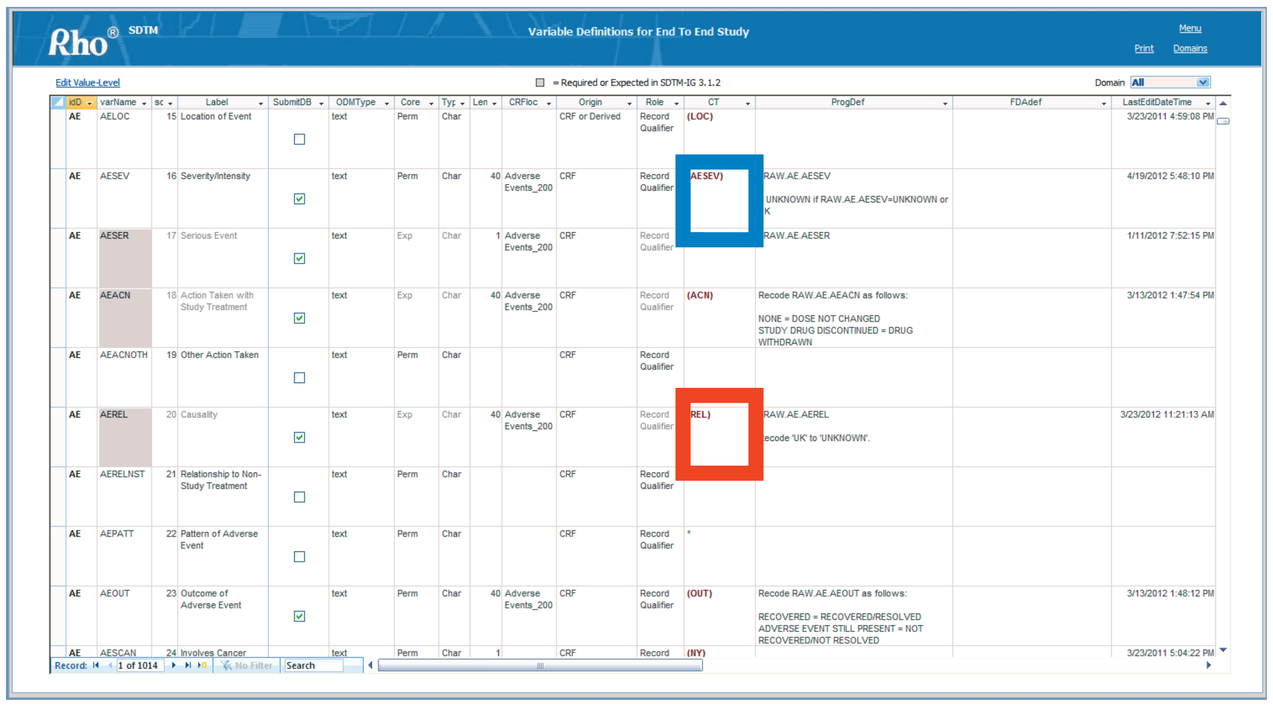
PP24: Controlling Controlled Terminology

Ryan Burns, Rho, Inc., Chapel Hill, USA

Making Metadata Easy For All To Work With:

Going Beyond Excel For a More Robust Solution

A system that collects and stores machine-readable metadata is critical to the operation of reliable and efficient clinical trial submission projects. In addition to storing metadata, the system at Rho creates programming specifications, sets variable attributes, and outputs define files. The system is comprised of an Oracle database with a Microsoft Access user interface.
Below is an example variable definition form. All variable-level metadata is entered in this form. It includes variable attributes, definitions and algorithms for programmers, and metadata destined for define files.



Rho's metadata database has four main tables—one each for study-level, dataset-level, variable-level, and value-level metadata. Three of these tables are presented below with data from the example AE form.

Studies table

IDSTUDY	STUDYTITLE	PROTOCOL	COMPOUND	PROJROOTDIR	FIELDSUSED-VARIABLESTABLE	FIELDSUSED-DATASETSTABLE	STUDYCREA-TIONDATETIME	STUDYCREAT-EDBY	RELSTUDS	SDTMVERSION-NUMBER	LBLVARS_CUSTOM_TXT1	LBLDSNS_CUSTOM_TXT1
1412	End To End	9999-ABC-999	ETOE	Q:\ETOE	VARNAME, IDDOMAIN, FDADEFINITION, CUS-TOM_TXT1,...	DOMAIN, PURPOSE, STRUC-TURE, KEYS, CLASS, CUS-TOM_TXT1,...	6/6/2012 4:44:40 PM	r70818		3.1.2	SAPRef	CRFPages

Datasets table

idDomain	Domain	submitDB	Purpose	Structure	Keys	class	Description	input_notes	output_notes	CRFPages	lastEdit-DateTime	lastEditedBy
122705	AE	Yes	Tabulation	One record per adverse event per subject	STUDYID, USUB-JID, AETERM, AESTDTC	Events	Adverse Events	Input dataset is RAW.AE. Subset where RAW.AE.AETERM is not missing.	Keep only subjects present in DM where ARMCD ne 'SCRNFAIL'.		9/6/2012 12:53:02 PM	r70818

Variables table

idVariables	idDomain	varName	sortOrder	label	submitDB	Core	ODMType	type	length	Origin	Role	CT	progDef	FDAdef	SAPRef
5673926	122705	AELOC	15	Location of Event	No	Perm	text	Char		CRF or De-rived	Record Qualifier	24			
5673941	122705	AESEV	16	Severity/ Intensity	Yes	Perm	text	Char	40	CRF	Record Qualifier	4	=RAW.AE.AESEV = UNKNOWN if RAW.AE.AESEV=UNKNOWN or UK		
5673913	122705	AEACN	18	Action Taken with Study Treat-ment	Yes	Exp	text	Char	40	CRF	Record Qualifier	51	Recode RAW.AE.AEACN as follows: NONE = DOSE NOT CHANGED STUDY DRUG DISCONTINUED = DRUG WITHDRAWN STUDY DRUG INTERRUPTED = DRUG INTER-RUPTED UNKNOWN or UK = UNKNOWN		
5673914	122705	AEACNOTH	19	Other Action Taken	No	Perm	text	Char		CRF	Record Qualifier				
5673932	122705	AEREL	20	Causality	Yes	Exp	text	Char	40	CRF	Record Qualifier	71	=RAW.AE.AEREL Recode 'UK' to 'UNKNOWN'.		
5673933	122705	AERELNST	21	Relationship to Non-Study Treat-ment	No	Perm	text	Char		CRF	Record Qualifier				
5673929	122705	AEPATT	22	Pattern of Adverse Event	No	Perm	text	Char		CRF	Record Qualifier	*			
5673928	122705	AEOUT	23	Outcome of Adverse Event	Yes	Perm	text	Char	40	CRF	Record Qualifier	18	Recode RAW.AE.AEOUT as follows: RECOVERED = RECOVERED/RESOLVED ADVERSE EVENT STILL PRESENT = NOT RE-COVERED/NOT RESOLVED ADVERSE EVENT RESOLVED WITH SEQUE-LAE = RECOVERED/RESOLVED WITH SEQUE-LAE DEATH = FATAL UNKNOWN or UK = UNKNOWN		
5673934	122705	AESCAN	24	Involves Cancer	No	Perm	text	Char	1	CRF	Record Qualifier	3			

Why Recreate Lists Again and Again? Links Make Your Database More Efficient and Your Terminology More Controlled.

The variable-level metadata does not contain complete controlled terminology metadata, but rather contains a link to a controlled terminology list. Controlled terminology information is stored in three tables (shown at right). Editing of controlled terminology lists is done using popup forms that are shown below.
Notice how the form will allow additions for the extensible list (REL, in orange), but not for the non-extensible list (AESEV, in blue).

IDCTNAMES	LISTNAME	NCILIST-CODE	LISTDESCRIPTION	EXTENSIBLE	LISTCUSTOM	LISTSYNONYM	LISTVERSION	NCITYPE	LISTNCIPREFTERM	LISTPREFTERM	FMTNAME
4	AESEV	C66769	A scale that defines the degree or state of disease existing in a patient as a result of the occurrence of an adverse event. (NCI)	N		Severity/ Intensity Scale for Adverse Events			CDISC SDTM Severity Inten-sity Scale for Adverse Event Terminology	AESEV	$AESEV
71	REL		Relationship	Y	1						

The CTNames table stores list-level metadata, including list name, NCI code, and whether it is extensible. Controlled terminology is linked to the Variables table by IDCTNames.

IDCTVALUES	IDCTNAMES	CVALVALUE	CVALCODE	CVALSORTORDER	CVALCUSTOM	NCICODE	CVALDEFINITION	CVALNCIPREFTERM	CVALPREFTERM
21	4	SEVERE		3	0	C41340	A type of adverse event that interrupts usual activities of daily living, or significantly affects clinical status, or may require intensive thera-peutic intervention.	Severe Adverse Event	SEVERE
20	4	MODERATE		2	0	C41339	A type of adverse event that is usually allevi-ated with additional specific therapeutic inter-vention. The event interferes with usual activi-ties of daily living, causing discomfort but poses no significant or permanent risk of harm to the research participant.	Moderate Adverse Event	MODERATE
19	4	MILD		1	0	C41338	A type of adverse event that is usually tran-sient and may require only minimal treatment or therapeutic intervention. The event does not generally interfere with usual activities of daily living.	Mild Adverse Event	MILD
18	71	UNKNOWN		6	1				
17	71	DEFINITE		5	1				
16	71	PROBABLE		4	1				
15	71	POSSIBLE		3	1				
14	71	UNLIKELY		2	1				
7	71	NOT RELATED		1	1				

The CTValues table stores all terms for all studies, whether from a standard source or added for a particular study. It is linked to the CTNames table by IDCTNames. In addition to the terms, it contains metadata from external sources (e.g., NCI) and default values for sort order and input code.

The structure of the table is one row per term per list—not per study.

IDCTUSED	IDSTUDY	IDCTVALUES	USEDCODE	USEDSORT
32	941	21		
31	941	20		
28	941	19		
30	941	18	99	0
33	941	17	4	
34	941	16	3	
35	941	15	2	
29	941	14	1	
27	941	7	0	

The CTUsed table (left) is the only controlled terminology table that references specific studies. It contains one row per term used per study—linking to the CTValues table by IDCTValues and to a study by IDStudy. It also has values of sort order and input code to override the default values.

Below is a view of the controlled terminology data from all three table for the AE example.

IDSTUDY	IDCTUSED	LISTNAME	CVALVALUE	CVALCODE	CVALSORTORDER	USEDCODE	USEDSORT	IDCTVALUES	NCILISTCODE	LISTVERSION	LISTCUSTOM	LISTDESCRIPTION	EXTENSIBLE	NCITYPE	LISTNCIPREFTERM	LISTPREFTERM	LISTSYNONYM	FMTNAME	NCICODE	CVALCUSTOM	CVALDEFINITION	CVALNCIPREFTERM	CVALPREFTERM
941	32	AESEV	SEVERE		3			21	C66769			A scale that defines the degree or state of disease existing....	N		CDISC SDTM Severity Intensity Scale for Adverse Event Terminology	AESEV	Severity/Intensity Scale for Adverse Events	$AESEV	C41340	0	A type of adverse event that interrupts usual activities....	Severe Adverse Event	SEVERE
941	31	AESEV	MODERATE		2			20	C66769			A scale that defines the degree or state of disease....	N		CDISC SDTM Severity Intensity Scale for Adverse Event Terminology	AESEV	Severity/Intensity Scale for Adverse Events	$AESEV	C41339	0	A type of adverse event that is usually alleviated with additional specific....	Moderate Adverse Event	MODERATE
941	28	AESEV	MILD		1			19	C66769			A scale that defines the degree or state of disease existing....	N		CDISC SDTM Severity Intensity Scale for Adverse Event Terminology	AESEV	Severity/Intensity Scale for Adverse Events	$AESEV	C41338	0	A type of adverse event that is usually transient and may....	Mild Adverse Event	MILD
941	30	REL	UNKNOWN		6	99	0	18			1	Relationship	Y							1			
941	33	REL	DEFINITE		5	4	5	17			1	Relationship	Y							1			
941	34	REL	PROBABLE		4	3	4	16			1	Relationship	Y							1			
941	35	REL	POSSIBLE		3	2	3	15			1	Relationship	Y							1			
941	29	REL	UNLIKELY		2	1	2	14			1	Relationship	Y							1			
941	27	REL	NOT RELATED		1	0		7			1	Relationship	Y							1			

CDISC Announcement Opens New Possibilities:

Define 2.0 Brings New Challenges and Opportunities.

In September, CDISC released a draft Define-XML 2.0 specification. According to CDISC, this draft is a major change in the way SDTM, SEND, and ADaM metadata are transmitted. Among the revisions are changes to how controlled terminology is represented. The chart to the right was created by CDISC and shows the revised controlled terminology codelist structure in define.xml.