PP06: Legacy Data Meets CDISC Standards
Experiences from Working with Evolving Standards in Submission Projects

Dirk Spruck, Beate Hientzsch, Accovion, Germany

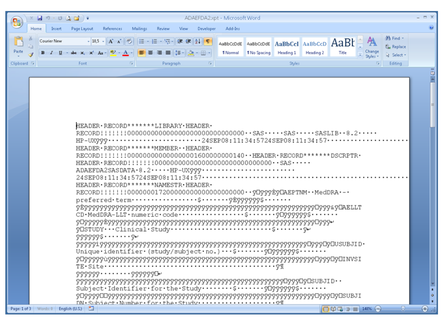
CDISC standards have become an integral part of the life science industry; nevertheless, we will have to continue to deal with data in different legacy formats for some time in the future. While the use of purely CDISC-formatted data from the very beginning of a submission project is unproblematic, combining data in legacy format with CDISC standardized data presents considerable challenges and therefore requires careful planning and special attention.

Example 1: „The files you sent us are kaput“

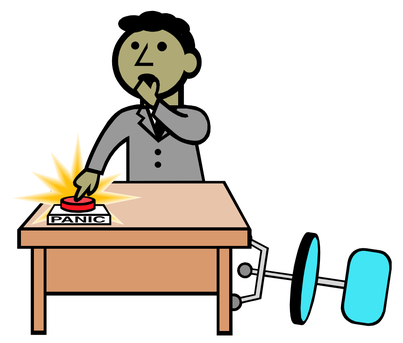
More than once we have explained upset callers that XPT files cannot be opened in WORD.

This may seem funny but illustrates a challenge we are constantly facing:

Outside the clinical data science world there is very little understanding of what needs to be done with clinical data for a regulatory (FDA) submission.



Regulatory Affairs departments are hesitant to approach the FDA outside the mandatory milestones. But with legacy data it is important to contact the agencies with a sound data concept early to leave enough time for data preparation.



The pre-NDA / BLA meetings are usually too late for this discussion and should focus on important science aspects rather than data structures.

Requests for the “full CDISC package” with a clean CDISC-Validator log often lead to un-necessary effort.

Example 2: Analysis Datasets ≠ XPT Compliant

Original Legacy Data

aCRF Raw Data in Format

⇨ pregnancy test: 9 not done ⇨ exclude patient (DEMOG → pregt_n)

1 yes ⇨ 0 neg.

1 pos. ⇨ exclude patient (DEMOG → pregr_n)

Analysis Data Preparation

```
select (pregt_n);
          when (9)  pregnancy_test="9=not done";
          when (1)  pregnancy_test="1=&yes.";
          otherwise pregnancy_test="n.a.";
      end;

      select (pregr_n);
          when (0)  pregnancy_test_result="neg.";
          when (1)  pregnancy_test_result="pos.";
          otherwise pregnancy_test_result="n.a.";
      end;
```

Starting Position

Data from multiple studies was analyzed using legacy formats.

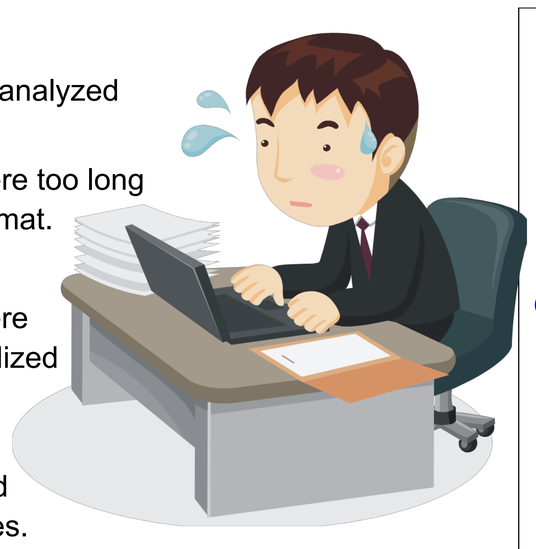
Dataset and variable names were too long for direct conversion to XPT format.

Possible Solution

Dataset and variable names were carefully renamed and standardized across all studies.

Programs were generated and submitted to map data back and forth between the data structures.

Old and new names were documented in the DEFINE document.



Submission Package

aCRF Analysis Data — Analysis dataset: ADEMO

⇨ pregnancy test: 0 not done (PREGTN) ⇨ exclude patient (PREGT)

1 yes ⇨ 0 neg. (PREGRN) ⇨ exclude patient

1 pos. (PREGR) ⇨ exclude patient

DEFINE.PDF Analysis Data

Variable	Label	Type	Format/Codes	Comments
PREGTN	Pregnancy test performed code	Num	0: no 1: yes	Demographic data Page 5 (Original variable name: PREGT_N)
PREGT	Pregnancy test performed	Char		Demographic data Page 5 (Original variable name: PREGNANCY_TEST)
PREGRN	Pregnancy test result code	Num	0: neg. 1: pos.	Demographic data Page 5 (Original variable name: PREGR_N)
PREGR	Pregnancy test result neg/pos	Char		Demographic data Page 5 (Original variable name: PREGNANCY_TEST_RESULT)

+ Data Preparation & Analysis Programs

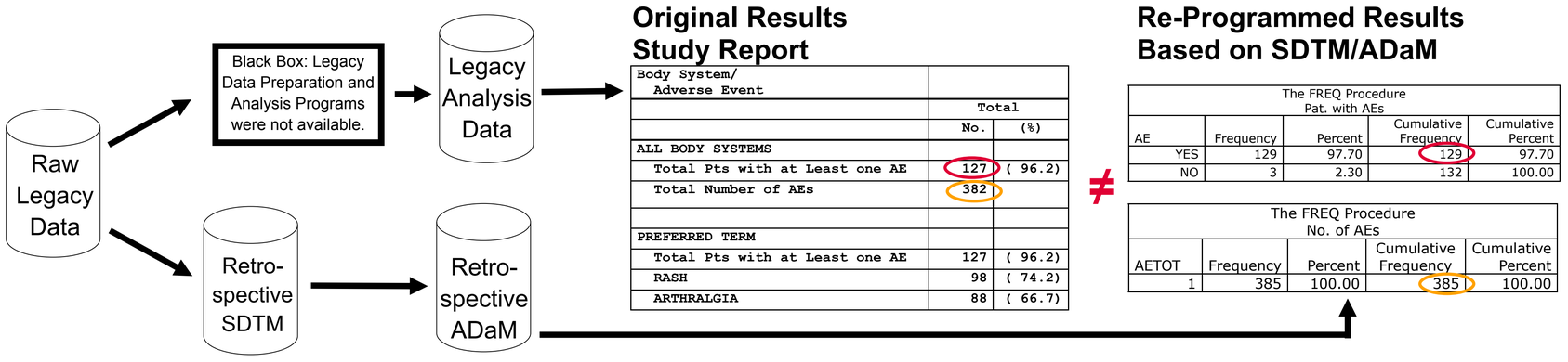
Example 3: Comparing Original Results against Mapped Data

Project Outline

For this project only legacy raw data, legacy analysis data and original analysis results were available. Data preparation, analysis programs, and data definition documentation were missing.

The customer requested re-mapping of the legacy raw data to SDTM followed by the creation of CDISC compliant ADaM datasets.

As a final QC step analysis results were recreated based on ADaM datasets and compared to original analysis results.



QC Result

Of course discrepancies between the original and the re-programmed analysis emerged. Because of the lack of additional information on the original analysis the resolution and documentation of findings was **extremely time consuming**.

Potential Sticking Points:

- Differences in the interpretation of the Analysis Plan.
- CDISC standard definitions differ from implementation in original programs.
- Programming error in the original programs.
- Programming error in the SDTM/ADaM mapping and re-analysis.

Example 4: Documentation

The picture shows the **eCTD** standard folder structure.

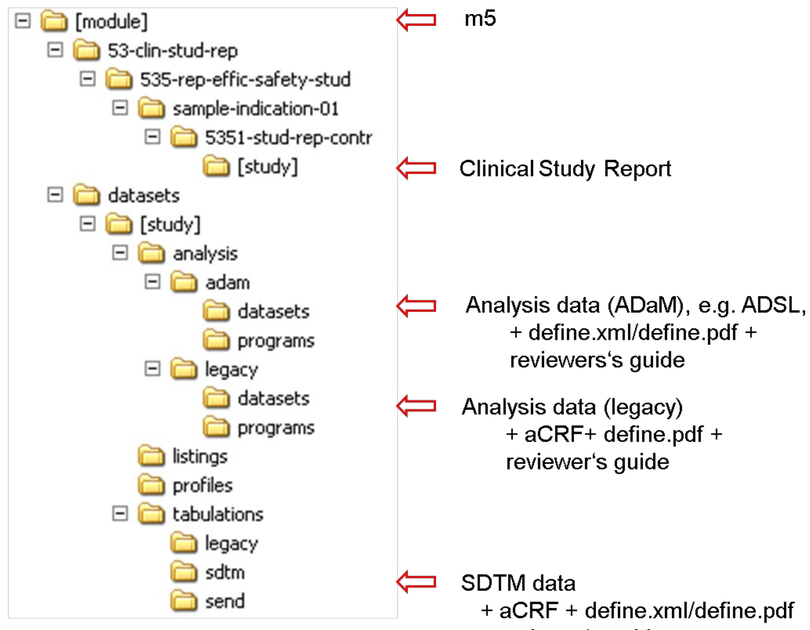
Empty folders should be removed from the folder tree.

Sufficient documentation is key for reviewers to understand where the data comes from and how it was processed.

Annotated CRFs and DEFINE documents are needed not only for SDTM data but also for legacy data.

Do not overload single documents. If more information is needed to understand certain aspects of the data, e.g. derivation of the key efficacy parameter, provide documents in addition to the reviewer’s guide and the define document.

KISS — **k**eep **i**t **s**hort and **s**imple and easy to understand.



Every submission project is unique and needs careful planning to avoid costly delays

- Preparing legacy data for submission must be done carefully to retain the scientific integrity of the data.
- Develop a sound plan how to deal with legacy data and get approval from the agencies ahead of time.
- During the preparation of legacy data for submission it is very likely that discrepancies will be found. Plan enough time and resources for investigation and resolution.
- Perform additional validation steps against the original study results based on a risk assessment if legacy data is retrospectively mapped.
- Include sufficient documentation with your submission data to make the preparation process, findings, discrepancies, and special cases transparent.
- Moving away from XPT files and further regulatory guidance will hopefully simplify data preparation for electronic submissions.

Prepare a sound data concept and convince Regulatory Affairs to submit it early to regulatory authorities for approval.

Accovion

Dirk Spruck
Softwarecenter 3
35037 Marburg
Germany

dirk.spruck@accovion.com

Beate Hientzsch
Helfmann-Park 10
65760 Eschborn
Germany

beate.hientzsch@accovion.com

References

FDA email addresses for data format questions:
cder-edata@fda.hhs.gov, cber.cdisc@fda.hhs.gov

FDA Guidance documents:
http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm