

Traceability and Data Flow

Current State Analysis

-August 21, 2013-

Background:

At the 2012 FDA/PhUSE Computational Science Symposium, a working group was set up to address “challenges of integrating and converting data across studies.” As part of this working group, a smaller sub-team was set up to examine the issue of traceability on integrated data and prepare a white paper on the subject.

The Traceability sub-team recognized that the issue of traceability on integrated data was highly dependent upon the data flow at the individual study level. As a result, a new project was kicked off at the 2013 FDA/PhUSE Computational Science Symposium called “Traceability and Data Flow”. This project falls under the “Optimizing the Use of Data Standards” working group and its high level scope is to define traceability considerations and best practices for study level and integrated dataset conversion for a variety of different data flow scenarios.

Traceability Definition:

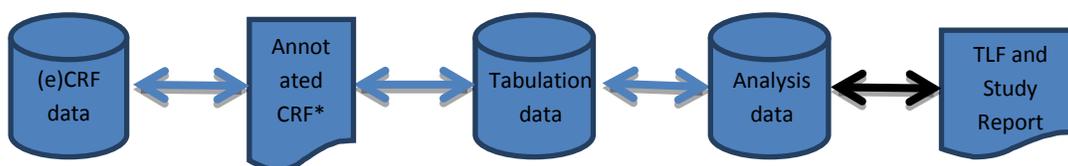
At the time this current state analysis was drafted, Wikipedia contained the following definition of traceability: “Traceability is the ability to verify the history, location, or application of an item by means of documented recorded identification.”

Information about the traceability expectations or requirements for electronic clinical datasets can be found in many places: FDA documents, CDISC documents, proceedings from industry meetings, conferences and webinars sponsored by associations like the DIA, PhUSE, SAS Global forum, etc. The information exists, but the level of traceability across electronic submissions of clinical data is still inconsistent.

Problem Statement:

Traceability is part of the process; it is not just for submission. Traceability provides documentation about the study that is useful along the way. It includes metadata for business rules, not just variables and datasets, to determine if data can be integrated. Decisions made to convert to standards need to include discussions around traceability and risks for each option. Communication, within the sponsor company and between sponsor and FDA reviewer, is imperative. The outcomes of the Traceability and Data Flow project are intended to help facilitate those discussions.

At the study level, traceability from one data processing step to another is more straightforward when the data flow follows a linear path. For example, in the scenario below, data moves linearly from one step to the next.



* Tabulation dataset annotations

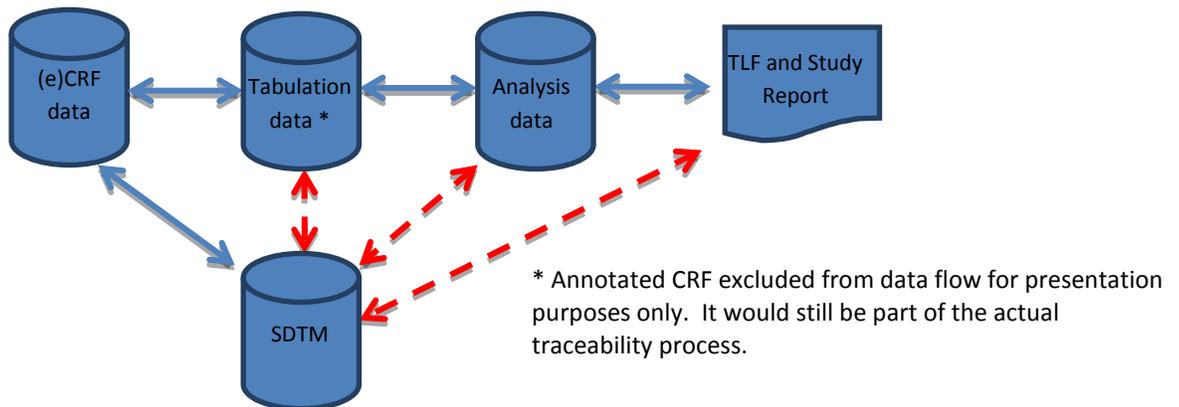
Traceability and Data Flow

Current State Analysis

-August 21, 2013-

In this scenario, it does not matter if the data standards used are legacy or CDISC. If the tabulation data are based on SDTM and the analysis data are based on ADaM, the data flow is still linear and traceability is more inherent in the process.

Traceability issues arise when there is a need to convert legacy data after the fact. If SDTM datasets are requested as part of an electronic submission, but the data used to generate analyses and the study report are based on legacy standards, then post process creation of the SDTM based datasets can result in traceability gaps. The visual below shows some of the traceability paths along this non-linear process. Blue arrows indicate traceability components that are inherent in the process. The additional step of creating SDTM data from (e)CRF data is part of the legacy conversion process. Red arrows indicate where traceability is not part of the process but might be useful or even necessary during review.



From an FDA perspective, reviewer needs vary, especially between medical and statistical reviewers. In the scenario displayed above, the medical reviewer may depend heavily on the SDTM datasets in order to use available review tools. The statistical reviewer may start with the analysis datasets and need to trace back to the eCRF data. If only the SDTM data are delivered as part of the submission, and not the legacy tabulation datasets, the traceability chain may be broken. When a traceability chain is broken, reviewers may not be able to replicate results, and this could lead to questioning the validity of the data. The CDER Data Standards Common Issues Document (version 1.1/December 2011) addresses this scenario:

It is very important that the results presented in the Clinical Study Report be traceable back to the original data elements as they were collected in the case report form and represented in the SDTM datasets. The SDTM datasets must be able to support the results in the Clinical Study Report, either directly for some results, or, for other results, indirectly through analysis datasets that are derivable from the SDTM datasets.

If a sponsor decides to convert trial data to SDTM that was originally collected in non-SDTM format, it is important to note that the resulting SDTM data should support the accompanying analysis data sets and

Traceability and Data Flow

Current State Analysis

-August 21, 2013-

sponsor's reports (study reports, etc.). CDER has received applications in which the converted SDTM data sets were not consistent with the submitted analysis datasets and study report analyses, thus causing confusion during application review.

Traceability and Data Flow Project Scope:

In an effort to ensure consistency in submission of electronic datasets and level of traceability documented as part of the process, the Traceability and Data Flow project is planning to produce the following deliverables:

1. Review of traceability references (FDA, CDISC, other industry organizations and documents) in order to consolidate pertinent information and make recommendations on usage.
2. Define 2.0 document review in order to identify solutions for how traceability information is documented.
3. Basic linear data flow white paper, regardless of use of CDISC standards, to identify fundamental best practices for traceability.
4. White paper: study level traceability for FDA needs
5. White paper: integration traceability for FDA needs

The following points of view will be taken into consideration throughout the deliverables.

- We need traceability at each point along the way, regardless of legacy/proprietary or CDISC standards.
- Traceability needs to be bi-directional at each point (if you can get from A to B, you can also get from B to A).
- Standards, CDISC or other, are not the focus of this paper. Some traceability issues might be easier to solve when working with full-CDISC standard data, but moving to standard data does not eliminate the need for traceability.
- Linear process, legacy and full CDISC will be taken into consideration along with transitions to CDISC
- If SDTM datasets are to be submitted, traceability of labeling pursued efficacy and safety variables between SDTM datasets and analysis datasets should be considered as a necessary part of the drug development process.

Audience:

- Agency reviewers (statistical and medical)
- Sponsor regulatory representatives
- Sponsor submission creators (usually data managers, programmers, statisticians, plus contractors and vendors in these areas)
- Technology providers