

PhUSE 2014

Paper CD03

What's new in ADaM?

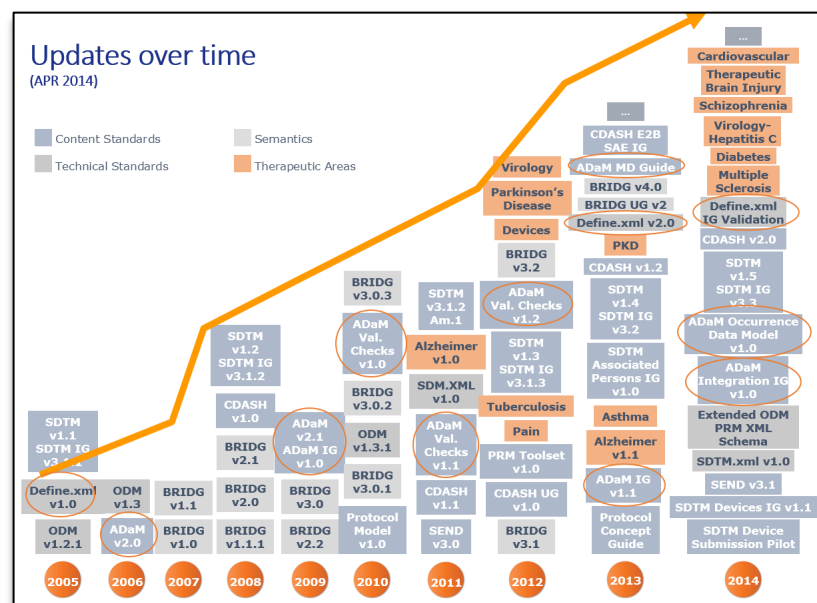
Author: Gavin Winpenny, Business & Decision Life Sciences, Brussels, Belgium

In keeping with CDISC's ongoing mission to develop and support global, platform-independent data standards, the ADaM data standards continue to evolve to meet the needs and experiences of a growing user base. This paper will explore the recent developments related to ADaM standards, documentation and regulatory position.

CDISC and Regulatory Submission Landscape

Over the last decade the rate at which CDISC data standards and processes have been released or updated has increased year-on-year. Keeping abreast of the changes and the implication for process has become an important consideration for all organisations that are involved in the use of CDISC data standards.

When people talk about 'providing the data to CDISC standard', it is in fact many, many standards and guides. Whilst the majority of updates that CDISC has released are focussed on SDTM and CDASH, there has been a number of significant releases over the last year, including a Draft ADaM IG, Draft ADaM Occurrence Data Structure, and Analysis Results Metadata document.



It is also worth remembering that a therapeutic area that did not have a TA standard in the past will have been mapped as a custom domain in SDTM, and subsequently mapped from the custom domain into ADaM.

With the release of a TAUG, e.g. Diabetes, the SDTM structure will change. Both are SDTM-compliant, both will be different. This can, and does, have an impact on ADaM mapping, especially for domain mapping processes

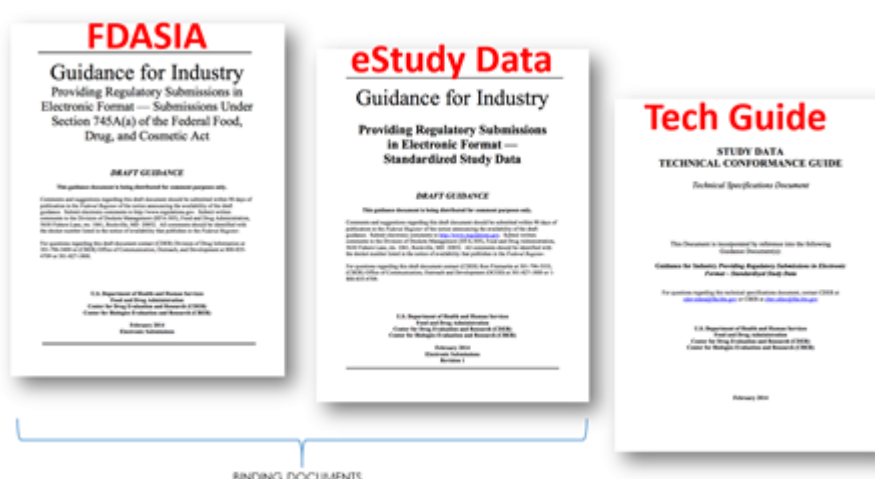
that may be automated. Being mindful of the changes and how and when these are applied, and their impact across time is essential in maximizing the value and usability of standards, and the traceability of data.

In addition to these releases from CDISC, there have been movements in FDA regulatory positions and also a release of an Analysis Data Reviewer's Guide (ADRG) from PhUSE/CSS.

Regulatory Submission Landscape

A question often asked by those discussing CDISC is 'When will the FDA mandate it. When must I *do* CDISC?' The answer to this question is clear(er)* in 2014 with the release of three draft documents from FDA in 2014:

- FDASIA – Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Submission under Section 745A(a) of the Federal Food, Drug and Cosmetic Act.
- eStudy Data - Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Data.
- Tech Guide – Study Data Technical Conformance Guide



The documents are expected to become finalised in 2015, and contain a timescale (24- and 36-months) for the mandating of data standards for electronic submissions.

* Consequently, 2017 is generally given as the milestone year the FDA when submissions will be given CDISC.

The question then arises, does 'standardized study data' also include ADaM, or is it limited to SDTM? It is true that the '[Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Data](#)' is principally focussed on SDTM, however, the '[Study Data Technical Conformance Guide](#)' contains the statement:

- 4.1.7.2. *General Considerations (page 12):*

"One of the expected benefits of analysis datasets that conform to ADaM is that they simplify the programming steps necessary for performing an analysis. ADaM datasets should be derived from the data contained in the SDTM datasets. There are features built into the ADaM standard that promote traceability from analysis results to ADaM datasets and from ADaM datasets to SDTM. Sponsors who provide the software programs used to create ADaM datasets help reviewers to better understand how the datasets were created (see section 4.1.7.8). Each analysis dataset that is shown in the define.xml file should be described."

Whilst this is not a perfect answer in terms of declaring that ADaM is definitely required, it does point to the direction of travel. It would be reasonable to assert that sponsors looking at submitting should be thinking beyond SDTM and making ADaM data standards part of their submission plans for current and up-coming studies.

As a final note on this document, although most organisations will have in place good programming practices, the statement also adds a further driver to organisations to promote, monitor and enforce good programming practices.

CDISC ADaM Team 2014 Goals

At the time of submitting the abstract for the paper in March 2014, the goals for the CDISC ADaM Team included:

- Document on the introduction of ADaM methodologies to support ISS/ISE and an integrated ADSL structure
- Explore integrating ADaM in SHARE
- Development of CFAST/Therapeutic Area ADaM standards
- Development of ADaM examples and best practices for questionnaire data which can be applied across TA standards
- Harmonize ADaM model document with Define 2.0 and better address representation of results-level metadata
- Update Compliance rules for consistency with the current version of ADaM-IG
- Continued work on ADaM-IG updates
- Publication of final versions of ODS v1 and ADaMIG v1.1

At the time of the PhUSE conference in October 2014, not all these activities had been achieved or were still in draft state. Consequently, timelines for completion of these activities will likely stretch into 2015, and it is advisable that regular visits be made to the [CDISC ADaM website](#) to check on latest progress and updates.

CDISC ADaM Implementation Guide (Version 1.1 Draft)



Released in May 2014, the Draft Implementation Guide contained a number of changes (full list available in Appendix B of the document). Generally, the changes were fairly minor, with some additions, clarifications, tidying up. Amongst the changes/updates was:

- Announcement of deprecation of PARAMTYP variable
- Increased padding of 'x' to "xx" in a variable names (e.g., TRTxxP, APxxSDT) where "xx" is replaced with a zero-padded two-digit [01-99].
 - Increased padding of 'z' to 'zz' in a variable name (e.g., ANLzzFL) where "zz" is replaced with a zero-padded two-digit integer [01-99].
- Noted that variable length can vary between SDTM and ADaM variable (the FDA has given several presentations over the last year in which they have raised the issue of 'empty space' in datasets due variable lengths being padded out);
- Added variables for ADSL - AGEGRY, ACTARM, TSEQPGY, DOSE;
- Added variables for BDS datasets - ASEQ, dose variables, MCRITY and corresponding flags
- Made record-level Population flags (RFL) and parameter-level Population flags (PFL) variables permissible instead of conditional;
- Clarification regarding when certain timing variables should be included in ADSL vs. BDS .

At the time of the PhUSE conference there is no official announcement as to when the Final release will be available for the updated Implementation Guide.

CDISC ADaM Occurrence Data Structure (ODS) v1.0 (Draft)



When the ADaM Data Structure for Adverse Events was released in May 2012, it was quickly picked up by the ADaM community as a framework that could be extended to other occurrence event data domains.

In recognition of this usage, with the release of the Draft ADaM Occurrence Data Structure in March 2014, the ADAE document was expanded to support Occurrence Data Models such as Medical History, Concomitant Medications, and Lab Events.

Typically, occurrence event data are focused on subject count analysis, where a subject may be represented multiple times in a category, and consequently AVAL or AVALC are not required.

Occurrence Data Structure: Adverse Events

What does this mean for ADAE? The good news is the changes are minor. ADAE under ODS structure has been mapped and is backwards compatible. The examples from ADAE Structure document are copied into new ODS structure document. For the dataset structure there are minor label changes to make them general rather than specific to AE, and 'Class of Dataset' has been made non-specific.

Table 4.1 Example of ADaM ADAE Dataset Metadata

Dataset Name	Dataset Description	Dataset Location	Key Variables of Dataset	Class of Dataset	Documentation
ADAE	Adverse Event Analysis Dataset	pathname/analysis/datasets/adae.xpt	USUBJID, AEDECOD, AESEQ, (optional)	ADAE*	ADAE.SAS Example: Dictionary used is

Table 4.1.1 Example of ADaM ODS Dataset Metadata

Dataset Name	Dataset Description	Dataset Location	Key Variables of Dataset	Class of Dataset	Documentation
ADXXXXXX	<Dataset label>	adxxxxxx.xpt	List variables, such as USUBJID, --SEQ	ODS	Example: Dictionary used is MedDRA V11.1

Occurrence Data Structure: Concomitant Medications

For Concomitant Medications, most variables come from:

- SDTM's CM, SUPPCM and ADaM ADSL.
- Instead of MedDRA, Concomitant Medications use of WHODRUG.

Table 9.2.1 Example of ADCM Variable Metadata

Dataset Name	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Term	Source / Derivation
ADCM	STUDYID	Study Identifier	text	\$3		CM STUDYID
ADCM	USUBJID	Unique Subject Identifier	text	\$11		CM USUBJID
ADCM	CMSEQ	Sequence Number	integer	3.0		CM CMSEQ
ADCM	CMTRT	Reported Name of Drug, Med or Therapy	text	\$200		CM CMTRT
ADCM	CMMODIFY	Modified Reported Name	text	\$200		CM CMMODIFY
ADCM	CMDECOD	Standardized Medication Name	text	\$200	WHODRUG	CM CMDECOD WHO Drug Dictionary March 2012
ADCM	PREFCODE	Preferred Term Code	text	\$200	WHODRUG	CM PREFCODE WHO Drug Dictionary March 2012
ADCM	ATC1C	ATC Level 1 Code	text	\$200	WHODRUG	ATC Level 1 Code WHO Drug Dictionary March 2012
ADCM	ATC2C	ATC Level 2 Code	text	\$200	WHODRUG	ATC Level 2 Code WHO Drug Dictionary March 2012
ADCM	ATC3C	ATC Level 3 Code	text	\$200	WHODRUG	ATC Level 3 Code WHO Drug Dictionary March 2012
ADCM	ATC1T	ATC Level 1 Text	text	\$200	WHODRUG	ATC Level 1 Text WHO Drug Dictionary March 2012
ADCM	ATC2T	ATC Level 2 Text	text	\$200	WHODRUG	ATC Level 2 Text WHO Drug Dictionary March 2012
ADCM	ATC3T	ATC Level 3 Text	text	\$200	WHODRUG	ATC Level 3 Text WHO Drug Dictionary March 2012
ADCM	AOCFFL	1st Occurrence of Any AE Flag	text	\$1	Y	<Sponsor will insert derivation here>
ADCM	ATC1FL	ATC Level 1 - First Occurrence Flag	text	\$1	Y	<Sponsor will insert derivation here>
ADCM	ATC2FL	ATC Level 2 - First Occurrence Flag	text	\$1	Y	<Sponsor will insert derivation here>
ADCM	ATC3FL	ATC Level 3 - First Occurrence Flag	text	\$1	Y	<Sponsor will insert derivation here>
ADCM	AOCFFL	1st Occurrence of Preferred Term Flag	text	\$1	Y	<Sponsor will insert derivation here>

To support analysis, Indicator and Occurrence flags have also been provided in the data structures.

Indicator and Occurrence flags needed for analysis have been provided.

Table 4.2.5.4 Concomitant Medications Indicator Variables

Variable Name	Variable Label	Type	Code List / Controlled Terms	Core	CDISC Notes
ONTRTFL	On-Treatment Flag	Char	Y	Cond	Character indicator of whether the observation occurred while the subject was on treatment. Example derivation: If ADSL TRTSDT <= ASTDT <= ADSL TRTEDT then ONTRTFL = 'Y' This variable is conditional on whether the concept of on-treatment is a feature of the study and used in analysis.
PREFL	Pre-treatment Flag	Char	Y	Cond	Character indicator of whether the observation occurred before the subject started treatment. Example derivation: If ASTDT < ADSL TRTSDT then PREFL = 'Y' This variable is conditional on whether the concept of pre-treatment is a feature of the study and used in analysis.
FUPFL	Follow-up Flag	Char	Y	Cond	Character indicator of whether the observation occurred while the subject was on follow-up. Example derivation: If ASTDT > ADSL TRTEDT then FUPFL = 'Y' This variable is conditional on whether the concept of follow-up is a feature of the study and used in analysis.

Table 9.3.1 Sample ADCM Data

Row	STUDYID	USUBJID	CMSEQ	CMTRT	CMMODIFY
1	ABC	ABC-001	1	TYLENOL	TYLENOL
2	ABC	ABC-001	2	TYLENOL	TYLENOL
3	ABC	ABC-001	3	TYLENOL	TYLENOL
4	ABC	ABC-001	4	TYLENOL	TYLENOL
5	ABC	ABC-001	5	TYLENOL	TYLENOL
6	ABC	ABC-001	6	TYLENOL	TYLENOL

Row	CMDECODE	PREFCODE	ATC1C	ATC2C	ATC3C	ATC1T
1	PARACETAMOL	N02BE	N	N02	N02B	NERVOUS SYSTEM
2	PARACETAMOL	N02BE	N	N02	N02B	NERVOUS SYSTEM
3	PARACETAMOL	N02				
4	PARACETAMOL	N02				
5	CONTACMS	N02				
6	FLUTICASONE PROPIONATE	R01				

Row	ATC2T	ATC3T
1	ANALGESICS	OTHER ANALGESICS AND ANTIPYRETICS
2	ANALGESICS	OTHER ANALGESICS AND ANTIPYRETICS
3	ANALGESICS	OTHER ANALGESICS AND ANTIPYRETICS
4	ANALGESICS	OTHER ANALGESICS AND ANTIPYRETICS
5	ANALGESICS	OTHER ANALGESICS AND ANTIPYRETICS
6	NASAL	NASAL

Row	ATC1FL	ATC2FL	ATC3FL	AOCCL	AOCCPFL	CMINDC
1	Y	Y	Y	Y	Y	HEADACHE
2						HEADACHE
3						HEADACHE
4						HEADACHE
5					Y	COLD
6	Y	Y	Y		Y	CHOLESTEROL

- The **Occurrence** Flags (AOCCzzFL) are permissible, and not required.
- The main purpose of these flags is to facilitate data point traceability between records in the dataset and unique counts in the summary displays.

Occurrence Data Structure: Medical History

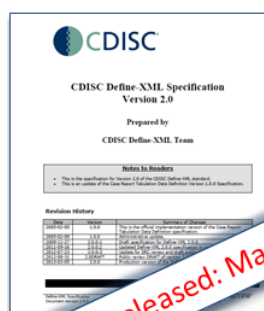
For Medical History, most variables come from:

- SDTM's MH and ADaM ADSL.
- MedDRA dictionary is used to code terms.
- Includes any variables required for analysis (e.g. could add severity of the History event).

Table 10.2.1 Example of ADMH Variable Metadata

Dataset Name	Variable Name	Variable Label	Variable Type	Display Format	Code List / Controlled Terms	Source / Derivation
ADMH	USUBJID	Unique Subject Identifier	text	\$11.		MH.USUBJID
ADMH	MHSEQ	Sequence Number	integer	3.		MH.MHSEQ
ADMH	MHSCAT	Category for Medical History	text	\$200.		MH.MHSCAT
ADMH	MHSCAT	Sub Category for Medical History	text	\$200.		MH.MHSCAT
ADMH	MHDECODE	Dictionary-Derived Term	text	\$200.		MH.MHDECODE
ADMH	MHBODSYS	Body System or Organ Class	text	\$200.		MH.MHBODSYS
ADMH	MHTERM	Reported Term for the Medical History	text	\$200.		MH.MHTERM
ADMH	MHSTDTC	Start Date/Time of Medication	datetime	\$16.	ISO \$601	MH.MHSTDTC
ADMH	ASTDT	Analysis Start Date	integer	date9.		From MH.MHSTDTC, converted to SAS Date. Any derivations to derive partial start dates are applied here and listed in comments.
ADMH	ASTTM	Analysis Start Time	integer	time5.		From MH.MHSTDTC, converted to SAS Time.
ADMH	ASTDTM	Analysis Start Date/Time	integer	datetime16.		From MH.MHSTDTC, converted to SAS Datetime.
ADMH	MHENDTC	End Date/Time of Medication	datetime	\$16.	ISO \$601	MH.MHENDTC
ADMH	AENDT	Analysis End Date	integer	date9.		From MH.MHENDTC, converted to SAS Date. Any derivations to derive partial start dates are applied here and listed in comments.
ADMH	AENTM	Analysis End Time	integer	time5.		From MH.MHENDTC, converted to SAS Time.
ADMH	AENDTM	Analysis End Date/Time	integer	datetime16.		From MH.MHENDTC, converted to SAS Datetime.
ADMH	MHENRF	Continuation Flag	text	\$20.		MH.MHENRF

CDISC Define-XML v2 and ADaM Results Metadata



CDISC Define-XML v2 [Specification](#) was released in March 2013, and is used to describe CDISC SDTM, SEND and ADaM datasets for the purpose of submissions to the FDA, as well as any proprietary (non-CDISC) dataset structure. With the release of Version 2 of define notable changes included:

- Value (“parameter” in ADaM) level metadata improved. 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 segments, with new machine readable “Formal Expression” element as part of Method Definitions (NB, Define.xml 2.0 states, “Comments are not intended to replace a properly defined computational algorithm, which is expected for derived variables.”)



At the time of its release, Define-XML v2 did not provide for the inclusion of Analysis Results Metadata to address this, CDISC has issued a draft Analysis Results Metadata document package can be downloaded from:

<http://portal.cdisc.org/CT/Review%20Documents/AResM-for-Define-XML-1.zip>

Although, the review period has closed, the document package is still available for download.

The inclusion of Analysis Results Metadata with define.xml will provide new opportunities to share key information related to an analysis display (aka TFL) in a well-structured file.

Analysis Results Display or Study CDISC-Sample		Analysis Results Display Description	
Table 14-3.01	1	2	
Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)		
Analysis Result	Dose response analysis for ADAS-Cog changes from baseline	3	Analysis Results Description
Analysis Parameter(s)	ACTOT = Adas-Cog(11) Subscore	4	Analysis Results Parameter
Analysis Variable(s)	CHG (Change from Baseline)	5	Analysis Dataset Variable
Analysis Reason	SPECIFIED IN SAP	6	Analysis Results Reason
Analysis Purpose	PRIMARY OUTCOME MEASURE	7	Analysis Results Analysis Purpose
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANLO1FL = "Y"]	8	Analysis Results Dataset
Analysis Results Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1	9	Analysis Results Where Clause
Programming Statements	[SAS version 9.2]	10	Analysis Results Documentation hyperlink
	proc glm data = ADQSADAS; where EFFFL='Y' and ANLO1FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT"; class SITEGR1; model CHG = TRIPN SITEGR1; run;	11	Analysis Results Programming Statements

Whilst the information from bullets 1-8 are readily available, the creation of “Analysis Results Where Clause” (bullet 9 in the above figure) will likely require organisations to further refine / restructure information track in their metadata process to support the underlying XML.

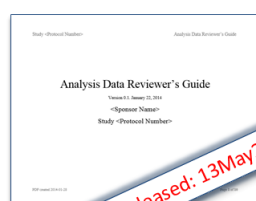
9

```
<def:WhereClauseDef OID="WC.Table_14-3.01.R.1.ADQSADAS">
  <RangeCheck Comparator="EQ" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.PARAMCD">
    <CheckValue>ACTOT</CheckValue>
  </RangeCheck>
  <RangeCheck Comparator="EQ" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.AVISIT">
    <CheckValue>Week 24</CheckValue>
  </RangeCheck>
  <RangeCheck Comparator="EQ" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.EFFFL">
    <CheckValue>Y</CheckValue>
  </RangeCheck>
  <RangeCheck Comparator="EQ" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.ANL01FL">
    <CheckValue>Y</CheckValue>
  </RangeCheck>
</def:WhereClauseDef>
```

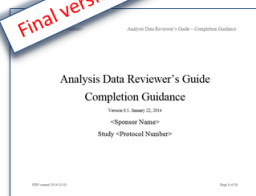
Additionally, the specification document does not define what constitutes “Analysis Results Programming Statements” (bullet 12). It could be argued that as the purpose of an ADaM dataset is to be ‘one PROC away’ then derivations are an upstream activity of the reporting stage, so are not an issue. However, the example given in the document shows a PROC GLM model, and makes use of the where clause in bullet point 9.

Agreeing what, and how much, code usefully represents an analysis is a matter of opinion and will likely vary from organisation to organisation. On top of this, should organisations choose to include statistical procedures such as the example above, there may be the additional complication of resolving parameterised macro code, and ensuring the resolved code stored in the metadata matches the SAS code used to create the data display.

Analysis Data Reviewer’s Guide



Final version released: 13 May 2014



Outside of the CDISC organisation, At PhUSE CSS symposium in 2013, a PhUSE working group was formed to create an Analysis Data Reviewer’s Guide (ADRG).

The Analysis Data Reviewer’s Guide plus Completion Guidance documents drafted during 2013, and released as Final on PhUSE Wiki Website:

http://www.phusewiki.org/wiki/images/0/0d/ADRG_V1.0_2014-05-13.zip.

Although ADaM provides a robust metadata framework, FDA Reviewers benefit from additional, human-readable, documentation of analysis methods, data sets, and programs that cannot be fully explained within the ADaM metadata.

The ADRG package contains a template to be used in submissions with completion guidelines and examples to help sponsors with their implementation. It provides an orientation to the submitted data in a consistent and usable format.

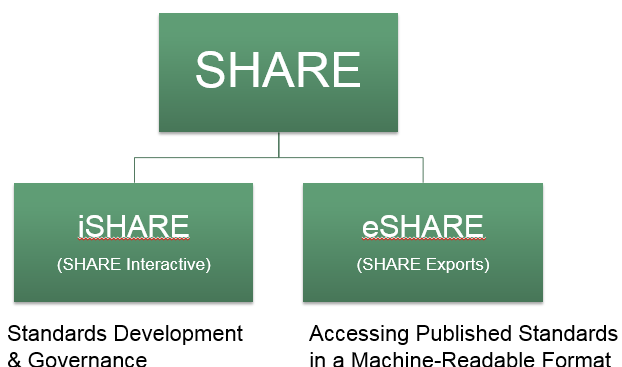
CDISC SHARE

As part of CDISC's standards implementation technical roadmap, the *Shared Health and Research Electronic Library* (SHARE) initiative will provide a global electronic repository for developing, integrating and accessing CDISC standards metadata in electronic format.

SHARE should dramatically improve the quality, reusability and integration across CDISC standards and controlled terminologies, and improve interoperability with healthcare.

Although not currently in SHARE, there is an ongoing stream within the SHARE initiative to add ADaM content for close of 2014.

The initial content to be included will be: ADSL; BDS as defined in ADaM IG version 1; and ADAE as defined version 1 of the ADAE document.



#	Milestone Description	Date
1	Production eSHARE content download site	July 25
2a	SDTM 3.1.3 Available	June 13
2b	SDTM 3.2 Available	July 25
3	Initial Value Level Metadata content in eSHARE	August 29
4	SEND 3.0 Available	September 26
5	Research concepts prototype	September 26
6	Initial ADaM content Available	Q4
7	Research concepts meta-model and initial content	Q4
8	RDF/OWL export format available	Q4

At present, access to CDISC eSHARE is restricted to Platinum CDISC members. This provides easy access to machine-readable standards published from the iSHARE production library. In the near term, only those developing and governing the standards will use the iSHARE software interactively. As CDISC adjusts to the new standards development processes, and loads more standards content into SHARE, CDISC will slowly grow the community of interactive SHARE users.

Further information or to volunteer on a CDISC SHARE sub-team, can be found at the CDISC website: <http://www.cdisc.org/cdisc-share>

How to manage this change?

First off – don't panic. The past is still valid, but the future should be planned for. Most of the changes are evolution, not revolution. Many documents formalise what is common practice (ODS' MH and CM being good examples).

For ADaM, documentation and specifications should demonstrate good understanding of the relationship between analysis results, analysis datasets and SDTM domains. It should support:

- metadata traceability - finding a relationship between an analysis result and analysis dataset(s), or a relationship of the analysis variable to its source dataset(s) and variable(s));
- data point traceability - finding the predecessor record(s)

For standards and process change, the implementation needs to be well managed. Implementation strategies, e.g. assessment, deployment, transition, traceability for ongoing and legacy studies / ISS and ISE will maintain control of standards and the business rules applied to them.

With the implementation of Define-XML v2 and the extension to include Analysis Results Metadata Specification there is the potential for substantial impact on programming structure and management, but again good practice in development of computational methods, can assist.

Conclusion

As sponsors are increasingly submitting their analysis data in ADaM data structures, the demands and focus placed on ADaM standards will encourage organisations to rethink both processes and procedures to maximise the benefits of improved standards and technologies available. The development of standards

CONTACT INFORMATION

Contact the author at:

Gavin Winpenny

Business & Decision Life Sciences

St Lambertusstraat 141 Rue Saint-Lambert

1200 Brussel – Bruxelles

Work Phone: +32 2 774 11 00

Fax: +32 2 774 11 99

Email: gavin.winpenny@businessdecision.com

Web: www.businessdecision-lifesciences.com