PhUSE 2013

Paper CD02

ADaM Standards – Organizing the Unorganized

Heike Reichert, Bayer Pharma AG, Wuppertal, Germany

ABSTRACT

Analysis datasets according to ADaM requirements still have a big scope for individual interpretation, implementation and variations across programmers and studies. Therefore standardization is of high importance for both, sponsors and agencies, in order to reduce the creativity and make the development and use of tools possible. This means a lot of effort for the pharmaceutical companies in redefining their processes and structures. But it also offers the opportunity to develop generally applicable standards and tools that can make life easier in the long run. The agencies are able to shorten their review time as the re-evaluation of the study results is facilitated.

INTRODUCTION

The expenses for developing new drugs are very high. Therefore pharmaceutical companies do not concentrate on only one single market but on multiple markets throughout the world. To get the marketing authorization in the US all drugs must be FDA (U.S. Food and Drug Administration) approved.

The FDA's "safety team monitors the effects of more than 3,000 prescription drugs on 200 million people" (1) from various pharmaceutical companies. This means a great deal of work especially as each company has its own submission standards.

Therefore authorities request submissions of standardized study data in an electronic format.

WHY STANDARDS?

Imagine your fridge looks like this:

This fridge is particularly disorganized. It is topsy-turvy.

There are not any identifiable preferred locations for different kinds of food, e.g. eggs and milk go into the door tray.

It is not even clear which food goes into the fridge at all. Do all fruits need a cool depository?

Moreover it is really tough to retrieve food. And it is almost impossible to locate expired food.

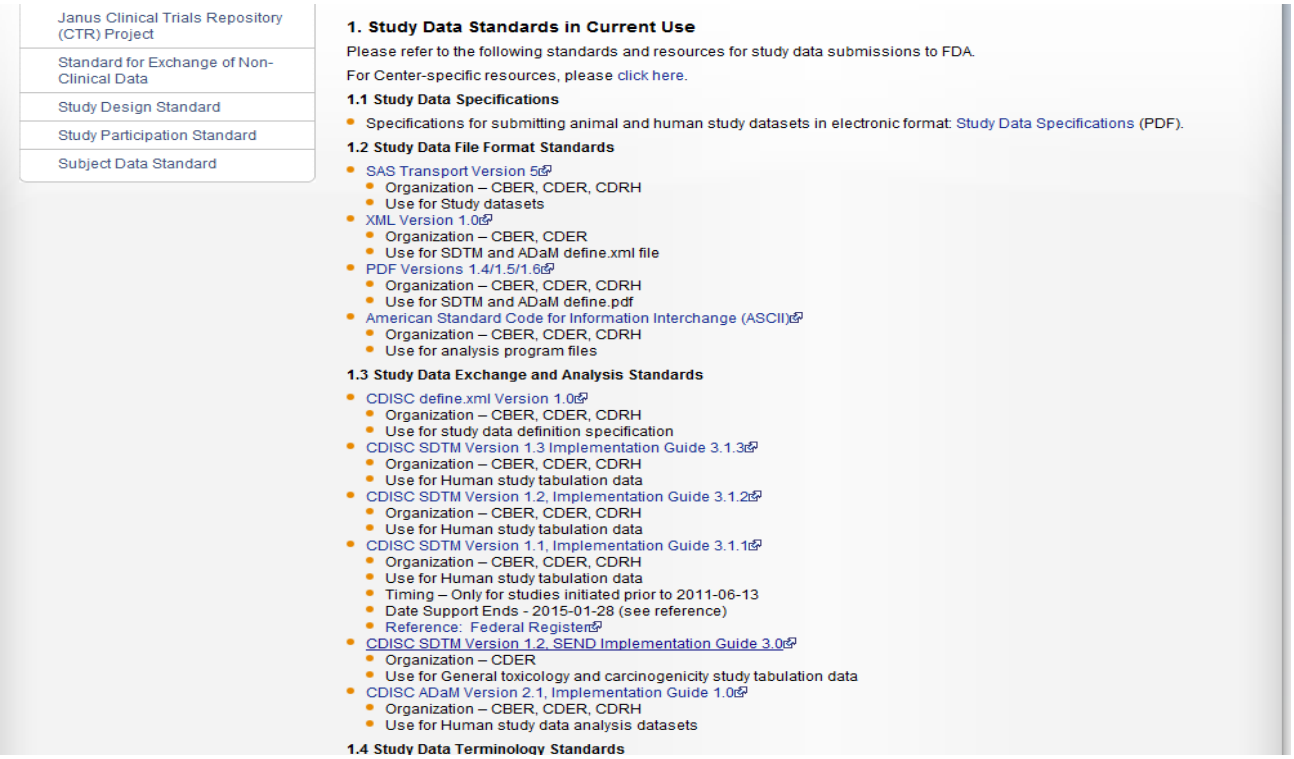
Conversely, please look at this fridge:

It is neat and well-arranged. All food has its well-defined storage place, so that it is quite easy to retrieve the desired food or to locate expired food for removal.

Of course the fridges are metaphors for non-standardized and standardized datasets. The food stands for data points of a study. Both, fridges and datasets are aimed to be well organized in order to retrieve, remove or store something new.
The bottom line is that standardization helps to specify unambiguous metadata. It is clearly defined which data goes to which dataset so that the retrieval of data and implementation of updates are made easier. Retired data can be identified quickly.

Standards at the FDA
Since 2007, FDA has collaborated with the Clinical Data Interchange Standards Consortium (CDISC) on the development of study data exchange standards. “CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website. “(2)
Amongst others CDISC provides standards for the study planning, data collection (e.g. CDASH), data tabulations (e.g. SDTM) and statistical analysis (ADaM) (3).

It is necessary to check the agency’s website (4) regularly to make sure that the appropriate standard supported by the FDA is used for submission. The supported standards may change over time. The FDA does support old versions or standards no longer than 2 years after a new version or new standard is available.



Furthermore the FDA developed a “Guidance for Industry Providing Regulatory Submissions in Electronic Format (DRAFT)” (5). This document describes the current and future requirements how data has to be provided to the agency. The FDA expects that submitted data fits validation rules, e.g. sponsors should apply the openCDISC validator (6). This tool is by the way a common validation checker also used by the FDA. Standards should be taken into account already at the study set-up to avoid a retrospective mapping. Retrospective mapping is error-prone and means additional workload. The FDA recommends to communicate and to discuss the planning of standard implementation for submission with the review division of the agency as early as possible, preferably even before the pre-IND meeting (eStudy Guidance and PDUFA, Amy Malla, MT(ASCP), CQA, PMP, CBER/Office of Review Management (7)).
As soon as this guidance is published - that will be 12 month after the public comment period – submissions for Biologics License Application (BLA), New Drug Application (NDA), supplements and amendments must comply within 24 months, for Investigational New Drug (IND) and some amendments (see section 561 of FD&C Act) (8) within 36 months.

Standards at the EMA

The EMA believes that a publication of clinical data establishes trust and confidence in the system. Therefore the agency published a draft policy “Publication and access to clinical-trial data” (9). This policy is the result of the “Clinical Trial Data Transparency Initiative” that aims to proactively publish clinical-trial data for interested parties.
In a preceding workshop (10) with participants of Academia, Media, Industry, Patients, a European Ombudsman’s representative and the assistant of European Data Protection Supervisor the interest, views and concerns were collected and evaluated.
The policy demands raw de-identified clinical trial data in an analyzable format. In future CDISC shall be the required standard.
Public comments could have been provided until 30 September 2013. It is expected that the policy will come into force on 1 January 2014. Even though the pharmaceutical industry should be prepared to deliver raw clinical trial data, the implementation of the policy is impacted by the outcome of some court cases related to the “European Medicines Agency policy on access to Documents” of 2010. Legal action was initiated in order to clarify the legal situation for the publication of commercially confidential information. In the ongoing legislative process to replace the current European directive on clinical trials the European Parliament’s Committee on Environment, Public Health and Food Safety stated that “in general the data included in clinical-trial study reports should not be considered commercially confidential once a marketing authorisation has been granted or the decision-making process on an application for marketing authorisation has been completed” (11).

ANALYSIS DATA STANDARDS AT BAYER

Analysis datasets at Bayer are specified according to CDISC ADaM standard. Starting point for the generation of the analysis datasets are the so-called SDTM+ datasets which contain, besides SDTM defined variables, also Bayer specific variables, e.g. the date parts day, month and year. For traceability reasons certain SDTM+ variables are taken over to the analysis datasets.

A characteristic of an SDTM domain are the character variables that are not very preferred by programmers. On the one hand selection via text strings is error-prone, on the other hand variability among different programmers especially when handling text string makes standardization delicate. Therefore it was decided to introduce a Code-Decode principle for most character variables. Mandatory for the coded variables are codelist/formats which are referenced by them. The coded variables serve as code, the decode variables serve as label. With this method the use of character strings are controlled via a "controlled terminology" beyond CDISC requirements. A study can only use codes and decodes that are available in a pre-defined codelist. If required codes are missing these must be officially requested. An expert team evaluates each request and approves or rejects the requested codes.

Bayer has a department that exclusively deals with pooling of clinical trials. This department decided to use analysis datasets for the data pools. A reproduction of the tables, listings and figures that are provided for each single study is facilitated and an integrated analysis can be provided within a short period of time. It is extremely important for the pooling that defined standards are met. Otherwise an integration is challenging and can only be done with additional workload.

STANDARDIZATION OF ANALYSIS DATASETS AT BAYER

The standardization at Bayer comprises both, dataset metadata and codelists. Bayer has a detailed system of different hierarchical levels.

The highest level is the global level. All lower levels must adhere to the analysis dataset and codelist specifications of the global level. At the moment about 30, mainly safety analysis datasets (e.g. ADAE- adverse events, ADVS – vital signs, ADLB – laboratory data) are defined on global level. But also efficacy datasets like ADTTE (Time-to-Event) are available.

Below global is the Therapeutic Area Standard (TAS) level, e.g. Oncology. In contrast to global level TAS specifies mainly efficacy datasets like ADMB (Microbiology) or ADTU (Tumor). Additional variable specifications of datasets defined on global level are possible. For example in Bayer's global ADAE the toxicity grades are not specified, but TAS Oncology contains these additional variables.

Within a TAS several project standards can be maintained. The project standard defines all datasets and variables that are used in studies within the project. Variables from global and TAS level as well as project specific new variables are part of the project level.

Programmers of a study have to use the metadata and codelists that are available on the appropriate project level. Only few adaptions are possible. These so-called permissible changes are well-defined, e.g. the method for date imputations can be modified in order to fulfill the requirements described in the SAP. Study specific definitions of datasets or codelists are not allowed.

ADaM rules leave a big scope for interpretation and it is necessary to align the handling among programmers, among different projects and among different studies.

This standard model limits the creativity for each programmer.

CREATION OF ANALYSIS DATASETS

The generation of analysis datasets is supported by a complex macro system which was developed and is still enhanced in-house. It is tailored to particular needs of the programmers to generate analysis-ready, ADaM conform analysis datasets.

Prerequisite for this macro system are pre-defined metadata as SAS datasets that are used as input-information for an automated process. In this metadata variables are defined that should be present in the appropriate analysis dataset with their origin and derivation methods.

In particular the macro system takes care of the creation of an analysis format catalog and the transfer of selected SDTM+ variables to the analysis datasets with their attributes by reading the information of the SAS metadata. If a derivation algorithm is fix, e.g. CHG=AVAL-BASE, this variable is automatically derived. Additionally variable attributes like label, format and length are defined according to the metadata which means a great reduction of workload as the programmers do not need to take care of all these ATTRIB-statements. Decode variables are automatically generated as well. And last but not least it checks the generated data against the metadata for completeness and compliance to rules.

On the one hand this macro system supports the generation of ADaM compliant analysis datasets with limited possibilities for programmers for own interpretations and data handling. On the other hand a prerequisite for the development of this system is a wide standardization of analysis datasets and SDTM+ datasets beforehand.

ADVANTAGES OF A STANDARDIZED ADAM SUBMISSION

The most outstanding advantage of a standardized ADaM submission is certainly the similarity of the data structure and terminology. The unity is not limited to one submission of a single sponsor, but applies as far as possible also to different sponsors. This again results in a facilitated review for authorities as they can develop their own standardized tools for data review and for re-evaluation of study results. And this ends hopefully in a shortened review time.

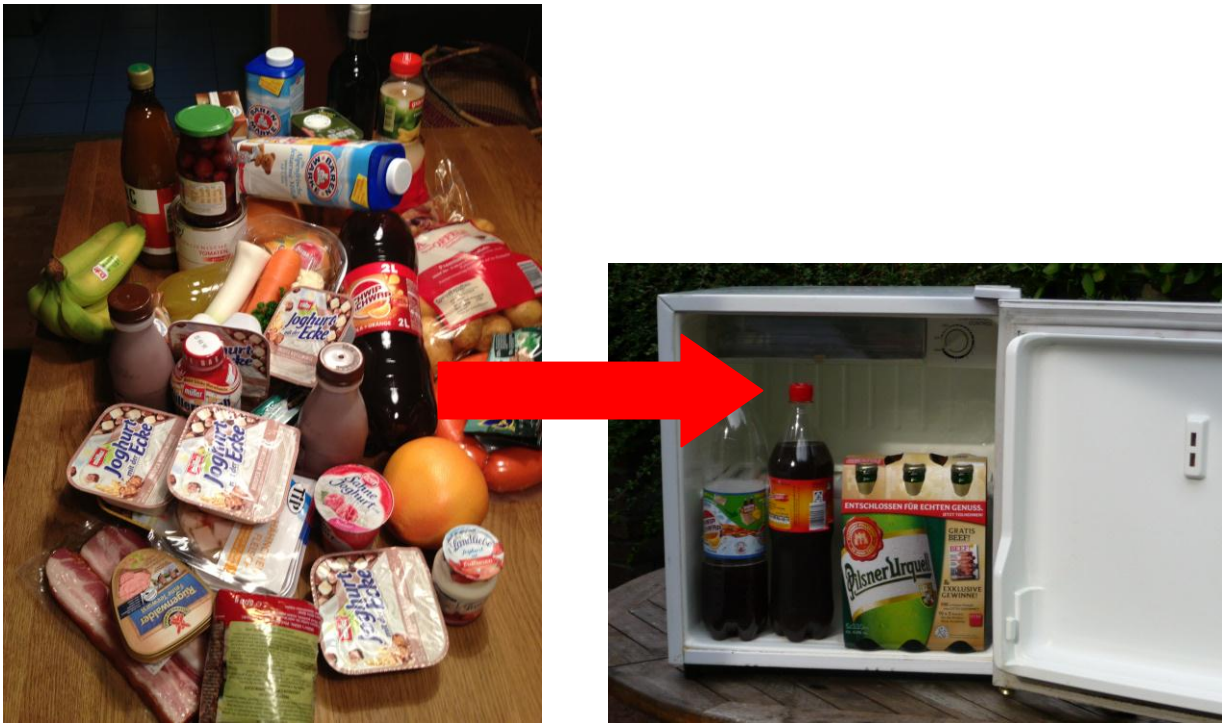
Similar advantages are not to be underestimated on the sponsor side. Standardized programs and tools for the generation of datasets, documentation and deliverables can be developed so that time can be saved for the deliverables in the long run.

CHALLENGES OF AN ADAM SUBMISSION

It is a tough job to implement all ADaM requirements in a standardized manner.

And coming back to the fridge model, it sometimes felt like:

How shall all this stuff go into this small fridge?



But seriously, there are a lot of challenging obstacles that have to be overcome.

To name just a few from the experiences with the ADaM standardization:

- It is very time consuming to prepare documentations for all types of metadata. Without going into detail here, some kind of metadata can be retrieved easily in the metadata, others require a lot of manual work (e.g. analysis result metadata).
- The programmers have to understand the principles of CDISC. Sometimes it is hard to understand why additional records should be added instead of additional variables. And the harmonization principle – same name, same value, same label – has to be considered in any case! Otherwise the required traceability gets lost.

- There is also a learning curve for the authorities. For one of our submissions with efficacy datasets according to ADaM standards, the authorities requested additional columns for LOCF and WOCF derivations, which is not according to ADaM rules. ADaM asks for additional rows marked as DTYPE=”LOCF” or DTYPE=WOCF”.
- Additionally changes in specifications made by CDISC have a high impact not only on the metadata and possible structures of dataset, but also on programs and on the time required until all deliverables are provided.

CDISC STANDARD ISSUES BY THE FDA

In a Webinar hosted by CDISC last year (12) the FDA informed about the main issues occurred with standardized submissions.

- One big issue, the FDA had, were the huge dataset sizes. They found out that there is a correlation between the dataset size and the allocated variable length. The allocated length of some variables was 200, even though the actual length is just 8. The column lengths should not be set to an arbitrary limit of 200.
- Another issue the FDA pointed out was the creation of sponsor defined domains and variables. SUPP- domains contained unnecessary information and the FDA asked to discuss on sponsor’s side whether all information in SUPP-datasets or in additional domains is really necessary.
- As third the FDA complains that submitted datasets still contain validation errors. The agency asks to use the OpenCDISC validator. Errors and warnings that can be fixed, should be fixed and not just simply addressed in the “Reviewer’s Guide”.
- For extended codelists/controlled terminology the FDA observed in many submissions that not all used values of a variable were added to the related codelist.
- Invalid ISO 8601 values caused some problems, especially with partial dates and times.
- Traceability between source data and dataset must be given. The Linkage has to be: CRF → SDTM → ADaM → CSR
- Last but not least the FDA figured out that inadequate documentation was provided. The agency asks for supporting documentation in the form of a “Reviewer’s Guide” to explain how the data standard was implemented, e.g. content of custom domains, unfixable errors and warnings, derivation of key analysis variables.

Standardization and development of standard tools help to prevent these errors as the individual handling of data and stand-alone decision-making of programmers are limited. A consistent approach is ensured.

To emphasize it here again: The FDA will reject submissions if submitted data does not adhere to standards as described in the “Guidance for Industry Providing Regulatory Submissions in Electronic Format”.

CONCLUSION

The FDA and other agencies around the world expect high quality data, metadata and documentation in a standardized format. Industry has to reconsider their submission deliveries. If the worst comes to the worst it is necessary to discard all processes and start from scratch.

And to use again the fridge model:

It may be necessary to bring the old fridge to the waste container!

The outer shell is provided by CDISC. Now the empty fridges have to be filled with content in order to make them look really nice for the agencies!

REFERENCES

(1) Center for Drug Evaluation and Research - Wikipedia, the free encyclopedia
(2) Mission Statement and Core Principles
(3) Standards & Implementations
(4) http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm
(5) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf
(6) OpenCDISC | An Open Source Community of CDISC Developers and Users
(7) http://www.cdisc.org/stuff/contentmgr/files/0/0b585db258d68c43cadc5fae135d473a/files/malla_cber_cdisc_session_8_oct_2012_malla.pdf
(8) Federal Food, Drug, and Cosmetic Act (FD&C Act)
(9) HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/OTHER/2013/06/WC500144730.PDF**)**
(10) European Medicines Agency - News and Events - Workshop on clinical-trial data and transparency
(11) http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2013/06/WC500144734.pdf
(12) CDISC Team Updates - Standards in the Electronic Submission Process - 26 Jan 2012 – Webinar

CONTACT INFORMATION

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

Heike Reichert
Bayer Pharma AG
Aprather Weg 18a
42113 Wuppertal
Work Phone:
Fax:
Email: heike.reichert@bayer.com
Web: