP has supported more than 200 clinical trials and epidemiologic studies including several high-profile international clinical trials. [1]

Our organization does this while operating from a unique business perspective compared to most clinical research trial data management centers: SCHARP is an academic center with a nonprofit business model. Our center relies on grant funding from major governmental and private sector sources. SCHARP is affiliated with many academic centers but is a nonprofit organization. As such, we have to find ways to work closely with our DAIDS Network Partners such as HIV Prevention Trials Network (HPTN), HIV Vaccine Trials Network (HVTN), and Microbicide Trials Network (MTN) in developing unique studies as well as meet varied individual sponsor requirements.

ABOUT CDISC STANDARDS

Clinical Data Interchange Standards Consortium (CDISC) is an organization that was created to develop data exchange standards to streamline clinical research. CDISC Foundational Standards are the basis of a complete suite of standards, supporting clinical and non-clinical research processes from end to end. These

standards allow researchers and sponsors to submit data to the FDA or other regulatory agencies in electronic format that is uniform. CDISC Foundational Standards focus on the core principles for defining data standards and include models, domains, and specifications for data representation. The standards SCHARP focuses on implementing are: Clinical Data Acquisition Standards Harmonization (CDASH) for data collection, Study Data Tabulation Model (SDTM) for structuring data that simplifies the collecting, managing, and reporting of results, and Analysis Data Model (ADaM) for creating data sets for analyses. For more information about CDISC and the different levels of standardizations, please visit their extensive website at www.cdisc.org.

GARBAGE IN-GARBAGE OUT

Standardization of data collection, structuring, and analysis is important. Why? The quality of our research output is contingent on data quality being held to highest standards while the data is collected, organized, stored, and analyzed in order to publish reproducible, complete, quality data. The best way to do this is to standardize data management following Industry Standards as published by CDISC -- from study start with the Case Report Forms using CDASH all the way to SDTM/ADaM data sets and accompanying quality assurance documentation. Benefits of having standards that follow CDISC [2] include:

- Fostered efficiency
- Complete traceability
- Enhanced innovation
- Improved data quality
- Facilitated data sharing
- Reduced costs
- Increased predictability
- Streamlined processes

Organizations should adhere to data standardization not just because it is FDA and often Sponsor-required, but also they can avoid Garbage In-Garbage Out issues with their data: The quality of analysis and study results is contingent on the quality of the data collected, stored, organized, and managed.

DATA COLLECTION STANDARDS

ALIGNING GLOBAL LIBRARY TO CDASH STANDARDS

One of the first areas in SCHARP where data standards were identified as urgently needed was in structuring how data is collected in Case Report Forms (CRFs) and with our Electronic Data Capture System, Medidata Rave. We have multiple studies from diverse Network partners with a dizzying array of questions and variable names that were not standardized. Sometimes there could be two identical questions in two different studies with identical responses, but the order would be different or there would be an extra response. Due to limitations in Medidata Rave, if a code list for the same question differs between implementations, it requires a new code list to be created. This can even happen within the same study build, requiring a new variable name for basically the same exact information.

With leaning on CDASH standards in creation of standard forms and common questions and variables names, we can create uniformity in not only how questions are asked, but possible responses that are analyzable within and across different studies. We created SCHARP Global Libraries (GLIBs) aligned with our network partners that has an oversight governance committee to make sure there is both alignment with CDISC standards as well as support of the differing needs of different types of study designs. This committee also makes sure any changes to the standards are vetted by different departments within SCHARP to make sure there are no unforeseen downstream consequences from the changes. To ensure there was buy-in with this structure, we worked closely with our Network Partners and Sponsors explaining the benefits of standardization and the ability, and this is key, to be flexible. Not all study designs are identical, and we recognize the need to allow flexibility within the implementation of standards. As I like to

say, GLIBs are like a buffet where there are many dishes to choose from that are made uniformly in the kitchen but each plate a customer puts together is as unique as they are. We are still in the process of completing the Global Library Project for our three biggest networks: HVTN, MTN, and HPTN. We were able to merge MTN and HPTN into one GLIB since their study designs have many similarities. With HVTN and the many vaccine studies, there is a need for a more extensive GLIB with forms that are used in multiple studies that wouldn't be used in other Networks.

TAILORING STANDARDS BY NETWORK

CDISC standards are wonderful launching pads for setting up clinical trial data collection and organization. In some cases, there are Implementation Guides (IG) specific to unique populations such as Associated Persons IG. There are also helpful and important Therapeutic Areas User Guides (TAUG) – the HIV TAUG [3] is one that SCHARP helped draft with CDISC and we use it often. Also, the CDASH and SDTM Controlled Terminology (CT) and Sample CRFs have an array of information that grow every year. That said, there are many different aspects of a research study that CDISC does not fully address. FDA Technical Conformance Guides [4] and CDISC itself often say the best approach towards working with data that isn't within the IGs, CTs, or TAUGs is to develop internal standardization plans and to stick to them. We work with our Network partners, as they are the subject matter experts, to make sure our standards make sense. For example, with collaborating closely with our network partners, we are in the process of standardizing our Global Libraries of variables and forms for electronic Case Report Forms (eCRFs) by network. We work closely with our network partners who develop the case report forms to get their input on which questions and variables we can standardize. We first work with CDISC standards to create the basis of commonly used forms but need to make sure anything that falls outside those channels are identified, defined, and implemented, too.

NEW VARIABLE AND FORM CREATION STANDARDS

Another component of the data standardization process that is important on the data collection side is how new variable and form names are created when they are not already in the Global Library. We have created a Standard Operating Procedure (SOP) where we have standardized the request and completion process. First we make sure the new form or question hasn't been used before. If so, we will use the old variable and form name, and amend if necessary to keep close to CDISC alignment. If there is no precedent, then we look to guidance from the CDASH and/or SDTM IGs and/or CTs. All this work on the front end of data collection bears fruit for when we move on to the next steps in our data standardization process: mapping data from various data sources into SDTM data sets.

NEXT STEPS FOR SCHARP DATA COLLECTION STANDARDS IMPLEMENTATION

We are working on a Division of AIDS (DAIDS) -driven focus working group to standardize pregnancy forms and to handle data collection for mother-infant studies. Working in close partnership with other sister organizations such as Frontier Science and the Office of HIV/AIDS Network Coordination (HANC), we are designing CRFs, variable names, and SDTM mapping for these unique patient populations in HIV and Vaccine studies. Additionally, at the completion of our Global Library project, our EDC team will design standard eCRFs from the GLIBs, streamlining the study build process.

OPERATIONAL AND STANDARD SDTM STANDARDS

ALIGNING STUDY DATA SETS TO SDTM STANDARDS

The next area in SCHARP where data standards were implemented was in the organization of study data after collection. The FDA has issued guidance recommending the use of SDTM for organization of study data that is to be submitted for drug approval. SCHARP has decided to use SDTM not only for study data that will be submitted to the FDA, but as a standardization bellwether for all new studies. Some aspects of SDTM compliance are not helpful in the maintenance of study data for internal use during the life of a study which is why the SDTM+/- standards and resultant data sets have been developed. This standard is to be

as close as possible to SDTM compliance, but still provide the data in a way that makes sense for operational use by SCHARP. These standards include how the data is described, defined in formats in the data sets, how they are named, and mapped. Most of these are within the SCHARP standard SDTM Specifications, which include Metadata and Conversion mapping.

STANDARD SPECIFICATIONS

First steps for developing SDTM data sets using SCHARP SDTM+/- standards is to update the standard Metadata and Conversion specifications to be study-specific. We have specifications that are DAIDS network specific: MTN, HPTN, and HVTN. The Metadata specifications are based off the SDTM Implementation Guide (IG). "The SDTM is built around the concept of observations collected about subjects who participated in a clinical study. Each observation can be described by a series of variables, corresponding to a row in a data set or table. Each variable can be classified according to its Role. A Role determines the type of information conveyed by the variable about each distinct observation and how it can be used." [5] These observations are then collected for each subject into domains. A domain is a collection of observations with a topic-specific commonality about a subject. The structure of each data set is a flat file representing a table. Each row is a single observation and each column is a variable.

At SCHARP we have a Metadata Specifications excel document that provides information about the variables used in all the data sets. And we have a separate Conversion Specifications spreadsheet that breaks down information by domain and contains mapping specifications, data dictionaries, controlled terminology and other relevant information that completes the traceability loop from CRF data sets to SDTM data set development. It is within these specifications that the Data Standards Analysts at SCHARP document, map, and describe decisions made that are study-specific in order to develop the Operational and SDTM data sets. These specifications also house aCRF mapping information sourced by the Automated aCRF process described later in this paper.

It is important to have standardization at this level within an organization for developing SDTM data sets in order to add efficiencies in specifications mapping, but it also allows us flexibility in adapting these processes not just for studies that will have drug submissions to the FDA, but to also organize the data in a way that can be used for other purposes within SCHARP for all studies. This creates uniformity and reproducibility within networks between different studies, while remaining flexible enough to allow variances as identified by the unique nature of many of our studies. Additionally, and importantly, this adds a level of data quality and assurance with transparency and strong traceability as data is collected, entered into EDC systems, organized into SDTM data sets, and then available for data set consumers on a secured datamart or through FTP file transfer.

SDTM DATA SET DEVELOPMENT USING DELPHI

SCHARP has created a platform internally to develop derived data sets called the Delphi System. If you reference PharmaSUG 2020 - Paper AD-130 "A Novel Solution for Converting Case Report Form Data to SDTM Using Configurable Transformations", you can learn about SCHARP's solution for SDTM conversion from CRF data sets in greater detail. This platform has a suite of tools that data set developers can use to transform CRF data from RAVE data sets into operational SDTM and SDTM data sets. The output of the data sets is in a centralized datamart. This validated system is 21 CFR Part 11 compliant, satisfying both our DAIDS sponsor and FDA submission requirements. SDTM data set developers specify and configure the data set using configurable transformations and expression mapping both using command line tools and JEXL/json programming within data set directories and study schemas. The consumers of the operational SDTM data sets are internal departments for report generation. The SDTM data sets are consumed by the Data Analytics Unit to develop ADaM data sets. Additionally, the final SDTM data sets are to be run through Pinnacle 21 for compliance checks and Define.xml and Clinical Study Data Reviewer's Guide (SDRG).

VALIDATION

Validation of data set development from raw study data sets to SDTM data sets is important. The industry standard use of Pinnacle 21 is also used at SCHARP where we can validate the data against FDA business rules and receive a Data Fitness score. This score describes SDTM Compliance, Controlled Terminology, Regulatory Conformance, Metadata, Data Quality and Analysis Support. This is also a powerful tool used while mapping data into SDTM data sets in helping finds errors or missteps. We also have a tool within Delphi called the “diff tool” that allows us to further identify differences between our Metadata conversion specifications and the SDTM data sets. Everything from variable format to missing domains can be caught with this tool during data set development.

AUTOMATING SDTM ANNOTATED CRFS

An important part of traceability from study data collection to mapping into SDTM data sets is the generation of the Annotated Case Report Form (aCRF). This document maps clinical data collection fields to the corresponding variables that are in the SDTM data sets. Typically, this is a very manual and time-consuming process. If you reference PharmaSUG 2020 - Paper SS-159 “Automating CRF Annotation using Python”, you will see efforts at SCHARP in using Python to automate the generation of the aCRFs. We developed SCHARP SDTM aCRF Standards in both the generation and formatting of the aCRFs, in order to adhere to a consistent and replicable process that can be carried on by different study protocols across different Networks.

NEXT STEPS FOR SCHARP SDTM +/- STANDARDS IMPLEMENTATION

Within SCHARP, we have aspired to create SDTM compliant data sets that are submission-ready. We are still in the process of operationalizing the strategy towards that end by identifying FDA requirements for Clinical Study Reports (CSR) in collaboration with a Study Sponsor for a pivotal study. We are using this opportunity to find ways to use the CSR process to further reinforce and strengthen our data standards processes. This includes the creation of the Define.xml and the Clinical Study Data Reviewer's Guide.

GOVERNANCE

Where CDISC standards and SCHARP align and the standards hit the ground, there is a strong need for governance. SCHARP has a Data Standards Committee (DSC) with representatives from multiple units within SCHARP. The Data Standards Committee members leverage their experience, subject matter expertise, and insight from their professional capacity and experience to oversee SCHARP data standards and maintain its alignment with the organization’s mission and objectives. This committee has three (3) sub-committees that focus on the governance of specific standard content, management of study-specific requests, and day-to-day requirements associated with the following:

- Data Collection Standards Subcommittee
    - CRF Content
    - Edit Checks
    - Custom functions
    - Data Dictionaries – standard codelists utilizing CDISC Controlled Terminology (CT) where applicable
    - Lab Analyte units and formats
    - Change control process for Global Library forms and associated Change Advisory Board (CAB)
- SDTM +/- Subcommittee
    - Standard SCHARP SDTM Conversion specifications and associated Metadata
    - Standard data set configuration file
    - Study-specific SCHARP SDTM Conversion specifications and associated Metadata
    - Study-specific data set configuration file
    - Trial Design templates and instructions
    - Change control process for SCHARP SDTM and associated Change Advisory Board (CAB)

- ADaM Subcommittee
    - SCHARP ADaM standards
    - SCHARP ADaM data set specifications and documentation
    - SCHARP ADaM and analysis data sets
    - Change control process for SCHARP ADaM Specifications and associated Change Advisory Board (CAB)

These subcommittees work in parallel to ensure that data standards are developed, implemented, and/or changed, taking into account different department needs and perspectives in a manner that is transparent and traceable. Changes to any of the approved standards are governed by the Change Advisory Board (CAB) and controlled through SOPs to maintain data quality assurance. For example, a change in a Global Library may cause a downstream effect in IND report development code. It is important to make sure such changes wouldn't have any negative effects and are communicated broadly so there aren't any surprises. If a change to the standards is indicated, a request has to be made, it is reviewed by members of the specific CAB depending on what standards need to be changed, a vote is taken and documented, and if approved, the Change Request can be fulfilled, typically by a member of the Data Standards Team. Communication about governance within and across departments at SCHARP as well as with our Network partners is important in leading an organization like SCHARP towards CDISC Standardization.

NEXT STEPS FOR SCHARP DATA STANDARDS STRATEGY

Our next steps at SCHARP for further incorporating CDISC standards is to implement ADaM using our SDTM data sets. Our ADaM 2.0 Project is going to standardize the use of ADaM across the Data Analytics Unit and potentially other areas of SCHARP. Additionally, it has become increasingly apparent, now that CDISC aligned data standardization at SCHARP has started taking effect, that there are additional areas where data standardization will be of benefit. On the horizon is defining an Operational Metadata Strategy at SCHARP with governance through the DSC. While not aligned with CDISC standards, finding solutions to consolidate operational metadata into centralized systems with governance through the newly chartered Metadata Subcommittee for streamlining reporting, milestone tracking, sharing of data sets, and other related processes is the next logical set for continuing to strive towards meeting our Network's needs and optimize data management. Finally, there is ongoing planning and development of pipelines of non-CRF data into Delphi. Currently assay and related laboratory data as well as qualitative survey instruments not in Rave but in other EDC systems like RedCAP are not transformed into SDTM data sets. Once the data pipelines are centralized into Delphi, these sources of data will then be organized into SCHARP SDTM +- data sets and later into ADaM.

CONCLUSION

Implementing data standardization at a unique organization such as SCHARP, which lies at the intersection of Academia and Industry, has proven to have interesting challenges. Due to the need to balance standard regulatory reporting requirements as well as specific sponsor needs with different types of stakeholders, we have tailored a multi-prong approach towards standardizing how data is collected, organized, and analyzed. While a work in progress, SCHARP has made great strides to maintain the center's Mission: "To provide world-class statistical and data management services". [6]

REFERENCES

1. Statistical Center for HIV/AIDS Research and Prevention (SCHARP) website. *Fred Hutchinson Cancer Research Center*. Accessed 20 March 2020. Weblink: https://www.fredhutch.org/en/research/divisions/vaccine-infectious-disease-division/research/biostatistics-bioinformatics-and-epidemiology/statistical-center-for-hiv-aids-research-and-prevention.html

2. CDISC Standards in the Clinical Research Process. *Clinical Data Interchange Standards Consortium (CDISC)*. Accessed 20 March 2020. . Weblink: https://www.cdisc.org/standards
3. HIV Therapeutic Area User Guide v1.0. *Clinical Data Interchange Standards Consortium (CDISC)*. Accessed 21 May 2020. Weblink: https://www.cdisc.org/standards/therapeutic-areas/hiv/hiv-therapeutic-area-user-guide-v1-0
4. Study Data Technical Conformance Guide v4.5. *Food and Drug Administration (FDA)*. Accessed 21 May 2020. Weblink: https://www.fda.gov/media/136460/download
5. SDTM Standards. *Clinical Data Interchange Standards Consortium (CDISC)*. Accessed 20 March 2020. Weblink: https://www.cdisc.org/standards/foundational/sdtm
6. About SCHARP. *Fred Hutchinson Cancer Research Center*. Accessed: 20 March 2020. Weblink: https://www.fredhutch.org/en/research/divisions/vaccine-infectious-disease-division/research/biostatistics-bioinformatics-and-epidemiology/statistical-center-for-hiv-aids-research-and-prevention/about.html

CONTACT INFORMATION

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

Kobie O'Brian
SCHARP, Fred Hutchinson
206.667.6645
kobrian@scharp.org
SCHARP Website

BIOS

Kobie O'Brian is the Lead Data Standards Analyst with the Statistical Center for HIV/AIDS Research and Prevention (SCHARP) at Fred Hutchinson and has been working in clinical research for 10 years and data management for 4 years.

Sara Shoemaker is a Technical Project Manager at SCHARP and has been working in the field of software development for 25 years.

Robert Kleemann is the Software Development Manager at SCHARP and has been designing and implementing software systems for 30 years and managing software teams for 10 years.

Kate Ostbye is the Director of Programming at SCHARP and provides leadership, oversight and mentorship to a dedicated team of lab, clinical and statistical programmers as well as SCHARP's CDISC and Data Standards program.