ICON
A Symbol of Excellence

Development Solutions

Use of Metadata in an Outsourcing Strategy

Gregory Steffens

ICON Development Solutions

PhUSE August 2011

Metadata Constituents

- A standard list of database attributes to include in any description of a database or of a data standard
- Put in a standard set of data structures that can be read by programming code
- The attributes must be highly structured in order to be usable by program code
- To define a standard for defining data standards
- Enables easy publication in different formats, html, word, pdf, xml, etc. Generate documents from metadata, not metadata from documents!

Standard Database Attributes

- Data Set Level
  - Short/long names, data set location, order
- Variable level
  - Short/long name, type, length label, primary key flag, format, value list name, suppqual flag, code/decode relationship, order, aCRF location, etc.
- Valid values
  - Value list name, start/end value, short decode, long decode, rank
- Descriptions
  - Source name, derivation description
- Row-level attributes
  - Identical to variable level attributes but for subsets of rows defined by a parameter variable value

Row-Level Metadata

- Necessary to fully describe tall-thin data set structures

USUBJID	SYSBP	BPSYSLOC	BPSYSU	HEIGHT	HEIGHTU	WEIGHT	WEIGHTU	BMI	BMIU
1	120	STANDING	mm Mg	185	CM	90	KG	26.3	Kg/m**2

USUBJID	VSTESTCD	VSLOC	VSORRES	VSORRESU
1	SYSBP	STANDING	120	Mm Mg
1	HEIGHT		185	CM
1	WEIGHT		90	KG
1	BMI		26.3	Kg/M**2

Metadata Structure

- Structured content to enable programmatic access
- Storage structure is separate from publication structure – maximize programmatic access and user friendly access by people
- Storage structure is also separate from the data entry format
- Maximize sharing of information within the metadata, e.g. values lists and descriptions. Normalize the metadata design.

Some Principles of Metadata Design

- Rigorously standardized for all database and standard descriptions, no metadata design change for database types!
- Maximize structured information and programmatic access, e.g. primary keys flagged instead of listed
- Enter once; use many. e.g. descriptions and values
- Derivation logic in descriptions though

Objectives of Metadata

ICON
A Symbol of Excellence
Development Solutions

- It is critical to explicitly define the objectives. Many disagreements arise from an unstated difference in assumed objectives
- Objectives allow evaluation of the success of the metadata design; e.g. retrospective description for esubmission or prescriptive enabler of automation.
- Data standards and metadata are a means to an end and that end includes automation and transparency of data flow

Objectives

- Include enough attributes to enable the automation of everything
- Don’t automate each domain, automate all domains and all standards with a single set of macros that read metadata that tells them what to do.
- Metadata greatly reduces the assumptions about the data that your automation code needs to make - meta-automation with code that does not assume any database standard.
- Maximize use of data standards and minimize the assumptions code makes and much greater coverage of data automation can be attained. This is essential for CROs and also for pharmaceutical companies.

Industry Metadata Standard

ICON
A Symbol of Excellence
Development Solutions

- We need a metadata standard to exchange information about data standards and data specifications
- Current practice is to use metadata that is quasi-standardized at each company or to use old-fashioned word documents
- This causes great inefficiencies
  - Translating between metadata standard structures and attribute lists causes large amounts of unnecessary work

Some of the Problems

ICON
A Symbol of Excellence

Development Solutions

- Excel often used, with un-typed columns, not 2-dimensional and confusion between storage, entry and presentation structures
- Inconsistent metadata structures even within a sponsor company, between different standards, specifications and versions of the same standard
- Unstructured information like
  - controlled terminology concatenated in large character variables
  - Primary key variables in lists instead of flags
- Inconsistent attribute list
- CDISC excel workbooks have these problems too

What could be

ICON
A Symbol of Excellence

Development Solutions

- An industry metadata standard does exist – the define.xml. This has a standard list of attributes and a standard structure
- But the standard structure is xml and difficult to access programmatically
- A solution is a standard relational metadata structure that contains the list attributed in the define.xml schema but in a programmatically accessible format. This approach was used in the two CDISC pilot projects with success.
- All data standards and specifications would be stored in this standard metadata structure
- Standard GUI for entry and modification of metadata content
- A set of standard presentations of metadata content

What to do with Standard Metadata

- Data standards published in a standard way
- Data specifications exchanged between organizations and software systems
- Automation that uses metadata to inform the code about the database, instead of the code making assumptions about the database. Metadata is code.
- A metadata standard is more important that data standards!

An Outsourcing Process

ICON
A Symbol of Excellence

Development Solutions

- Submit corporate data standards to all CROs in an industry standard metadata structure.
- Create a study data specification by subsetting the metadata-resident data standard
- Compare the study specification to an IDB standard so that integrating the study data will be easier. Using multiple CROs for different studies is less of a problem.
- Create the define.xml / pdf from metadata in minutes, including all the hyperlinks
- Send the source data and specification to a CRO
- The CRO uses automation to build and validate the database
- The sponsor validates the data from the CRO by automated comparison of the data to the metadata-resident specification

Principles of the Process

- Metadata is ***prescriptive*** rather than merely descriptive
  - Prescriptive metadata created at the start has much more value than descriptive metadata created at the end
- Metadata is populated at the start of the project and supports automation throughout the process from creation to FDA submission
  - Publish the plan
  - Check compliance to standard
  - Build the database
  - Validate the data
  - Create define file for the FDA
  - Metrics – measure compliance of requirements to standard and the data to requirements
- Enter once; use many!
- Metadata structure is identical in all applications to support sharing of content

Other Kinds of Metadata

- After metadata comes map metadata that supports even more complex automation of the transformation of data from source to target structures, like creating SDTM or integrated databases to support ISS/ISE
- A Data Transformation Engine requires metadata and map metadata and provides huge efficiency gains and transparency in the data flow (transforms not “hidden” in code or documents)
- The term “metadata” is often used more broadly to also mean data that describes trial design, treatment arms, tables, figures and listings, titles/footnotes, etc. A more general term is “data driven applications”, which include metadata driven applications.

Examples of macros

List of some of the macros and their functionality which help us to achieve efficiency and ensure good quality:

Mdprint/md2odm	Publish in html or xml format
Mdatribs	Apply attributes defined in metadata to a data library
Ut_find_decodes	Finds decode variables and their attributes
Dt_make_decodes	Creates decode variables
Dt_copy_headers	Copies header variables from source to target data sets
Mdcompare / mdcompare_print	Compares metadatabases to each other, such as a study requirement to a standard or a study to a study
mdcheck	Checks data and reports discrepancies with the metadata
mdbuild	Builds metadata to describe an existing data library
mdfreqvals	Creates the values metadata set (supplements mdbuild)

Examples of macros

dtmap	Top level macro that users call to transform data from one format to another, e.g. raw to SDTM to ADaM to IDB
Dt_thin2wide	Convert tall-thin to short-wide
Dt_wide2thin	Convert short-wide to tall-thin
Tool_code_lib	Documents program code
Ut_saslogcheck	Checks SAS logs for disallowed messages
Ut_age_years	Computes age in years
Ut_truncate_long_chars	Truncates long character variable lengths to least length to hold longest value
mdport	Creates a transport file of a metadatabase to archive versions
md2excel / excel2md	Converts metadata between SAS and excel
mdmkdsn	Creates 0-observation data sets as defined in metadata

Examples of macros

Suppqual_make	Creates the suppqual data sets, by reading the suppqual flag in the metadata to identify supplementary qualifiers
Suppqual_get	Gets supplementary qualifier variables from the suppqual data sets and adds them to their proper domain
Dtmap_values	Changes the value of variables by reading value map metadata
Mdformats	Create user formats from values metadata set
Missvars	Report variables that have a missing value in all observations
Missobs	Report observations where all variables have a missing value