PharmaSUG 2023 - Paper MM-272

Traceability: Not just about Data

Rohit Nagpal, Kite Pharma, Inc.
Kavitha Mullela, Experis Solutions

ABSTRACT

Sometimes in clinical trials we see gaps between data, documentation and process which leads to questions from cross-functional teams and Regulatory Agencies. To overcome these issues, we can consider traceability, not just in the context of data, but as a combination of data, documentation, and process involving all cross functional teams from data collection to submission.

Traceability has many facets and is a joint effort that requires collaboration across all cross-functional teams. To accomplish this during the lifecycle of clinical trials, we use different documentation trackers like change control, database migration change, study status, process checklists and delivery approval form besides the standard documents. By using these additional documents, we can easily track database migration changes, changes during study lifecycle, recreate outputs at any point of clinical trials, ensure quality outputs, provide information to cross-functional teams, and can provide crucial information for integrated analysis.

This paper provides examples of traceability concepts that can be used to produce a robust trail to achieve better outcomes which helps us in improving compliance, risk mitigation, better data integrity, reliability, and less follow-up questions from Regulatory Agencies.

INTRODUCTION

Traceability is to track something as it moves through its product lifecycle, and it is data in the case of clinical trials. It helps to create transparency, accountability and keep tracks of what, who and when tasks are performed. It can be viewed as a two-pronged approach of data and document tracking within an organization.

Well established traceability standards not only help achieve better overall compliance, but also enhance data integrity and reliability, reduce risk, and improves the process to achieve greater quality controls.

Figure 1. Traceability

In the Pharmaceutical and Biotech industry data traceability standards are now well-known, whereas establishing traceability is still one of the most challenging aspects associated with any data conversion. To address this, the document traceability can play a key supporting role. Document traceability is about progress and delivery tracking, process and file level tracking and help support submissions.

The ADaMIG states: To assist review, ADaM datasets and metadata must clearly communicate how the ADaM datasets were created. Data traceability permits an understanding of the relationships between the analysis results (tables, listings, and figures in the study report), analysis datasets, tabulation datasets, and source data. These links are established by metadata and datapoint traceability.

The successful implementation of traceability in clinical trials, helps to trace where the drug development process has been at any given time. Reliable information and appropriate data access can provide visibility and can streamline regulatory responses to safeguard patients and the drug development process. This paper will briefly talk about data traceability and its two levels: metadata and datapoint and introduce examples for document traceability.

TRACEABILITY IN THE CLINICAL TRAIL DATA FLOW

In the clinical trial process, forward traceability is the ability to produce analysis result from raw data to study submission. The ability to trace processed data back to its source and the logic that processed it is known as backward traceability.

A key component of data quality is traceability, which is a requirement for submission to regulatory agencies. It is essential for maintaining the accuracy of source data and supporting clinical research findings from data collection to final analysis. Sponsors must showcase, in regulatory submissions, that the information in a submission package can be directly linked to the original source data in an unbroken chain, considering any possible data transformations or derivations used to process the data.

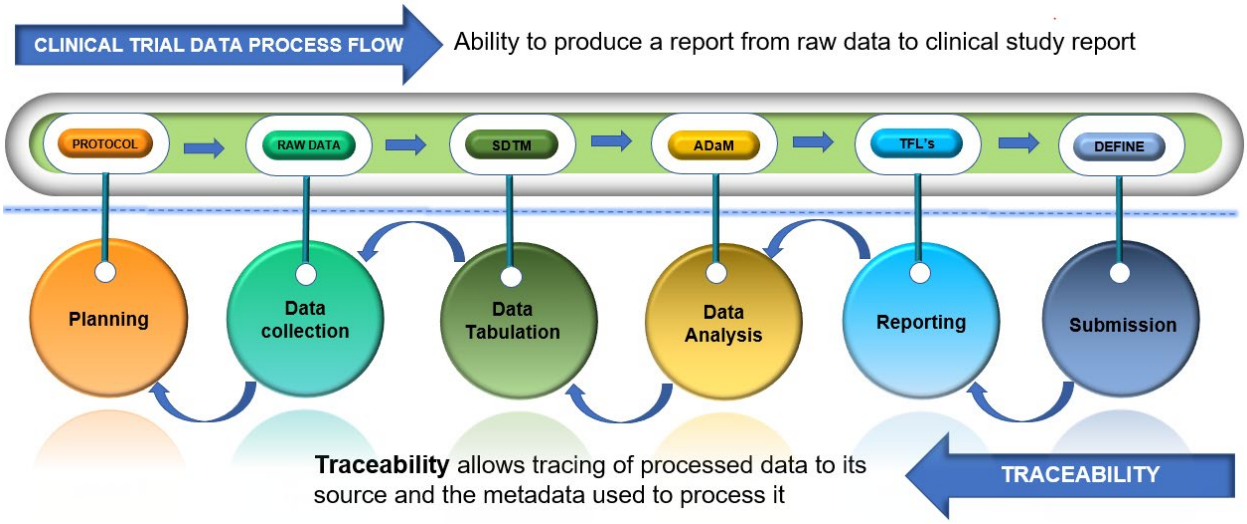


Figure 2. Traceability in the Clinical Trail Data Flow

SCOPE OF TRACEABILITY

TRACEABILITY CAN BE ACHIEVED ON DIFFERENT LEVELS:

- **Data Level Traceability**
  - Metadata Traceability
  - Datapoint Traceability

- **Document Level Traceability**
  - Tracking the Progress and Delivery
  - Process Documentation
  - File Level Traceability

Please see Table 1 below for few examples of data/document level traceability, which demonstrates how it can be achieved on different levels. Some documents may fall across different levels for example Protocol / SAP.

DATA LEVEL TRACEABILITY		DOCUMENT LEVEL TRACEABILITY		
METADATA	**DATAPOINT**	**TRACKING THE PROGRESS AND DELIVERY**	**PROCESS DOCUMENTATION**	**FILE LEVEL**
aCRF	Copy/retain SDTM variables	Data Transfer Plan (DTP) Trackers	Protocol	Program File Tracker
Data Transfer Plan (DTP)	SRCDOM/SRCVAR/SRCSEQ	Status Trackers	SAP	Change Control Tracker
Protocol	ASEQ	Decisions made by Statisticians/ Study team	SOP	
SAP	DTYPE	Database Migration	Trainings Documents	
cSDRG/ADRG	ANLxxFL	CRF annotation change log	Quality Control Checklists	
Lookup	Occurrence Flags in OCCDS	Raw to Targeted SDTM domain		
SDTM/ADaM specifications	TFL Shell Title and Footnotes	Study submission relevant information and compliance log		
Define.xml		Deliverable Approval Form		
Analysis Results Metadata (ARM)				

Table 1. Traceability Overview

DATA LEVEL TRACEABILITY

Traceability establishes across-dataset relationships as well as within-dataset relationships. There are two levels of data traceability.

METADATA TRACEABILITY

- Relationship between an analysis result (e.g., a p-value) and analysis dataset(s).
- Relationship of the analysis variable to the other variables within SDTM or ADaM source datasets.

Example 1: Analysis Result Metadata (ARM) Traceability

CDISC standardized the description of ARM for describing tables, listings, and figures. This references the data in standardized ADaM datasets, making it easier to re-use analysis results metadata across different studies.

Metadata traceability establishes traceability by describing the algorithm used to derive or populate an analysis value from its predecessor via metadata. Well defined and detailed programming specification document (Define.xml) is the means of building metadata traceability. Table 2 below is a sample of a programming specification document that enables the user to understand the relationship of an analysis variable to its source dataset(s) and variable(s).

Display	Table 14.X.X Summary of Overall Survival
Analysis Results	Survival Rate 95% CI by Kaplan-Meier
Analysis Parameter(s)	PARAMCD='OS'
Analysis Variables(s)	AVAL (Analysis Value)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	SECONDARY OUTCOME MEASURE
Data Reference (Incl. Selection Criteria)	ADTTE [AVAL ^=. and PARAMCD= 'OS' and SAFFL='Y']
Documentation	SAP section x.x
Programming Statements	(SAS version 9.2) proc lifetest data=adtte (where aval ^=. and paramcd= 'OS' and saffl='Y') method=km conftype=loglog alpha=0.05 timelist=12 24 36 reduceout outsurv= sci; time aval*cnsr (1); run;

Table 2. Analysis Results Metadata Sample Text

Example 2: SDTM Metadata Traceability

- Using aCRF, DDT, Define.xml, cSDRG.
- A variable with multiple origins (CRF, eDT, Derived, Assigned). Examples: Tumor responses, QNAM etc.
- A variable with multiple data types (Float, Integer, Text, Date, and Datetime). Examples: Lab results, Tumor Results etc.
- Algorithm/derivation of derived variable values. Example: PPD (Product of perpendicular Diameters).

DATAPOINT TRACEABILITY

Enables users (Agency reviewers, QC programmers, Biostatisticians, etc.) to go directly to the specific predecessor record(s) used to derive an analysis value. It can be established by providing clear links in the data to the specific data values used as an input from predecessor to derive an analysis value.

- Points directly to the specific predecessor record(s).
- The BDS and OCCDS structures are designed to enable datapoint traceability back to predecessor. Example: -SEQ variables.

Example 1: ADaM Datapoint Traceability

Below are few variables to establish data point traceability in ADaM:

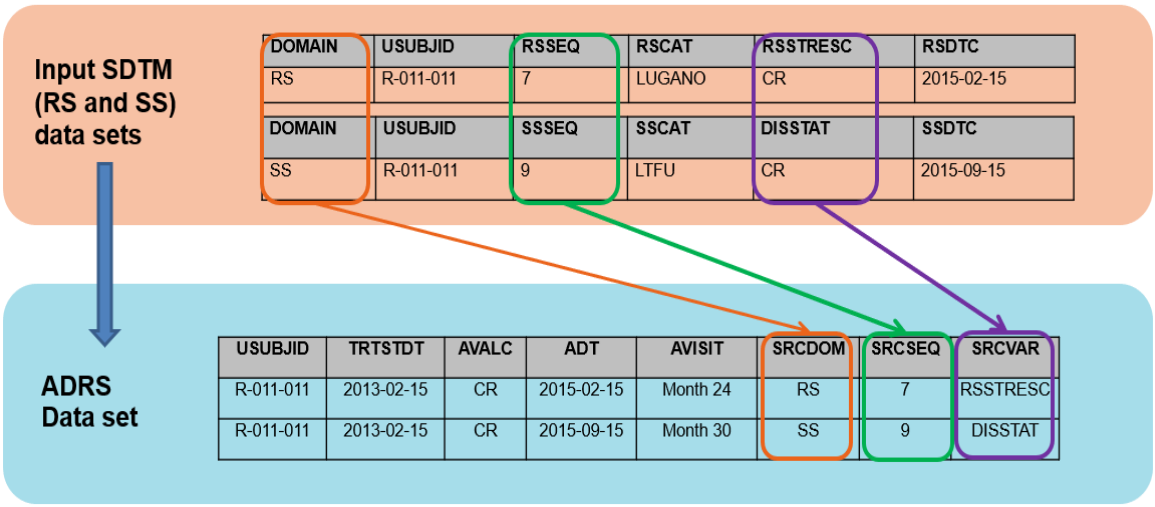
SRCDOM, SRCVAR AND SRCSEQ TRIPLET:

SDTM DOMAIN variable value, the name of the SDTM source variable, and the relevant SDTM domain -- SEQ value serves as primary candidates for data point traceability. ADaM implementation guide

recommends using SRCDOM, SRCVAR and SRCSEQ triplet along with derived analysis variable so that one can link back to the source SDTM records used to derive the analysis value. Table 3 below provides the derivation rules.

Variable	Source Data	Derivation
SRCDOM	RS SS	Comp: If PARAMCD = OVRLRESP and AVALC is selected from Disease Assessment then SRCDOM= RS; else if PARAMCD = OVRLRESP and AVALC is selected from LTFUP response then SRCDOM= SS
SRCVAR	DISSTAT RSSTRESC	Comp: If PARAMCD = OVRLRESP and AVALC is selected from Disease Assessment then SRCVAR = RSSTRESC; else if PARAMCD= OVRLRESP and AVALC is selected from LTFUP response then SRCVAR= DISSTAT.
SRCSEQ		Comp: If PARAMCD= OVRLRESP and AVALC is selected from Disease Assessment then SRCSEQ is equal to corresponding RS.RSSEQ; else if PARAMCD= OVRLRESP and AVALC is selected from LTFUP response then SRCSEQ is equal to corresponding SS.SSSEQ.

Table 3. SRCDOM, SRCVAR AND SRCSEQ Specification for ADRS



Display 1. SRCDOM, SRCVAR AND SRCSEQ Tracking

DOCUMENT LEVEL TRACEABILITY

The efficient conduct of clinical trials is largely dependent on the proper documentation of each clinical trial. Further, clinical trials results are created and stored in a way that makes it simple for an outside agency to revalidate it and provide an audit trail for any future investigations. Quality assurance monitors or regulatory agencies may audit or inspect essential documents to verify the accuracy of the study and the integrity of the data collected. Consequently, the investigator, sponsor, and monitor's compliance with GCP requirements (GCP) is demonstrated by all documentation.

According to ICH E6 (R2) Good Clinical Practice, essential documents are those documents which individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced. These documents serve to demonstrate the compliance of the investigator, sponsor and monitor with the standards of Good Clinical Practice and with all applicable regulatory requirements.

Essential documents are grouped in three sections according to the stage of the trial during which they will normally be generated:

- Before the clinical phase of the trial commences,
- During the clinical conduct of the trial, and
- After completion or termination of the trial.

TRACKING THE PROGRESS AND DELIVERY

The efficient conduct and management of clinical trials is aided by documentation that is accurate, brief, complete, timed, dated, and comprehensive. Tracking documents allow us to monitor status of a clinical trial and stay informed of the progress. Below are some examples of tracking the progress and delivery.

Example 1: SDTM Tracking:

With the use of various trackers, the data from CRF design to SDTM submission can be tracked and documented during study. Figure 3 illustrates a list of trackers like, Database migration updates, CRF annotation change log, Data Transfer Plan (DTP) trackers, Raw (CRF/external data) to targeted SDTM domains, Study submission relevant information and Compliance log to maintain and achieve traceability.

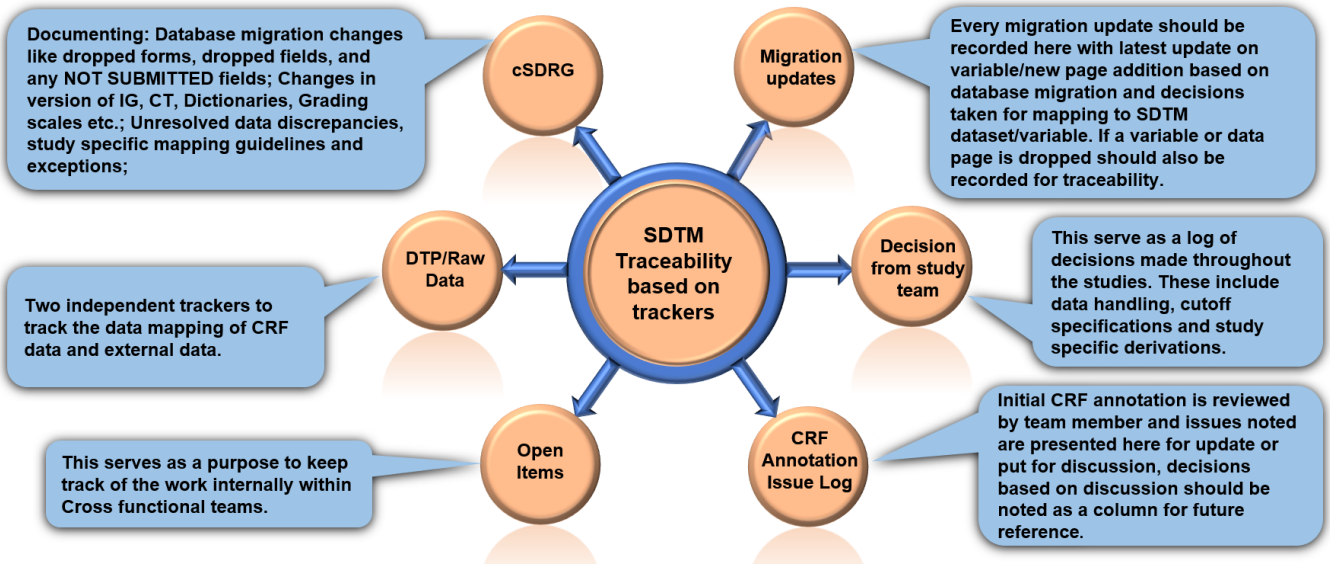


Figure 3. Example of Tracking progress and Delivery

The above example illustrates about SDTM trackers and ADaM team use similar trackers to help maintain document traceability.

Example 2: Ad-hoc request form:

During the life cycle of clinical trials adhoc requests are made to support the analysis from cross functional teams like biostatistics, patient safety, regulatory affairs to maintain compliance, and over time it becomes difficult to track these requests. Information gets lost in emails, a well-designed adhoc request form can help in answering the 5W1H (What, Why, Where, When, Who, and How) questions. In this usage example (Display 2), details about requestor information, timelines, purpose, list of studies to be included, detailed description of the request and any available past references are captured.

Ad Hoc Request Form

Section A (Requestor Information)			
Title			
Requester Details		Department	
Statistician Name			
Date Requested		Due Date	
Additional Information Requested		Due Date	
Purpose of Request			
Studies to be included (Please specify data location with cutoff date)			
General Information			
MedDRA Version		WHODrug Version	
Description			
Shell Location (If applicable)			
Statistical Procedure(s) (If applicable)			
Any previous reference			
Request Name			
Location			
Please Attach any Relevant File(s)			

Section B (Statistical programming information)	
Lead Programmer	
Support programmer(s)	
Programming location	
Comments (Programmer initial and date)	

Display 2. Ad Hoc request form

Example 3: Prep folder: A clone to eCTD (Electronic Common Technical Document) M5

Statistical programmers are responsible for preparing any Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) files required as part of the Module 5 [m5] Case Report Tabulations package for electronic submission (esubmission) to the FDA and/or PMDA, per the regulatory esubmission guidelines. The prep folder is used for preparation while Module5 (m5) is used for submission. New folders are added under the prep folder (please refer Figure 4). The primary objective of the additional folders in the prep folder is to improve submission traceability. However, the storage of files in the additional folders are not part of Module 5 eCTD. Additional folders are categorized based on the purpose (programs, documents, data sets) for any future reference. Please check the example below.

- Macroprog and sasprog folders contains sas programs with .sas extension.
- Docs folder contains editable copy of cSDRG, ADRG and aCRF, Control terminology document, and other important documents.
- Reports folder contains Pinnacle 21 reports for SDTM and ADaM.
- Sasdata folder contains sas data sets in sas7bdat format.

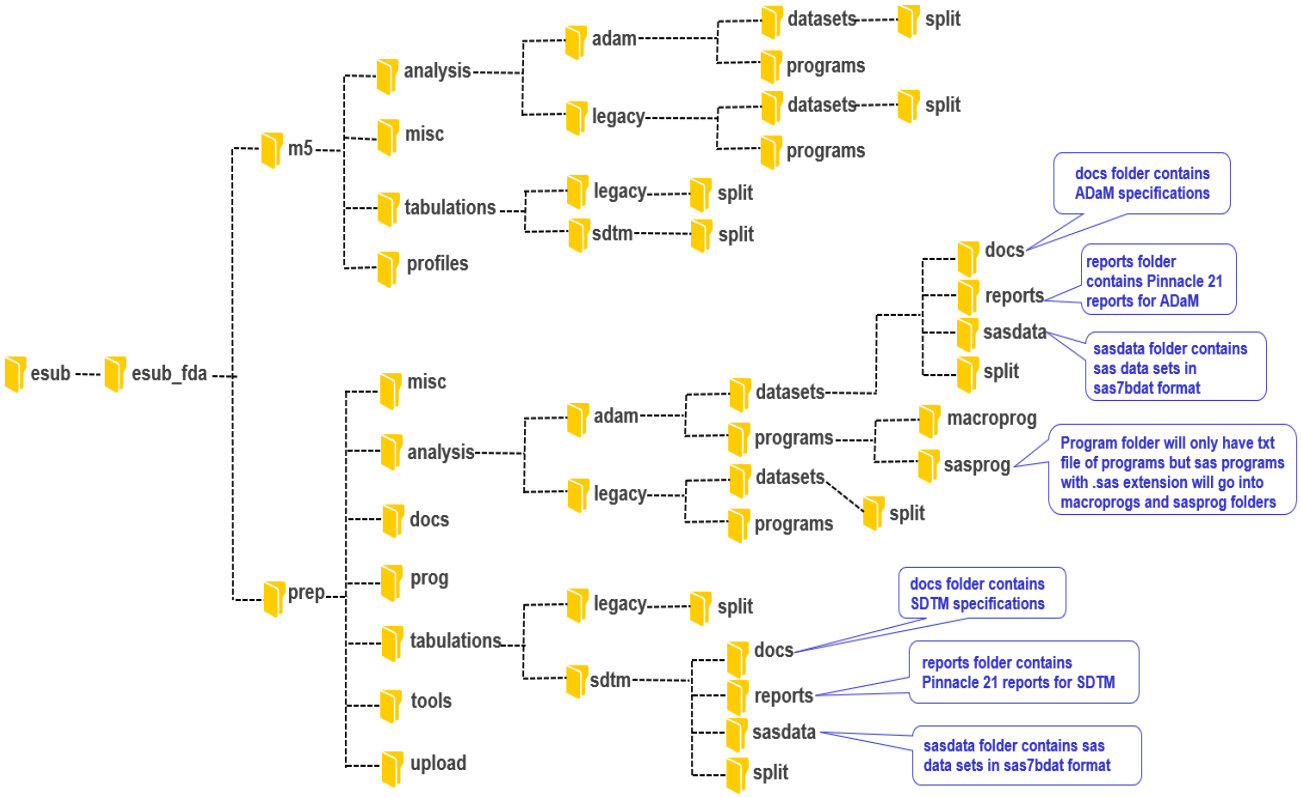


Figure 4. Prep folder: A clone to eCTD m5 folder

PROCESS DOCUMENTATION

Process planning documents play an important role, Protocols and SAP outline how a trial is conducted which include, steps followed, scope of the project and its timelines, data collection and analysis, resourcing, the process involved, submission-related guidelines, and instructions. Trainings, SOP's, and processes are well established which are assigned based on functional needs. Besides these, additional documents like output review checklist shown in Display 3 below, is an example of a quality control

Output Review Checklist

This checklist should be viewed purely as a tool to help biostatistics and statistical programming in the efforts to ensure high quality deliverables and does not replace any other quality control steps or processes that are already in place or those enforced in future

I. Format and Layout

Layout of the output matches specifications in the SAP and TFL shells		
1.	**Number of titles, footnotes, and columns are as per TFL shells**	☐
2.	**Proper population titles, by -titles are used**	☐
3.	**Footnotes are relevant to report contents** *e.g., Treatment emergency adverse event summary with inclusion definition*	☐
4.	**Correct title depending on the analysis is being displayed** *e.g., Title "Dummy Treatment Codes" should be displayed for the blinded outputs. Also check "Draft" vs. "Final" status*	☐
5.	**Spelling and clarity** *e.g., alignment, flow, breaks, accurate description of the data presented) of column headers, titles, and footnotes*	☐
6.	**No truncation has occurred**	☐
7.	**Proper sorting order is used**	☐
8.	**Ensure that name/location and the date stamp in the footnote is up to date** *e.g., for final run, date should be after database lock date*	☐

Display 3. Output review checklist

process document which is used by the study lead programmer to ensure all steps are followed, to maintain traceability. Other examples of review checklist are, pre snapshot activities checklist that can help lead programmers easily navigate between cross functional teams, say SDTM team and study team where activities need to be performed in sequential order. One more example is of look-up tables, which may be required pre or post ADaM data creation and input is needed from cross functional teams. In all these instances a quality control check list process document can be a powerful tool to keep track.

FILE LEVEL TRACEABILITY

Clinical trials take a large amount of time which can run over years and multiple analysis, it becomes imperative that a strong file level structure be in place, or the decisions made along the way can fall through the cracks as people move, so the question of what, where, when, why, who and how can be clearly answered.

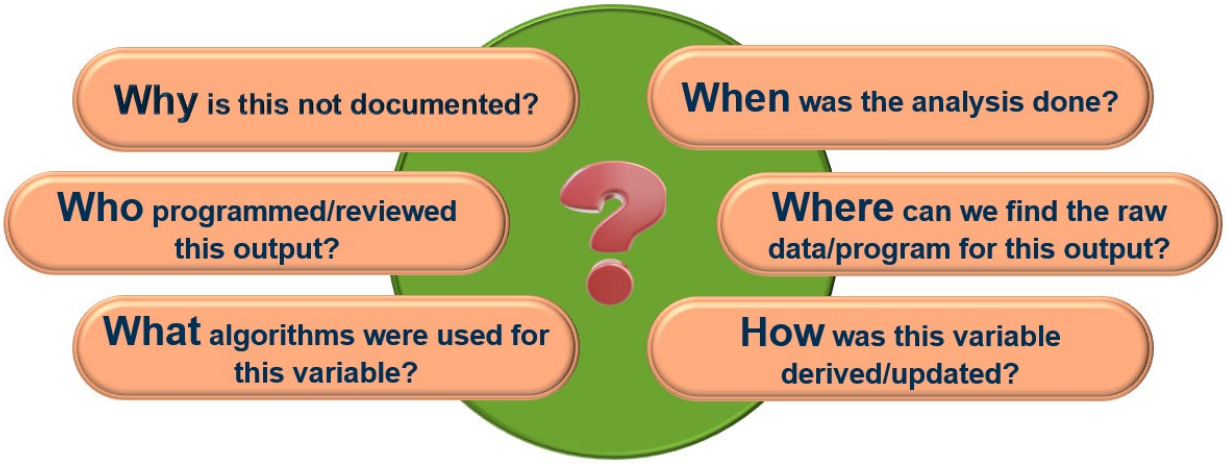


Figure 5. 5W-1H questions to track a file

Example 1: Program file tracker

A Program file tracker contains tabs like Cover Page, SDTM ADAM, and TFL. The Cover Page has analysis information (i.e., folder locations, cutoff/snapshots dates, programmers, statisticians, etc.). The SDTM ADaM tab tracks the progress of data set production. The TFL tab like the one shown below in Table 4 contains details of the attributes displayed in a TFL. Columns for these attributes are Category, Output No, Title, Population, Footnote, Source, cutoff date, Program Name and Output Name. Columns for tracking are Programmer, Production Date, Program Status, QC Programmer, QC Program Date and Review/Comment (filled by Statistician). Program status column provides the progress of the output being developed. Having the title and footnotes within a single document and calling them in the program is a better strategy than coding them in individual programs for tracking. Program headers provide useful information but get limited with the kind of information that is entered.

A new tracker is added to every analysis and can help with tracking within and across analysis, even when questions are raised after a few years.

Category	Output No	Title	Population	Footnote	Source	Cutoff date	Program Name	Output Name	Production Programmer	Production Date	QC Programmer	QC Program	QC Program Date	Program Status	Review/ Comments
Table	14.x.x	Demographics	safety	xxx	ADSL	01JAN2000	t_demog.sas	demog_saf	Rohit	10JAN2000	Kavitha	v_t_demog.sas	12JAN2000	100%	xxx

Table 4. Program file tracker

Using this document, study status can be generated and reviewed at any stage of the trial. Status of multiple ongoing studies can be reviewed if so desired.

Example 2: Change control tracker

A change control document is used to record changes made after a formal data snapshot, those impact the pre-defined algorithms, analysis, or the statistical outputs, which leads to modification to documents such as SAP, DDT, TFL shell, etc. The change control document contains tabs such as General, Raw data, captures information like outstanding raw data discrepancies. SDTM/ADaM tab captures domain level updates and TFL tab captures output level updates or group/global updates like an update, for all the Adverse Drug Reaction (ADR) tables. It is used to capture all the information from team communications and to reconstruct the output at any given time point. The Table 5 below shows an example of a change control document.

Dataset /TFL name	Variable Name	Change Request	Requested by	Date	Decision/ Resolution	Completed by	Date	Support document
ADTTE	EVNTDESC, CNSDTDSC	Update the derivation for subjects reported as lost to follow up prior to data cut off and after date last known to be alive	Statistician	7/1/2022	Updated	Programmer	7/05/2022	

Table 5. Change control document

CONCLUSION

In conclusion we can say, document traceability helps (to) create a clear path across the clinical trial flow including who completed what task, when and how. Document and process traceability helps to track the data to its source and to reconstruct the process at any given time point with the relevant information provided and (In) appropriate use of the data is easier to determine. Hence overall traceability should be envisioned as not just data traceability, but a combination of data, document, and process flow. Further traceability leads to efficient usage of resources and reduced operational cost, improved quality, higher levels of data reliability and integrity, better regulatory compliance, and less follow-up questions from regulatory agencies.

REFERENCES

CDISC. 2021. “Analysis Data Model Implementation Guide v1.3” https://www.cdisc.org/system/files/members/standard/foundational/ADaMIG_v1.3.pdf

CDISC. 2022. “Analysis Data Model (ADaM) Examples of Traceability v1.0” https://www.cdisc.org/system/files/members/standard/foundational/ADaM_Traceability_Examples_v1.0.pdf

CDISC. 2015. “Analysis Result Metadata Specification v1.0 for Define-XML v2.0” https://www.cdisc.org/standards/foundational/define-xml/analysis-results-metadata-arm-v1-0-define-xml-v2-0

ICH HARMONISED GUIDELINE. 2016 “Integrated Addendum to ICH E6 (R1): Guideline for Good Clinical Practice E6 (R2)” https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf

U.S. Department of Health and Human Services Food and Drug Administration Center for Biologics Evaluation and Research (CBER) Center for Drug Evaluation and Research (CDER). “STUDY DATA TECHNICAL CONFORMANCE GUIDE”, March 2023 https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document

ACKNOWLEDGMENTS

The authors would like to acknowledge colleagues who generously provided their support, insight, and perspective to improve this paper in particular Abhishake Singh, Vamsi Kandimalla and Murali Kanakenahalli.

CONTACT INFORMATION

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

Rohit Nagpal
Kite Pharma, Inc.
rnagpal@kitepharma.com

Kavitha Mullela
Experis Solutions
kmullela1@kitepharma.com