

Want Submission-Ready Datasets Package from the Get-Go?

Kishore Pothuri, Vita Data Sciences, Waltham, MA, USA
Bhavin Busa, Vita Data Sciences, Waltham, MA, USA

ABSTRACT

As per the Study Data Technical Conformance Guide, each submitted SDTM/ADaM dataset should have its contents described with complete metadata in the define.xml file. It is expected that an aCRF be included along with the SDTM domains per the Metadata Submission Guidelines. In addition, it is strongly recommended to have a Study Data Reviewer's Guide (cSDRG) for SDTM and Analysis Data Reviewer's Guide (ADRG) for analysis data submission. And so, to get to a 'submission-ready' stage for study datasets, it is not just about the datasets itself, but also requires one to check for data validation rules and to prepare all other submission components that go along with it. In this paper, we will provide a detailed approach and our recommendations to build efficiencies in developing CDISC datasets and all required submission documents. And provide step-by-step process to create submission-ready dataset package towards NDA submission from the get-go.

INTRODUCTION

Study Data Standards are required in clinical and nonclinical studies that start after December 17, 2016 [1]. Technical rejection criteria are being added to the existing eCTD validation criteria to enforce the deadlines. The FDA may refuse to file (RTF) for NDAs or refuse to receive (RTR) for ANDAs, an electronic submission that does not have study data in conformance to the required standards specified in the FDA Data Standards Catalog.

As per the Study Data Technical Conformance Guide, each submitted SDTM/ADaM dataset should have its contents described with complete metadata in the define.xml file. It is expected that an aCRF be included along with the SDTM domains per the Metadata Submission Guidelines. In addition, it is strongly recommended to have a Study Data Reviewer's Guide (SDRG) for SDTM and Analysis Data Reviewer's Guide (ADRG) for analysis data submission. The submission of the study datasets and any associated files should follow a specific file directory structure per the eCTD requirements.

So, it is very important for the study programmer(s) to adhere to the standards at each and every step of the development phase.

TYPICAL SDTM DEVELOPMENT LIFE CYCLE

The Figure 1 below demonstrates typical SDTM development life cycle [2]. The number annotated on each box represents the process and sequence in which they are developed and implemented. Upon development and QC of the SDTM datasets, a programmer passes the datasets through Pinnacle21 validator to check for the compliance. In cases where the study specific define.xml is available, it is also passed through the validator along with the SDTM datasets. This ensures the datasets and the define document are in sync and that both the components are checked at the same time for the compliance. Also, one should note that every time SDTM datasets are generated during the study period (interim data snapshot, blinded data review, dry run TLFs, DMC, etc.), one may or may not run their datasets through Pinnacle 21 validator to check for compliance in an ongoing manner. However, if one chooses to run each time, this additional step of checking for the compliance adds up significant amount of time to the SDTM development life cycle.

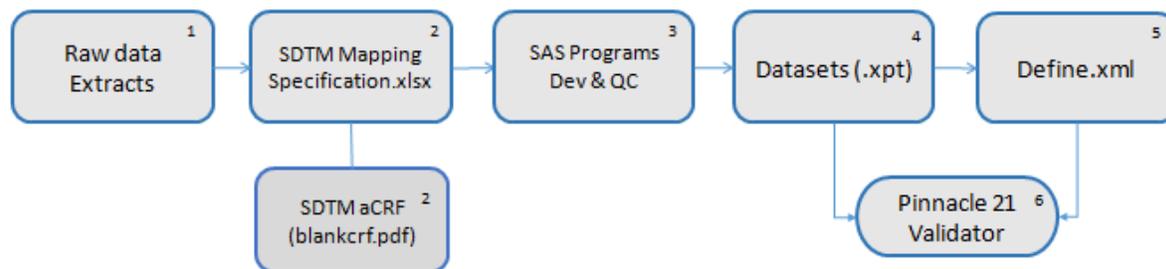


Figure 1. Typical process followed in the industry for SDTM development

A COMPREHENSIVE APPROACH TOWARDS A SUBMISSION-READY DATASETS PACKAGE

In this paper, we shall discuss a step-by-step approach and our recommendations for creating a submission-ready SDTM datasets package from the get-go. Same process could be applied for ADaM datasets package.

Below 3 main steps illustrate the process involved in creating a submission-ready SDTM datasets package:

1. SDTM Mapping Specifications Development

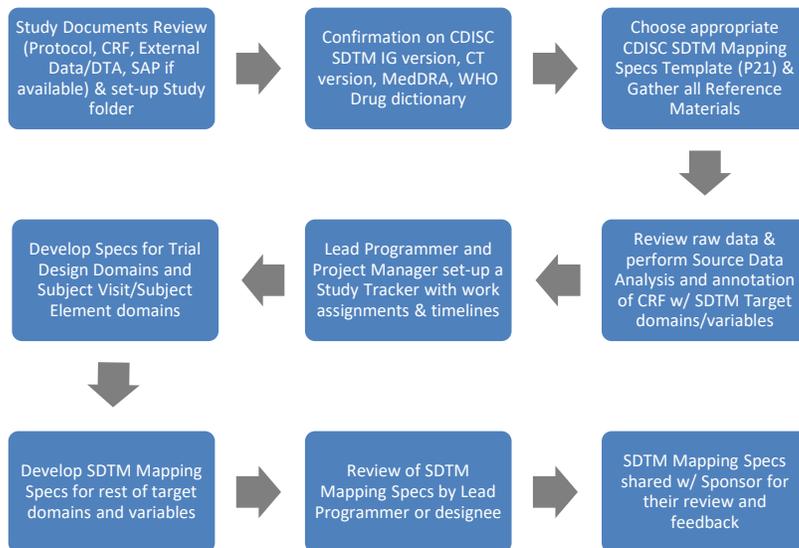


Figure 2. Process Flow for SDTM Mapping Specifications Development

1.1 Source Documents & Folder Set-up: The Source documents including but not limited to protocol, CRF, external data, data transfer specifications if any and SAP should be collected and stored in the study folder after it is set-up and then should be reviewed by the programming team.

1.2 Confirmation for the Standard Documents: Before the commencement of SDTM mapping specifications development, the programmers should be aware of the SDTM IG, SDTM CT, MedDRA and WHODrug Dictionary versions to be used for a particular study.

1.3 Mapping Specifications Template: Based on the SDTM IG version, the appropriate version of mapping specifications template from Pinnacle21 Community should be selected along with all the necessary documents.

1.4 Raw Data Extraction & Review: Even though most of the databases have edit checks built in, the raw data should be checked further for any inconsistencies.

Source Data Analysis: The data should be assessed in conjunction with the study protocol, CRF and SAP (if available) for creating a list of possible SDTM domains for a particular study.

CRF Annotations: It is recommended to do hand annotations first and then followed by the annotations using the Adobe Acrobat DC.

- Hand Annotations:
 - Based on the source data analysis, a hard copy of the latest available unique version of CRF should be hand annotated first, to get a clear picture of all the SDTM domains needed for a particular study.
 - The hand annotations should include SDTM domain/variable names and labels.
 - The hand annotations should also undergo a visual QC by the team members for content check and accuracy.
- SDTM Annotations:
 - Based on the hand annotations, the CRF should be annotated using Adobe Acrobat DC.
 - The aCRF annotations should be based on the guidelines provided in MSG (Metadata Submission Guidelines) [3].
 - The Bookmarks (By Domain & By Visit) should be included only after the final version of eCRF is available and preferably after the database lock.

NOTE: The latest version of CRF available is used for the annotations and during the course of trial there is every chance of CRF getting updated. So, until the version of CRF is finalized, this will be an ongoing process.

1.5 Mapping Specifications Development for Non-CRF domains and Subject Visits: SDTM Mapping Specifications for Trial Design Domains (TDD), Subject Visits & Subject Element domains should be developed first.

1.6 Mapping Specifications Development Continued: Development of SDTM mapping specifications for rest of the domains while referring to the mapping specifications for TDD & Subject Visits wherever applicable.

- All the required metadata specification components should be properly arranged into a single file and any additional columns that you may need should be added as and where necessary.
- Supplemental documentation if necessary should also be included as a pdf, which can be accessed from the define.xml

1.7 Review of SDTM Mapping Specifications: The SDTM mapping specifications should be reviewed thoroughly before they are sent for the review by the sponsor.

- Cross-functional review & feedback by the programming team
- Review & feedback by co-lead programmer
- Senior level review & feedback by lead programmer

1.8 SDTM Mapping Specifications for Sponsor Review: When the SDTM mapping specifications are updated based on the internal review process, they are sent for the review & feedback by sponsor.

Benefits of this approach: The programming team will be well educated about the study design and purpose of the study, will have a thorough understanding of the data, the folder structure, the CRF design, required SDTM domains, variables, derivations. Also, this approach will yield a high quality SDTM mapping specifications, which can be used to kick start programming without any hurdles.

2. Programming and Validation of SDTM Domains:

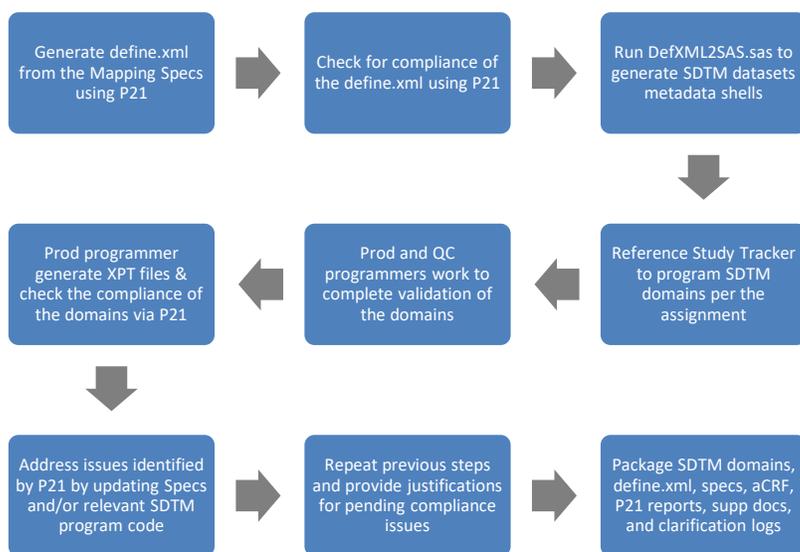


Figure 3. Process Flow for Production and Validation of SDTM Domains

2.1 Define.xml Generation

- Based on the review & feedback by sponsor, the SDTM mapping specifications should be updated if needed.
- **Development and Validation of Define.xml:**
 - The completed mapping (define) specifications should be passed into Pinnacle21 for generating define.xml [3].
 - After the define.xml is generated, it is to be checked for all the hyperlinks, navigation within the document and as well as navigation to the aCRF, supplemental documentation, cSDRG or ADRG.

2.2 Compliance Checks

- For compliance check, the define.xml should be again passed into Pinnacle21 and then validated.
- Based on the P21 report, the define specifications should be updated as needed to fix the ERRORS and WARNINGS and then entire process can be repeated until the most compliant version of the define.xml is ready.

NOTE: This version of define.xml will only be a draft version and is most likely to be updated during the course of trial as and when needed.

2.3 Metadata Shells: A utility macro or tool can be used to generate SDTM domain metadata shells, which should be used in the programming for variable names, labels and attributes.

2.4 SDTM Domains Development: As per the study tracker, the assigned SDTM domains should be programmed by the production and validation programmer respectively.

2.5 Validation of SDTM Domains: The production and validation programmers should work on any discrepancies, mismatches that resulted after PROC COMPARE, to complete the validation of the SDTM domains.

2.6 Datasets (. XPTs):

- The production programmer should generate the XPT files using the validated SDTM domains.

Compliance Checks for SDTM Domains using Pinnacle 21 Validator:

- The XPTs along with define.xml should be passed into P21 Validator for compliance checks.
- The appropriate SDTM IG, SDTM CT and MedDRA versions should be selected from the drop down in the tool.

2.7 Updates based on Compliance Report: The compliance report should be checked for any ERRORS/WARNINGS and if required necessary updates should be made to the SDTM mapping specifications or SDTM programming or both.

2.8 Justifications for Pending Issues: In order to address the compliance issues, if the SDTM mapping specifications are updated, then define.xml should be generated again to have these changes reflected and validated. For any pending compliance issues, a column can be added on the issue summary sheet of compliance report for justifications.

Define.xml Finalization:

- Based on the compliance report, if needed the define.xml should be updated and validated by two ways
 - Validation of define.xml alone using P21 Validator
 - Validation of define.xml along with XPTs using P21 Validator

aCRF Finalization:

- Based on the updated SDTM domains and define.xml, if needed the annotations should be also updated on the latest CRF version available.

2.9 Package Development: All the components including validated SDTM domains, SDTM mapping specifications, define.xml, supplemental docs, P21 compliance reports, aCRF and clarification logs should be packaged together.

Benefits of this approach: This approach will yield in CDISC complaint SDTM domains, define.xml and aCRF. All these components are packaged together and are almost submission-ready. This process will avoid the back n forth review and updates between the vendor and the sponsor towards the end.

3. Final Delivery Folder Set-Up

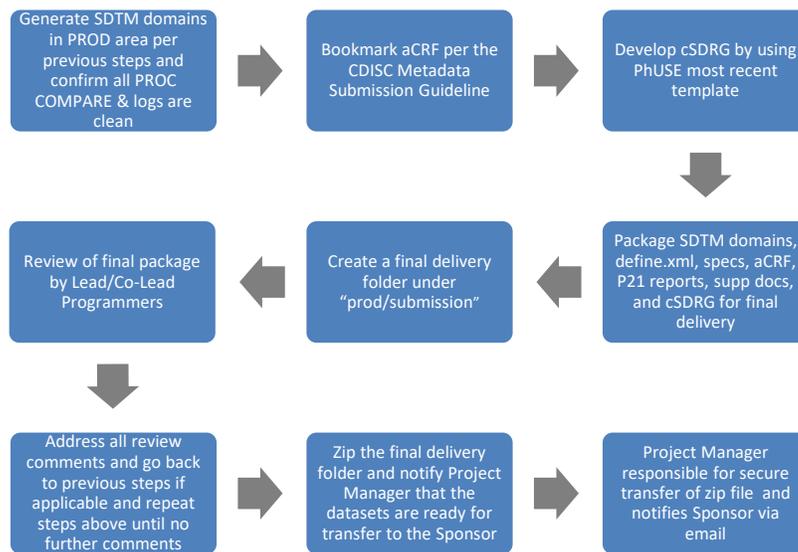


Figure 4. Process Flow for Development of Final Delivery Folder

3.1 Final SDTM Run: The SDTM domains should be generated in the PROD environment as per the process described previously. The programming team must confirm that all PROC COMPARE outputs and logs are clean.

3.2 Bookmarking of aCRF: Once the aCRF is finalized, then it is bookmarked (by domain, by visit) as per CDISC Metadata Submission Guidelines.

3.3 Clinical Study Data Reviewers Guide (cSDRG): By using the most recent template from Phuse, the cSDRG should be developed and reviewed internally [4].

3.4 Package Development: Within in the PROD environment, all the components including validated SDTM domains, SDTM mapping specifications, define.xml, supplemental docs, P21 compliance reports, aCRF, clarification logs and cSDRG should be packaged together.

3.5 Final Delivery Folder: A final delivery folder should be created under submissions folder on PROD environment where the final package is placed.

- 3.6 Final Package Review:** The final package should be reviewed by the co-lead and lead programmers.
- 3.7 Final Internal Review:** Based on the review & feedback provided by the co-lead and lead programmers, if necessary the programming team will repeat all the steps discussed above until no further comments given.
- 3.8 Submission Zip File:** The lead programmer should create a zip file for the final delivery folder and notify the project manager that the SDTM dataset package is ready for transfer to the sponsor.
- 3.9 Delivery to the Sponsor:** The project manager should ensure a secure transfer of the zip file and should notify the sponsor via email.

Benefits of this approach: This approach will help in doing multiple levels of reviews and streamlining the submission process which ultimately result in developing submission-ready datasets package with highest quality.

CONCLUSION:

In order to develop a submission-ready datasets package, there is no need for any innovative methods or sophisticated tools, but adherence to standards throughout the process and good programming practices play a key role in the development of a submission-ready datasets package.

The step by step approach discussed in this paper will not only provide guidance to develop a submission-ready dataset package of highest quality, but also streamlines the process which can be implemented for any number of studies.

REFERENCES

- [1] Ron, Fitzmartin and Ginny, Hussong. "Required Electronic Submissions to CDER / CBER." fda.gov. 10-08-2015 Available at <https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM511237.pdf>
- [2] Kishore Pothuri, Bhavin Busa. "Building Efficiency and Quality in SDTM Development Cycle" <https://www.pharmasug.org/proceedings/2016/PO/PharmaSUG-2016-PO24.pdf>
- [3] Vara Prasad Sakampally, Bhavin Busa. "Consider Define.xml Generation during Development of CDISC Dataset Mapping Specifications" <https://www.pharmasug.org/proceedings/2016/AD/PharmaSUG-2016-AD18.pdf>
- [4] Clinical Study Data Reviewer's Guide Template http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

CONTACT INFORMATION

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

Author Name: Kishore Pothuri
Company: Vita Data Sciences, a division of Softworld, Inc.
Address: 281 Winter St
City / Postcode: Waltham, MA 02451
Work Phone: 781-833-0259
Email: kpothuri@vitadatasciences.com
Web: www.vitadatasciences.com

Author Name: Bhavin Busa
Company: Vita Data Sciences, a division of Softworld, Inc.
Address: 281 Winter St
City / Postcode: Waltham, MA 02451
Work Phone: 781-373-8455
Email: bbusa@vitadatasciences.com
Web: www.vitadatasciences.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.