PhUSE Single Day Event 2013

Considerations for an ADaM Submission

Jack Shostak

Chris Holland

AMGEN

Duke Clinical Research Institute

Considerations for an ADaM Submission

A collection of ADaM considerations:

- Initial
- Dataset Size
- Define.xml
- Data Reviewer's Guide
- Integration

Initial ADaM Considerations

Initial ADaM Considerations: Should We Use ADaM?

Must I submit ADaM datasets?

Per PDUFA V performance goals:

- FDA shall develop standardized clinical data terminology through open standards development organizations (i.e., the Clinical Data Interchange Standards Consortium (CDISC)) with the goal of completing clinical data terminology and detailed implementation guides by FY 2017.
- FDA shall develop a project plan for distinct therapeutic indications, prioritizing clinical terminology standards development within and across review divisions. FDA shall publish a proposed project plan for stakeholder review and comment by June 30, 2013. FDA shall update and publish its project plan annually.

Initial ADaM Considerations: Should We Use ADaM?

Per FDA CDER Data Standards Strategy – Action Plan:

- The project, “Guidance on Electronic Submission of Applications” will “Issue draft Guidance in fiscal year (FY)2013 specifying the requirement of electronic submission of applications” estimated draft by December 2012.
- The project, “Guidance on Electronic Standardized Study Data” will “Issue draft Guidance to industry specifying requirements for electronic data standards” estimated draft by June 2013.

Initial ADaM Considerations: Should We Use ADaM?

PDUFA V allows the FDA to mandate data standards and CDER is working on it, but CDISC ADaM isn't quite required now. So, you don't have to submit ADaM yet.

However, it is fairly clear from PDUFA V, the Study Data Specifications document, and the CDER Data Standards Strategy document that you ***should*** at least be planning for ADaM compliance.

Contact your review division during planning

- Let them know if you are filing CDISC ADaM files
- Ask if they have ADaM preferences
- Consider contacting the Computational Science Center at CDER for help.

Initial ADaM Considerations: Should We Use ADaM?

Is ADaM worth doing for PK based Phase 1 studies?

Where are you in the development cycle? Are you just starting or do you have “legacy” studies to submit?

By design, you need SDTM as precursor to ADaM.

- Now more widely accepted that SDTM to ADaM is a linear data flow.
- CDER Common Data Standards Issues document says, “Analysis datasets should be derivable from the SDTM datasets”

Initial ADaM Considerations: Should We Use ADaM?

We recommend ADaM when:

- SDTM exists
- Non-trivial derivations and/or imputations are made based on the statistical analysis plan
- Not a lot of legacy studies to convert or integrate

ADaM Considerations After Deciding on ADaM

Published Standards and Guidance

- CDISC: ADaM 2.1, IG 1.0, ADTTE, ADAE, Examples document
- FDA: Study Data Specifications, CDER Common Data Standards Issues document

ADaM Considerations After Deciding on ADaM

What comprises your ADaM (and ADaM compliance)?

- CDER's Common Data Standards Issues document says to submit "key" efficacy and safety analyses.
- FDA's Study Data Specifications document says:

  "While the ADaM provides a valuable representation that may facilitate review, it does not always provide data structured in a way that supports all analyses needed for review. Sponsors should therefore augment their ADaM analysis datasets with analysis datasets based on requirements specified by the review division."

ADaM Considerations After Deciding on ADaM

What comprises your ADaM (and ADaM compliance)?

- What about concomitant medications (e.g., ADCM) that has no ADaM model yet?
- What if you need another “flat” dataset like ADSL?
- ADaM team is working on more models (e.g., proposed “Hierarchical Occurrence Data Structure (HODS)”), but what do you do in the meantime?
- What data to submit as ADaM compliant and what not? Metadata class=“OTHER” datasets.

ADaM Considerations After Deciding on ADaM

CDISC Model versions?

- ADaM 2.1 and IG 1.0 for now
- define.xml – version 1 or 2?
- CDISC controlled terminology
- Other related standards like SDTM

Shifting standards presents implementation challenge

ADaM Considerations After Deciding on ADaM

Metadata

- What metadata and documentation to include in your define files and data review guides?
- ADaM parameter (value) level metadata granularity. How deep do you go?
- ADaM results metadata. What do you include?
- Have a metadata plan. Metadata management is critical.

Clearly there are many technology and process considerations after deciding on ADaM.

ADaM Data Set Size Considerations

Size Considerations

The problem:

- Large text variables (e.g., $200) create a waste of space and large data sets
- Large data sets (e.g., >1 GB) can be slow to load and process on a laptop

The proposed fix:

- Truncating text variables to the minimum size required to avoid data loss can save a lot of space
- Data sets that are still too large should be split

Size Considerations – FDA Opinions:

From the CDER Common Data Standards Issues Document *:

Column Length/Size

For both CDISC and non-CDISC datasets, in order to significantly reduce dataset file sizes, the allotted character variable length/size for each column in a dataset should be the maximum length used. Lengths/Sizes of columns should not arbitrarily be set to 200. For example, if your USUBJID column has a maximum length of 18 being used throughout the dataset, the USUBJID's column size should be set to 18, not to 200. Alternative solutions to this problem that involve some inclusion of a small amount of padding to column width may be acceptable as long as they don't result in significant increases in file size due to the padding.

*** Also mentioned with similar text in the FDA Study Data Specifications Document**

Size Considerations – FDA Opinions:

From the CDER Common Data Standards Issues Document:

Dataset Splitting

If datasets are greater than 1 gb in size, please split the datasets into smaller datasets no larger than 1 gb in size. Datasets should be resized to the maximum length used prior to splitting. This will ensure split datasets have matching variable lengths for future merges. Split data should be noted in the data definition document, clearly identifying the method used for the dataset splitting. For additional information or clarification on file size limitations or splitting of datasets submitted to CDER, contact eData@fda.hhs.gov.

Size Considerations – FDA Opinions:

From the CDER Common Data Standards Issues Document:

LB Domain (Laboratory):

The size of the LB domain is often quite large and can exceed the reviewers' ability to open the file using standard-issue computers. This size issue can be addressed by splitting the large LB dataset into smaller data sets according to LBCAT and LBSCAT, using LBCAT for initial splitting. If the size is still too large, then use LBSCAT for further splitting. For example: use the dataset name lbc.xpt for chemistry lbh.xpt for hematology and lbu.xpt for urinalysis. Splitting it other ways (by subject or file size, etc) makes the data less useable. Sponsors should submit these smaller files **in addition to** the larger non-split standard LB domain file. The smaller split files should be submitted in a separate Sub-directory/SPLIT which is clearly documented in addition to the larger non-split standard LB domain file in the CRT directory. Please see File Size section for information about file size limits.

Size Considerations – FDA Opinions:

From the Study Data Specifications:

CONTENT OF DATASETS AND SIZE OF DATASETS

Each dataset is provided in a single transport file. The maximum size of individual dataset is dependent on many factors. In general, datasets other than SDTM datasets, should be less than 400 MB; SDTM datasets should not be divided. Datasets divided to meet the maximum size restrictions should contain the same variable presentation so they can be easily concatenated.

Datasets which are divided should be clearly named to aid the reviewer in reconstructing the original dataset, e.g., xxx1, xxx2, xxx3, etc. The files that have been divided and need to be concatenated should be noted in the data definition document. This documentation should identify the range of subject numbers (or other criteria used for division) in the label for each of the divided datasets. For further information on file size limitations for files submitted to CDER, contact eData@fda.hhs.gov, for files submitted to CBER, contact CBER.CDISC@fda.hhs.gov

Reducing Variable Length

- Despite all of this guidance, there still seems to be some resistance to heeding the advice
- Admittedly, it is hard to know in advance how much room is needed as files tend to grow
- Tools and a process are needed
  - A SAS-based proposal was presented at the PhUSE/FDA Computational Science Symposium:

    http://www.phuse.eu/download.aspx?type=cms&docID=5124

Reducing Variable Length

- Consider a macro or automated process that can programmatically determine minimum size needed to avoid truncation *within a study*
  - Later we will talk about this problem in the integration context
- Some variables may exist across data sets (e.g., VISIT), so to avoid truncation in the case of a merge, all data sets that a variable appears in need to be considered
- *Implementing CDISC Using SAS: An End-to-End Guide* contains a SAS macro %MAXLENGTH2 to do this at

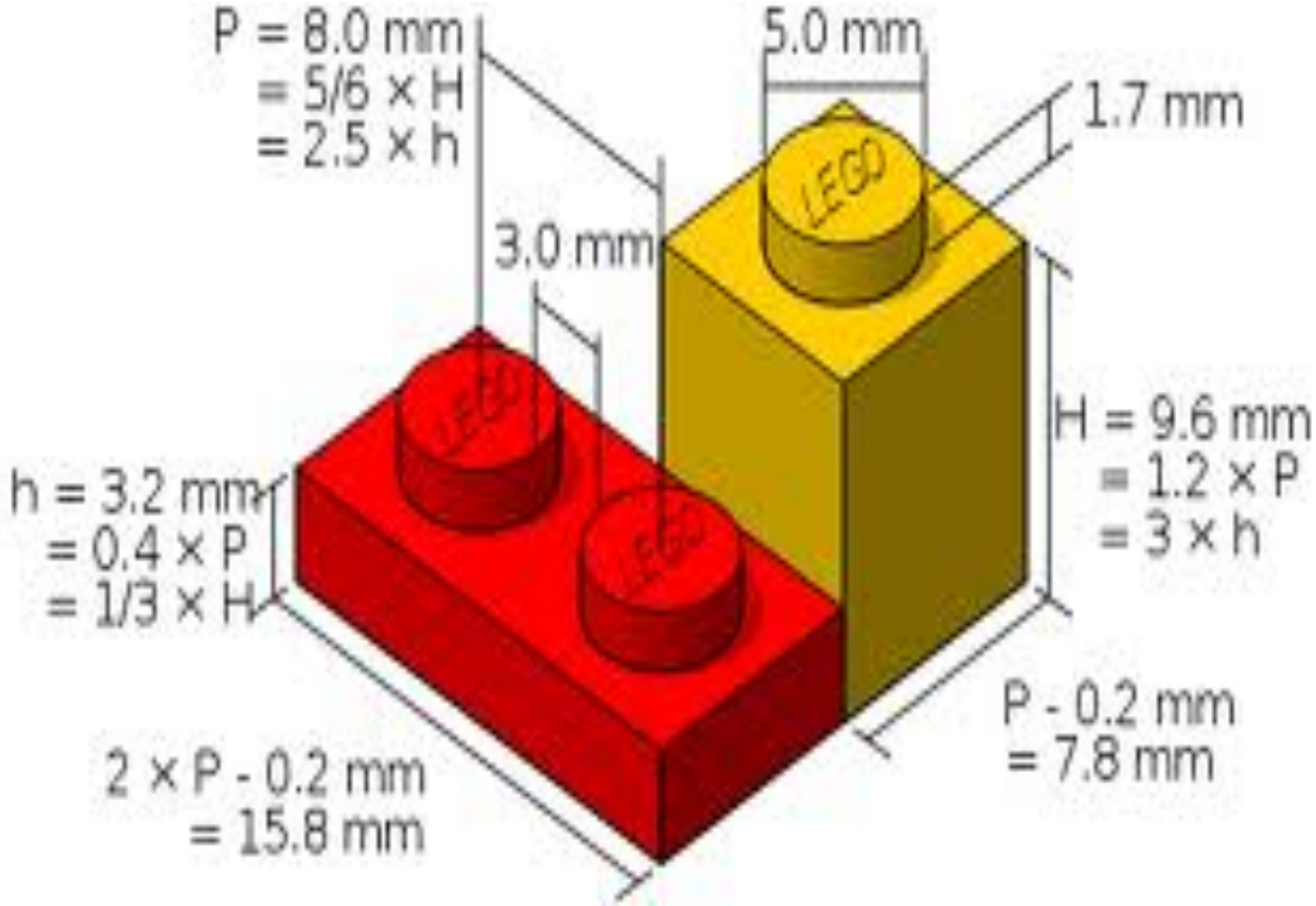
  http://support.sas.com/publishing/authors/extras/64409ecd.pdf

Using %MAXLENGTH2

```
libname sdtm "c:\studyXYZ\sdtm";
libname sdtmnew "c:\studyXYZ\sdtm\new";
%maxlength2(sourcelib=sdtm, outlib=sdtmnew);
```

- Data sets in SDTMNEW will contain variables with a reduced size
- Data set size reductions can be around 20-40%
- Some sort of loop-back process will be needed for the meta-data in define.xml (if you care about accurately stating variable lengths in your define.xml)

A Final Word on Size Considerations



ADaM define.xml Considerations

ADaM and define.xml

New define.xml 2.0 features

- Value (“parameter” in ADaM) level metadata improved. Now instead of just pointing at AVAL the value level metadata can point at any variable if needed.
- Provides where clause machine metadata and “slices” (collection of where clauses) for parameter level metadata definitions
- Old ADaM “source/derivation” metadata can be broken into smaller and more useful chunks.
  - New machine readable “FormalExpression” element as part of Method Definitions.
  - Define.xml 2.0 says, “Comments are not intended to replace a properly defined computational algorithm, which is expected for derived variables. ”

ADaM and define.xml

ADaM Results Metadata

- Unique to ADaM and not formalized in define.xml 2.0
- We still need results metadata ODM extension with define 2.0.
- ADaM results metadata needs enhancement and maturing. What in fact is an analysis result? A whole figure? N? Fisher's exact test?

Data Reviewer's Guide Considerations

"Reviewer's Guide"

FDA draft guidance on Standardized Study Data* referred to reviewer guides as "an integral part of a…data submission":

> Standardized electronic datasets should generally be accompanied by documentation typically referred to as a *Reviewer's Guide*. The Reviewer's Guide describes any special considerations or directions that will facilitate a reviewer's use of the datasets and can help the reviewer understand the relationships between the study report and the data. The Reviewer's Guide is an integral part of a standards-compliant study data submission. No additional guidance is available regarding the format and content of a Reviewer's Guide at this time.

*** Note that this guidance has since been withdrawn so that it can be aligned with PDUFA performance goals**

Study Data Reviewer's Guide

- A number of examples are out there
- A PhUSE working group created a SDTM reviewer data guide template and published the final version of that on 3/18/2013 at

  http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

Study Data Reviewer's Guide

Contents [hide]

Overview & Scope

CGD-010: define.XML does not adequately document SDTM mapping decisions, sponsor-defined domains, sponsor-defined controlled terminology, and sponsor extensions to CDISC controlled terminology. A standardized Data Guide would help to address this documentation gap. The content of the Data Guide should be standardized and developed jointly between CDER, Industry, and CDISC.

Scope: The Study Data Reviewer's Guide (SDRG) will be proposed as a component of the eCTD's Module 5 SDTM data documentation. This document will be defined for SDTM data descriptions only. An Analysis Data Reviewer's Guide will be handled in a future project.

SDRG Core Team

Helena Sviglin, FDA
Douglas Warfield, FDA
Gail Stoner GStoner@its.jnj.com, J&J
Joanna Koft joanna.koft@biogenidec.com, Biogen Idec
Scott Bahlavooni scottab@gene.com, Genentech
John Brega, Pharmastat

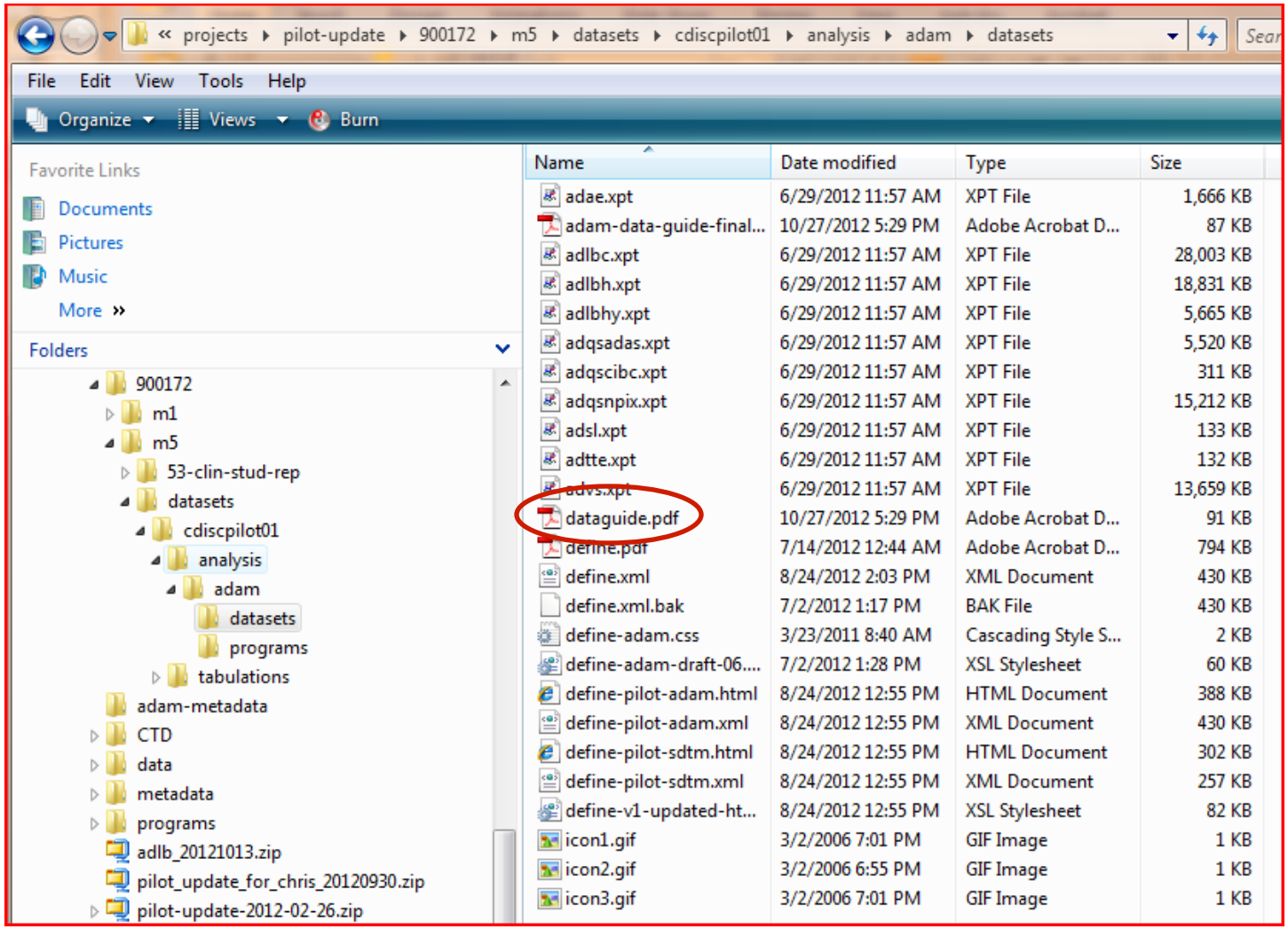
Final SDRG Work Package

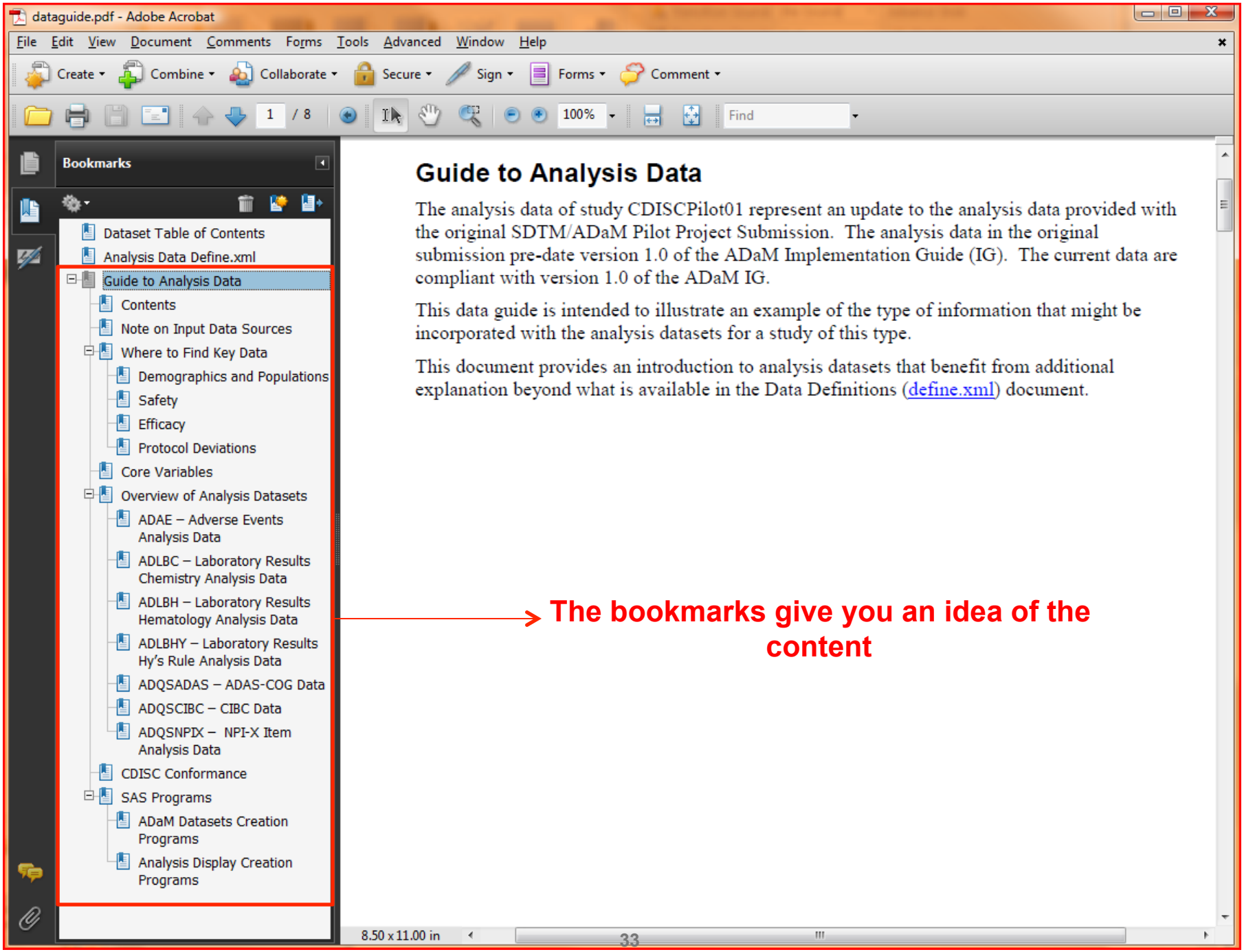
Final SDRG Work Package v1.0 - Includes: SDRG Completion Guidelines, SDRG Template, and Example SDRGs
Please review Readme.txt for a complete description of all files in the work package.

ADaM Pilot Project Update
ADaM Reviewer's Guide

- The original SDTM/ADaM pilot project was recently updated and posted on the CDISC web site
  - Available on the members-only page ☹
- It contains an ADaM-specific data guide
  - The SDTM-specific data guide was not provided so as not to conflict with the efforts of the PhUSE group

ADaM Pilot Project Update
ADaM Reviewer's Guide




dataguide.pdf - Adobe Acrobat
File Edit View Document Comments Forms Tools Advanced Window Help
Create Combine Collaborate Secure Sign Forms Comment
1 / 8
100%
Find
Bookmarks
Dataset Table of Contents
Analysis Data Define.xml
Guide to Analysis Data
Contents
Note on Input Data Sources
Where to Find Key Data
Demographics and Populations
Safety
Efficacy
Protocol Deviations
Core Variables
Overview of Analysis Datasets
ADAE – Adverse Events Analysis Data
ADLBC – Laboratory Results Chemistry Analysis Data
ADLBH – Laboratory Results Hematology Analysis Data
ADLBHY – Laboratory Results Hy's Rule Analysis Data
ADQSADAS – ADAS-COG Data
ADQSCIBC – CIBC Data
ADQSNPIX – NPI-X Item Analysis Data
CDISC Conformance
SAS Programs
ADaM Datasets Creation Programs
Analysis Display Creation Programs
Guide to Analysis Data
The analysis data of study CDISCPilot01 represent an update to the analysis data provided with the original SDTM/ADaM Pilot Project Submission. The analysis data in the original submission pre-date version 1.0 of the ADaM Implementation Guide (IG). The current data are compliant with version 1.0 of the ADaM IG.
This data guide is intended to illustrate an example of the type of information that might be incorporated with the analysis datasets for a study of this type.
This document provides an introduction to analysis datasets that benefit from additional explanation beyond what is available in the Data Definitions (define.xml) document.
The bookmarks give you an idea of the content
8.50 x 11.00 in
33

Analysis Data Reviewer's Guide

- There is an older sample analysis data guide document under "Archived Content" on the PhUSE Study Data Reviewer's Guide page.

- At the recent PhUSE CSS symposium in DC, a new PhUSE working group was formed to create an Analysis Data Reviewer's Guide (ADRG) to be a companion to the Study Data Reviewer's Guide (SDRG).
  - Targeted for completion at end of 2013

ADaM Integration Considerations

Integration Considerations

- One of the main motivations for data standards is to facilitate integration…

- For quick-and-dirty integrations, this is true
- However, for an integrated database used for an ISS or ISE in a regulatory submission, matters can get rather complicated…

Integration Considerations

1. Which data to integrate? ADaM only or SDTM and ADaM?
2. How to represent patients who appear in multiple studies in a patient-level dataset (e.g., ADSL)?
3. How to deal with multiple coding dictionaries across studies?
4. For a given assessment, how should one handle varying categories/controlled terminology across studies?

Integration Considerations

1. Which data to integrate? ADaM only or SDTM and ADaM?
   - To stat programmers, integrating ADaM only probably makes the most sense
   - For a reviewer conducting a safety review using tools that are built off of the SDTM, integrated SDTM data might be useful
   - Other integration considerations appear in *Implementing CDISC Using SAS: An End-to-End Guide*

Integration Considerations

1. Which data to integrate? ADaM only or SDTM and ADaM?
2. How to represent patients who appear in multiple studies in a patient-level dataset (e.g., ADSL)?
   - Consensus appears to be building for adding extra rows for the same patient
   - CDER Common Data Standards Issues document supports this approach, at least for DM

Integration Considerations

1. Which data to integrate? ADaM only or SDTM and ADaM?
2. How to represent patients who appear in multiple studies in a patient-level dataset (e.g. ADSL)?
3. How to deal with multiple coding dictionaries across studies?
   - Same dictionary version should be used across all studies within an integrated database
   - ADAE has variables to capture coding values based on earlier versions used in earlier analyses (e.g., DECODRGy)
     - Easy to find differences: `WHERE AEDECOD ne DECODRG1;`

Integration Considerations

1. Which data to integrate? ADaM only or SDTM and ADaM?
2. How to represent patients who appear in multiple studies in a patient-level dataset (e.g. ADSL)?
3. How to deal with multiple coding dictionaries across studies?
4. For a given assessment, how should one handle varying categories/controlled terminology across studies?
   - Study 1 Investigator Assessment: 1=None, 2=A Little Better, 3=Better, 4=A Lot Better, 5=Completely Better
   - Study 2 Investigator Assessment: -1=Worse, 0=No Improvement, 1=A Little Better, etc.
   - You're on your own here! ☺

Considerations for an ADaM Submission

This talk was a collection of ADaM considerations:

- Initial Considerations
- Dataset Size
- Define.xml
- Data Reviewer's Guide
- Integration

Considerations for an ADaM Submission

Questions?

Constructed with standardized building blocks….