



PhUSE US Connect 2019

Paper DS12

Let's Create Standard Value Level Metadata

David Fielding, PRA Health Sciences, Brantford, Canada

ABSTRACT

Value Level Metadata is an important part of the Define-XML that allows us to explain our study data. Currently there isn't any formal value level metadata, but it does exist in various forms. On the CDISC website under controlled terminology you will find Codetable Mapping Files. These files contain relationships between codelists. Codetable mapping files exist for CV, ECG, Oncology, SC, TS and VS. Questionnaires, Ratings and Scales have similar metadata included in their guides. Therapeutic Area User Guides (TAUG's) also have this type of information in their guides. Creating Codetable mappings provides the basis for standard value level metadata that will lead to a programmable define and standard data validation.

INTRODUCTION

CDISC Controlled Terminology provides a set of values that can be used for variables within datasets. It doesn't contain relationships across different codelists. Codetables provide the relationships across codelists. In the CDISC Controlled Terminology, TESTCD's and TEST's are provided as separate codelists. Using the concept codes and the submission values, a simple Codetable can be created to match up the TESTCD with the appropriate TEST.

CODETABLE

CDISC Codetable Mapping files can be found on the CDISC website (www.cdisc.org) under Controlled Terminology and expanding the "Codetable Mapping Files" section.

VS CODETABLE

The following example from the VS Codetable (2018-12-21) for VSTEST=Height and Temperature.

C-code (Concept Code)	Vital Signs Test Code (VSTESTCD) (codelist code = C66741)	Vital Signs Test Name (VSTEST) (codelist code = C67153)	C-code (Concept Code)	Units for Vital Signs Results (VSRESU) (codelist code = C66770)
C25347	HEIGHT	Height	C49668	cm
C25347	HEIGHT	Height	C48500	in
C25206	TEMP	Temperature	C44277	F
C25206	TEMP	Temperature	C42559	C

A typical study would collect the Height in centimetres or inches and Temperature in Celsius or Fahrenheit and allow the investigator to select the unit they used for the original results. To appropriately explain how the unit is being collected a subset codelist of VSRESU would need to be created for each test code.

The codelist 'Units for Vital Signs Results (HEIGHT)' would have values of cm and in; and a second codelist 'Units for Vital Signs Results (TEMP)' with values F and C.

VSORRESU would have value level metadata with where clauses for each VSTESTCD with the codelist subsets.

Variable	OID	Where Clause	Controlled Terms	Codelist Values
VSORRESU	VSORRESU.VSTESTCD_HEIGHT	VSTESTCD="HEIGHT"	VSRESU_HEIGHT	cm in
VSORRESU	VSORRESU.VSTESTCD_TEMP	VSTESTCD="TEMP"	VSRESU_TEMP	F C



ECG CODETABLE

The following table is a sample from the ECG Codetable (2015-09-25) file. There is a one -to-many relationship between the ECG result (EGSTRESC) and the ECG test code and test name.

C-code (Concept Code)	ECG Test Code (EGTESTCD) (codelist code = C71153)	ECG Test Name (EGTEST) (codelist code = C71152)	C-code (Concept Code)	ECG Result (EGSTRESC) (codelist code = C71150)
C111131	AVCOND	Atrioventricular Conduction	C111088	1ST DEGREE AV BLOCK
C111131	AVCOND	Atrioventricular Conduction	C62016	2ND DEGREE AV BLOCK
C111131	AVCOND	Atrioventricular Conduction	C111091	3RD DEGREE AV BLOCK
C111132	AXISVOLT	Axis and Voltage	C102628	EARLY R WAVE TRANSITION
C111132	AXISVOLT	Axis and Voltage	C71035	ELECTRICAL ALTERNANS
C111132	AXISVOLT	Axis and Voltage	C102701	INDETERMINATE QRS AXIS
C111280	MI	Myocardial Infarction	C71065	ACUTE ANTERIOR WALL MYOCARDIAL INFARCTION
C111280	MI	Myocardial Infarction	C102591	ACUTE ANTEROLATERAL WALL MYOCARDIAL INFARCTION
C111280	MI	Myocardial Infarction	C102592	ACUTE ANTEROSEPTAL WALL MYOCARDIAL INFARCTION
C111307	RHYNOS	Rhythm Not Otherwise Specified	C116130	ASYSTOLE
C111307	RHYNOS	Rhythm Not Otherwise Specified	C111120	BRADYCARDIA
C111363	STSTWUW	ST Segment, T wave, and U wave	C92228	BORDERLINE QTCB
C111363	STSTWUW	ST Segment, T wave, and U wave	C92229	BORDERLINE QTCF

This EGSTRESC codelist is often used in collecting the abnormality. This may be provided by a vendor or entered on the CRF. The test and test code are often not included; however, the result can be mapped to the test and test code.

Data collection (CRF) and vendor data is often set up in a way that only the abnormal result is captured. In the example below a picklist that includes the controlled terminology for EGSTRESC is used, but then needs to be mapped to EGTESTCD.



Sample ECG eCRF

What was the interpretation of the ECG?	☐ NORMAL ☐ ABNORMAL EGORRES WHERE EGTESTCD="INTP"
If Abnormal, what was the abnormality?	☐ 1ST DEGREE AV BLOCK ☐ 2ND DEGREE AV BLOCK ☐ 3RD DEGREE AV BLOCK EGORRES WHERE EGTESTCD ="AVCOND" ☐ EARLY R WAVE TRANSITION ☐ ELECTRICAL ALTERNANS ☐ INDETERMINATE QRS AXIS EGORRES WHERE EGTESTCD ="AXISVOLT" ☐ ACUTE ANTERIOR WALL MYOCARDIAL INFARCTION ☐ ACUTE ANTEROLATERAL WALL MYOCARDIAL INFARCTION ☐ ACUTE ANTEROSEPTAL WALL MYOCARDIAL INFARCTION EGORRES WHERE EGTESTCD ="MI" ☐ ASYSTOLE ☐ BRADYCARDIA EGORRES WHERE EGTESTCD ="RHYNOS" ☐ BORDERLINE QTCB ☐ BORDERLINE QTCF EGORRES WHERE EGTESTCD ="STSTWUW"

If the site enters 1ST DEGREE AV BLOCK, then using the Codetable the result can be mapped to EGTESTCD=AVCOND and EGTEST=Atrioventricular Conduction in the SDTM dataset.

Utilization of result to test and test code mapping during data acquisition could facilitate more complete data being transferred to SDTM datasets. Central vendors should use the ECG Test names and codes when providing results. CRF designs could be set up with subsets of results by test names.

VALUE LEVEL METADATA

The value level metadata for EGSTRESC will have a where clause for each ECG Test Code (EGTESTCD). To create the value level metadata a codelist subset will need to be created for each ECG test. In the sample below the subset codelists EGSTRESC_AVCOND, EGSTRESC_AXISVOLT, EGSTRESC_MI, EGSTRESC_RHYNOS, and EGSTRESC_STSTWUW were used.

Variable	OID	Where Clause	Controlled Terms
EGSTRESC	EGSTRESC.EGTESTCD_AVCOND	EGTESTCD="AVCOND"	EGSTRESC_AVCOND
EGSTRESC	EGSTRESC.EGTESTCD_AXISVOLT	EGTESTCD="AXISVOLT"	EGSTRESC_AXISVOLT
EGSTRESC	EGSTRESC.EGTESTCD_MI	EGTESTCD="MI"	EGSTRESC_MI
EGSTRESC	EGSTRESC.EGTESTCD_RHYNOS	EGTESTCD="RHYNOS"	EGSTRESC_RHYNOS
EGSTRESC	EGSTRESC.EGTESTCD_STSTWUW	EGTESTCD="STSTWUW"	EGSTRESC_STSTWUW



Here is how the define.xml would appear.

Value Level Metadata - EG [EGSTRESC]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format
EGSTRESC	EGTESTCD = "AVCOND" (Atrioventricular Conduction)	text	200	["1ST DEGREE AV BLOCK", "2ND DEGREE AV BLOCK", "3RD DEGREE AV BLOCK"] < ECG Result (AVCOND) >
EGSTRESC	EGTESTCD = "AXISVOLT" (Axis and Voltage)	text	200	["EARLY R WAVE TRANSITION", "ELECTRICAL ALTERNANS", "INDETERMINATE QRS AXIS"] < ECG Result (AXISVOLT) >
EGSTRESC	EGTESTCD = "MI" (Myocardial Infarction)	text	200	["ACUTE ANTERIOR WALL MYOCARDIAL INFARCTION", "ACUTE ANTEROLATERAL WALL MYOCARDIAL INFARCTION", "ACUTE ANTEROSEPTAL WALL MYOCARDIAL INFARCTION"] < ECG Result (MI) >
EGSTRESC	EGTESTCD = "RHYNOS" (Rhythm Not Otherwise Specified)	text	200	["ASYSTOLE", "BRADYCARDIA"] < ECG Result (RHYNOS) >
EGSTRESC	EGTESTCD = "STSTWUW" (ST Segment, T wave, and U wave)	text	200	["BORDERLINE QTCB", "BORDERLINE QTCF"] < ECG Result (STSTWUW) >

QUESTIONNAIRES, RATINGS AND SCALES

There is a lot of value level metadata in the questionnaires, ratings and scales guides. The codelists for the results (QSORRES/QSSTRESC) are usually not included in CDISC controlled terminology. To determine the values, we need to use the Questionnaire Supplements.

To create value level metadata for the results (QSSTRESC) a codelist is created for each set of responses. Then each QSCAT/QSTESTCD is used for a where clause to show the available responses for that question. This can be time consuming for a study with multiple questionnaires. Wouldn't it be nice when you purchased a questionnaire that you also received a value level metadata file?

The first step is to establish conventions for create where clauses and OID's. This will allow for standard creation and validation. Some considerations are the variables that are included and the order they appear. In a Findings domain Grouping Qualifiers (--CAT, --SCAT) followed by the Topic (--TESTCD) would be a logical choice.

A convention for QSSTRESC could be QSSTRESC.QSCAT_CDISC *Synonym for Category of Questionnaire.QSTESTCD_CDISC Submission Value for the question.*

Here are some examples of ECOG101 and CSS0501A and CSS0501B questions.

OID	Where Clause
QSSTRESC.QSCAT_ECOG1.QSTESTCD_ECOG101	QSCAT="ECOG" and QSTESTCD="ECOG101"
QSSTRESC.QSCAT_CSS05.QSTESTCD_CSS0501A	QSCAT=" C-SSRS ALREADY ENROLLED SUBJECTS" and QSTESTCD="CSS0501A"
QSSTRESC.QSCAT_CSS05.QSTESTCD_CSS0501B	QSCAT=" C-SSRS ALREADY ENROLLED SUBJECTS" and QSTESTCD="CSS0501B"



Here is an example of the Eastern Cooperative Oncology Group Performance Status (ECOG).

DOMAIN	QSTESTCD	QSTEST	QSCAT	QSORRES	QSSTRESC
QS	ECOG101	ECOG1- Performance Status	ECOG	Fully active, able to carry on all pre-disease performance without restriction	0
QS	ECOG101	ECOG1- Performance Status	ECOG	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work	1
QS	ECOG101	ECOG1- Performance Status	ECOG	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours	2
QS	ECOG101	ECOG1- Performance Status	ECOG	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours	3
QS	ECOG101	ECOG1- Performance Status	ECOG	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair	4
QS	ECOG101	ECOG1- Performance Status	ECOG	Dead	5

The table shows the possible rows that would be in our study data. The values of 0-5 are the possible standard result for the performance status question. Data acquisition would be set up using a pick list for the possible responses in QSORRES.

The following define.xml shows how this is represented in value level metadata. The where clause starts with the category followed by the test code as QSCAT=ECOG and QSTESTCD=ECOG101. A subset codelist ECOG1-Performance Status is created to show the values 0-5 and their display value as the original result.

Value level metadata is also created for QSTEST and QSTESTCD to show the different questions available for each questionnaire (QSCAT=ECOG).

One of the challenges with the questionnaire, rating and scales test codes and tests is locating them in the current structure of the controlled terminology file. To display the QSTEST value in brackets beside the QSTESTCD value in the where clause your stylesheet may require you to also create codelists for QSTEST and QSTESTCD. QSTESTCD="ECOG" (ECOG1 – Performance Status) is shown, but without those codelists then it would appear as just QSTESTCD="ECOG".



Value Level Metadata - QS [QSSTRESC]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
QSSTRESC	QSCAT = "ECOG" and QSTESTCD = "ECOG101" (ECOG1-Performance Status)	text	200	ECOG1-Performance Status	eDT	

Value Level Metadata - QS [QSTEST]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
QSTEST	QSCAT = "ECOG"	text	200	["ECOG1-Performance Status"] <Eastern Cooperative Oncology Group Performance Status Questionnaire Test Name>	eDT	

Value Level Metadata - QS [QSTESTCD]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
QSTESTCD	QSCAT = "ECOG"	text	200	["ECOG101" = "ECOG1-Performance Status"] <Eastern Cooperative Oncology Group Performance Status Questionnaire Test Code>	eDT	

Go to the [top](#) of the define.xml

Controlled Terms

Eastern Cooperative Oncology Group Performance Status Questionnaire Test Code [CL.ECOG1TC, C101815]

Permitted Value (Code)	Display Value (Decode)
ECOG101 [C102408]	ECOG1-Performance Status

Eastern Cooperative Oncology Group Performance Status Questionnaire Test Name [CL.ECOGTTN, C101816]

Permitted Value (Code)
ECOG1-Performance Status [C102408]

Category of Questionnaire [CL.QSCAT, C100129]

Permitted Value (Code)
ECOG [C102116]

ECOG1-Performance Status [CL.QSSTRESC, ECOG101]

Permitted Value (Code)	Display Value (Decode)
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
5	Dead

Validation of metadata should check that a codelist with values 0-5 was created for QSSTRESC for the where clause QSCAT=ECOG and QSTESTCD=ECOG101. The data should be validated against the value level metadata that the data matches the one of the records in the table above.

Questionnaires, ratings and scales also have value level metadata for associated qualifier and timing information that should also be included as part of the validation, including QEVAL (Evaluator), QSEVLINT (Evaluation Interval), and QSEVINTX (Evaluation Interval Text).

CREATING STANDARD VALUE LEVEL METADATA

Creating standard value level metadata starts with the Data Acquisition and is passed on to SDTM.

DATA ACQUISITION

Data should be collected in a way that subset codelists are automatically associated with the result and easily passed on to SDTM metadata.

- Consider the naming conventions of the codelist. Store both the original and standard results in the codelist metadata and in the data.



- Create categories and test names or codes as part of the metadata or data. For example, create QSCAT as a pre-populated value that can be hidden to data entry.
- Create reusable questionnaires by establishing a standard naming convention for form identifiers, suggest using 'QS' followed by the CDISC Synonym value for example QSCSS01, QSGCGI01, or QSECOG1.

TABULATION

To create a complete library of value level metadata there are a lot of sources to consider.

- Establish rules for creating where clauses and then create value level metadata. Incorporate the value level metadata into validation of the define-xml and datasets.
- Using the Questionnaires, Ratings and Scales supplements create codelist subsets for the results (QSORRES/QSSTRESC). Create Codetables for the questions, results and other qualifier and timing variables.
- In the TAUG's, identify and create the subset codelists using the latest controlled terminology and then create Codetables from the information provided in the TAUG's.
- Use existing Codetables to create standard value level metadata that can be used in the define.xml.

STANDARD CHALLENGES

The Codetable mappings are not published at the same time as CDISC Controlled Terminology. The ECG Mapping was created using 2015-09-25 controlled terminology. The 2018-09-28 version has 5 more additional terms and one different value. It is important that available Codetable Mappings are maintained with the release of new controlled terminology. Most of the Codetables were updated with the release of the 2018-12-21 release of controlled terminology.

Define-XML version 2 allows us to use value level metadata to show the available responses for a specific where clause. It is important the available responses conform to controlled terminology in the Define-XML document. This involves the creation of codelist subsets from controlled terminology. There are a few subsets in the latest controlled terminology, unfortunately maintenance of a set of sponsor's standards also requires maintenance of subset codelists for each quarterly release of CDISC Controlled Terminology.

Value level metadata allows for other comparators then just equals. The IN comparator can be used to group test codes with the same controlled terminology. For example, a questionnaire could have many Yes/No response type questions and instead of having multiple records for value level metadata, an IN statement would be more appropriate.

CONCLUSION

A database implementation for CDISC Controlled Terminology may ease maintenance of updates and associated subsets as controlled terminology continues to expand and become more specific.

As CDISC develops Questionnaires, Ratings and Scales; more TAUG's and more Codetable mappings, value level metadata should also be included in a way that allows for easy implementation using SHARE. They should be up to date and maintained with the latest releases of controlled terminology. The value level metadata should include standard naming conventions for OID's and display.

Vendors who provide validation software should include them when validating the Define-xml and datasets.

REFERENCES

The National Cancer Institute (NCI) Terminology: <https://www.cdisc.org/standards/semantics/terminology>
Clinical Data Interchange Standards Consortium (CDISC): <https://www.cdisc.org/>

ACKNOWLEDGMENTS

I would like to express my special thanks to Kent Letourneau and Craig Mistal for their valuable suggestions.



RECOMMENDED READING

Anthony Chow, Erin Muhlbradt, Donnas Sattler (2018) CDISC Public Webinar: Controlled Terminology Mapping/Alignment Across Codelists.

CONTACT INFORMATION

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

Author Name	David Fielding
Company	PRA Health Sciences
City / Postcode	Brantford, ON, Canada N3P 1T7
Work Phone:	(250) 483-4477
Email:	fieldingdavid@prahs.com
Web:	www.prahs.com

Brand and product names are trademarks of their respective companies.