Define-XML Office Hours

Sam Hume, VP, Data Science, CDISC
Lex Jansen, Senior Director, Data Science Development, CDISC
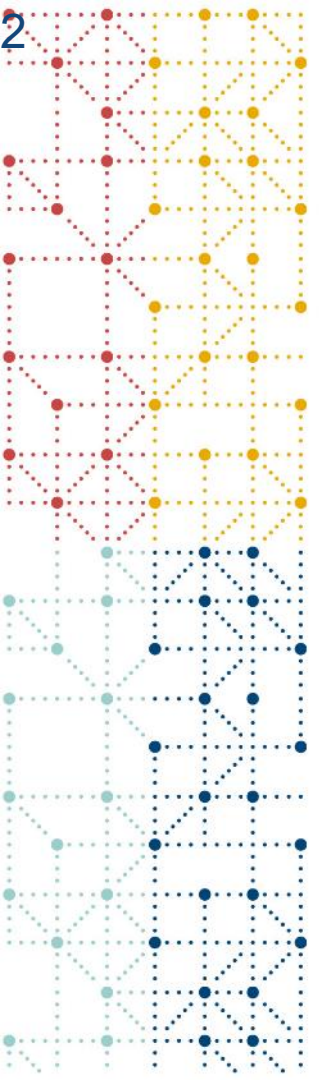Jozef Aerts, Director, XML 4 Pharma
Marcelina Hungria, Managing Director/Owner, DIcore Group
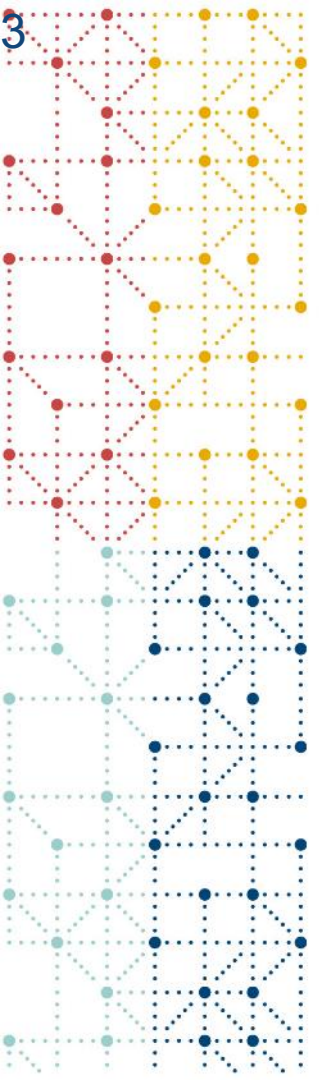Dmitry Kolosov, Expert Statistical Programmer, Parexel

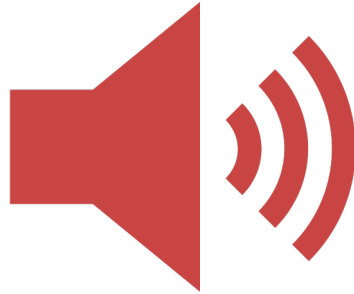**cdisc**

TUE 17 FEB 2022
11:00AM-12:30PM ET

Today's Agenda

1. Housekeeping
2. Speaker Introductions
3. Feature Presentation
4. Upcoming Learning Opportunities & Events
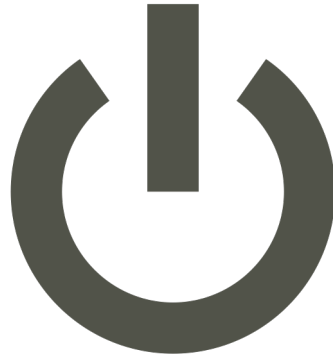
Housekeeping

Housekeeping

You will remain on **mute**

Housekeeping

Submit questions at any time via the Questions tool on your GTW app

Housekeeping

Audio issues?

Shut down & restart GoToWebinar app

cdisc

Housekeeping

A recording of this webinar and the slides will be available in the **Members Only** section of CDISC website

Today's Presenters

Sam Hume

VP, Data Science
CDISC

Lex Jansen

Senior Director, Data Science
Development CDISC

Jozef Aerts

Director
XML 4 Pharma

Marcelina Hungria

Managing Director/Owner
DIcore Group

Dmitry Kolosov

Expert Statistical Programmer
Parexel

Define-XML Office Hours

Sam Hume, VP, Data Science, CDISC
Lex Jansen, Senior Director, Data Science Development, CDISC
Jozef Aerts, Director, XML 4 Pharma
Marcelina Hungria, Managing Director/Owner, DIcore Group
Dmitry Kolosov, Expert Statistical Programmer, Parexel

cdisc

TUE 17 FEB 2022
11:00AM-12:30PM ET

Define-XML Office Hours webinar – 17 February 2022

- Sam Hume, VP, Data Science, CDISC

- Lex Jansen, Senior Director, Data Science Development, CDISC

- Jozef Aerts, Director, XML 4 Pharma

- Marcelina Hungria, Managing Director/Owner, DIcore Group

- Dmitry Kolosov, Expert Statistical Programmer, Parexel

cdisc

Define-XML Office Hours webinar

1. Would like to learn how to create define Xml, any links?

2. Which one is the best approach to create define when having multiple annotated CRFs? Should we only have 1 CRF (page links)

3. xxCAT were assigned on the CRF not collected from Rave data. Is it fine to for origin be Crf Page xx.

4. What length should I report in define-xml for character variables? Actual length or maximum length based on documentation.

5. I have a dataset in my submission for which intended data were to be collected, but none actually were. How do I handle this situation in my Define-XML file? Do I still include the dataset definition?

6. Define-XML has a "Mandatory" attribute for variables. How does this relate to the Core attribute that I see in ADaM, SDTM and SEND Implementation Guides?

7. Should I use the XLS stylesheet that was published as part of the Define-XML publication?

8. I am trying to open my Define-XML file in Chrome, but it does not open. How can I solve this?

9. When to create Value Level Metadata?

10. Do I need to create a codelist for all variables with a finite number of values?

cdisc

Would like to learn how to create define.xml, any links?

- Training, Webinars, Conferences:
  - CDISC offers many possibilities
- Understanding the Define-XML Specification
  - CDISC Standards and Control Terminologies
- Other documentation
  - PhUSE Working Groups Deliverables
- Regulatory Agencies
  - Guidelines
- Recommended creating metadata definitions at front
  - Use of metadata repository
    - Improves standardization on sponsor conventions
  - Ready for study startup
  - Metadata definitions need maintenance

cdisc

Would like to learn how to create define Xml, any links?

- CDISC training: https://www.cdisc.org/education

- Define-XML Specification: https://www.cdisc.org/standards/data-exchange/define-xml

- SDTM Metadata Submission Guidelines v2.0
  https://www.cdisc.org/standards/foundational/sdtm/sdtm-metadata-submission-guidelines-v2-0

- Define-XML Version 2.0 Completion Guidelines:
  https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Define-XML+Version+2.0+Completion+Guidelines.pdf
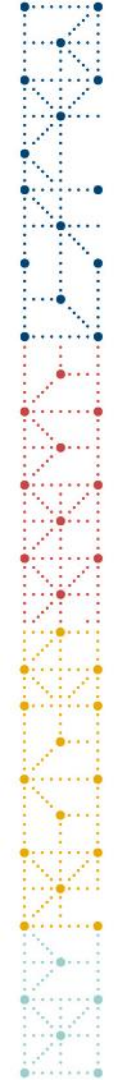
- Define-XML stylesheets: https://wiki.cdisc.org/display/PUB/Stylesheet+Library

cdisc

Which one is the best approach to create define when having multiple annotated CRFs? Should we only have 1 CRF (page links)

Define-XML supports Multiple Annotated CRFs, but if you submit to – for example – the FDA, it can only be one Annotated CRF with name "acrf.pdf".

xxCAT were assigned on the CRF not collected from Rave data. Is it fine to for origin be Crf Page xx.

All of the types below allow references to CRF pages, if it makes sense.
For "Collected" it is required.

Collected	A value that is actually observed and recorded by a person or obtained by an instrument. Note that a collected entry translated to a synonymous controlled term still has a type Collected.
Derived	A value that is calculated by an algorithm or reproducible rule, and which is dependent upon other data values, including data values available within the dataset or externally provided data values. MethodDef must be used to document the algorithm or rule used for a derived value.
Assigned	Data that is either: • Determined by individual judgment as provided by an evaluator, or • Coded terms supplied as part of a coding process, or • Values set independently of any subject-related data value in order to complete a dataset.
Protocol	Data that is defined as part of the study protocol, investigator instructions, standard operating procedures or trial design preparation
Predecessor	An entry that is copied from a variable in another dataset. The Description child element identifies the dataset and variable that is copied.
Not Available	Used when the origin is not available and cannot be determined. Sponsors should specify additional details that may be helpful to the reviewer in the Comments section of the data definition file.

cdisc

Table copied from the CDISC Define-XML Specification 2.1 (Final)

What length should I report in define-xml for character variables? Actual length or maximum length based on documentation.

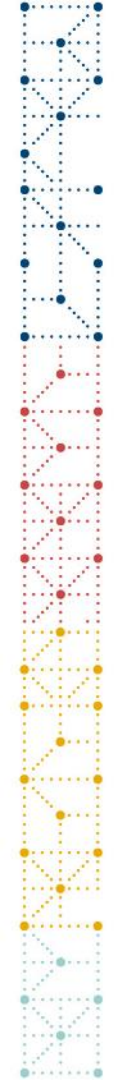- Depends on the context a Define-XML document is created for.
  - Is it for Sponsor – CROs metadata and data exchange?
    - Used for database setup? E.g., maximum length expected in text data values; variable need to hold all potential values from associated codelist
    - Used for data aggregation?
    - Validation rules may need to be considered; e.g., detecting issues from different systems in data transfers

  - Is it for submission to regulatory authorities?
    - Need to follow current guidelines

cdisc

What length should I report in define-xml for character variables? Actual length or maximum length based on documentation.

Define-XML has always been about planned metadata definitions.

In the FDA Study Data Technical Conformance Guide (Sept 2021)* we read in section 3.3.3:
"The allotted length for each column containing character (text) data should be set to the maximum length of the variable used across all datasets in the study except for suppqual datasets. For suppqual datasets, the allotted length for each column containing character (text) data should be set to the maximum length of the variable used in the individual dataset. This will significantly reduce file sizes. For example, if USUBJID has a maximum length of 18, the USUBJID's column size should be set to 18, not 200."

This is a statement about the length in the data, not the Define-XML

*) https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

cdisc

I have a dataset in my submission for which intended data were to be collected, but none actually were. How do I handle this situation in my Define-XML file? Do I still include the dataset definition?

Empty datasets are not going into a submission. Whether to include it in Define-XML documents, depends on the version of Define-XML standard:

- For Define-XML v2.0 empty datasets should not be included in define.xml

- In Define-XML 2.1 a special attribute was added to address such case: **HasNoData**. It can be used to indicate that there is no data for a dataset or a particular variable.
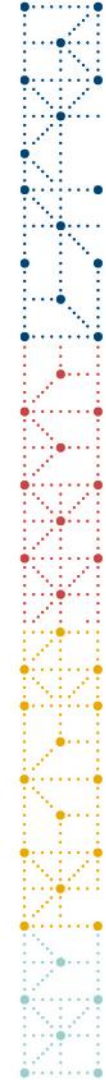
cdisc

Define-XML has a "Mandatory" attribute for variables. How does this relate to the Core attribute that I see in ADaM, SDTM and SEND Implementation Guides?

A general purpose of the **Mandatory** flag is to indicate whether a variable can be blank or not. When a variable has the **Mandatory** flag set to "Yes", then it must not have null (blank) values.

In ADaM there is no direct relationship between values of the Core attribute and emptiness of a variable (e.g., BDS AVAL). In SDTM, value of the Core attribute impacts whether a variable can be blank or not:
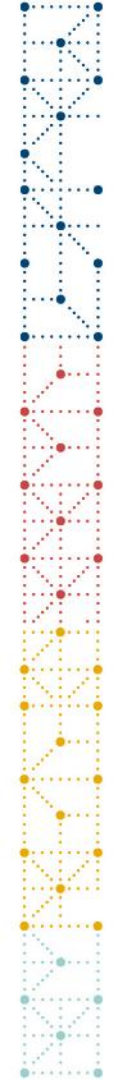
- SDTM variables with "Require" Core value cannot be blank, therefor Mandatory must be set to "Yes" for those variables.

- SDTM variables with "Expected" or "Permissible" Core values can be blank, therefor Mandatory should be set to "No" for those variables. If the sponsor requires that any of these variables cannot be blank, then the Mandatory flag should be set to "Yes".

cdisc

Should I use the XLS stylesheet that was published as part of the Define-XML publication?

The Define-XML standard does not dictate how a stylesheet should display a Define-XML document. An example stylesheet is provided with the CDISC Define-XML publication package; this can be altered to satisfy alternate visualization needs.

A PhUSE working group published an updated Define-XML v2.0 stylesheet in 2018:
https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Define-XML+v2.0+Stylesheet+Recommendations.zip

cdisc

I am trying to open my Define-XML file in Chrome, but it does not open. How can I solve this?

Modern browsers have strict security settings and one of the security policies blocks access to local files. This policy is enabled in the current versions of such browsers as Chrome, Firefox, Edge. Ability to render Define-XML document with a stylesheet requires that access. A solution is to enable it when for reviewing Define-XML. The approach below is explained for Windows, similar method can be used in Linux/MacOS.

cdisc

Chrome

Define-XML Viewer

1. Create a new link for the Chrome browser. It is suggested to name it differently, e.g., **Define-XML Viewer**.

2. Edit Link details and in the **Target** field add **--allow-file-access-from-files** at the end. If it does not work you can try **--disable-web-security --user-data-dir=c:/temp** parameters (c:/temp needs to be an actual folder).

3. Save the changes. After that you can drag and drop define.xml files on the Chrome browser window.

Use this link only to open trusted versions of Define-XML and stylesheet files.

Target type:	Application
Target location:	Application
Target:	plication\chrome.exe" --allow-file-access-from-files
Start in:	"C:\Program Files\Google\Chrome\Application"
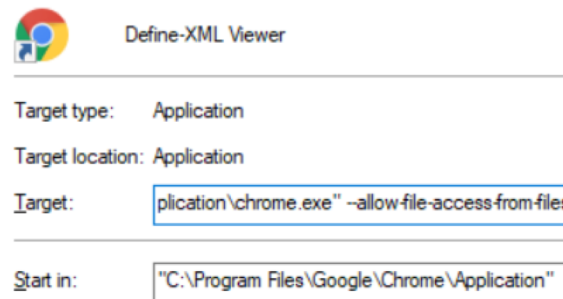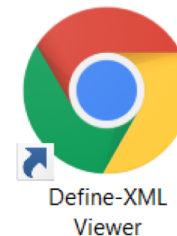
Firefox

With Firefox, the process is more complicated.

1. Create a new link for the Firefox browser, e.g., **Define-XML Viewer**.

2. Edit Link details and in the **Target** field add **-ProfileManager** at the end.

3. Open it and create a new profile for Define-XML. Uncheck "Use the selected profile without asking at startup".

4. Click Start Firefox and in the address bar, open the following URL: **about:config**

5. Search for **privacy.file_unique_origin** and set it to False.

Use this link only to open trusted versions of Define-XML and stylesheet files.
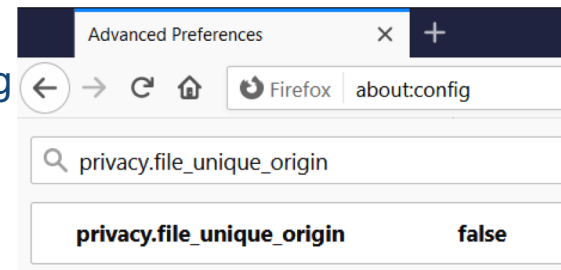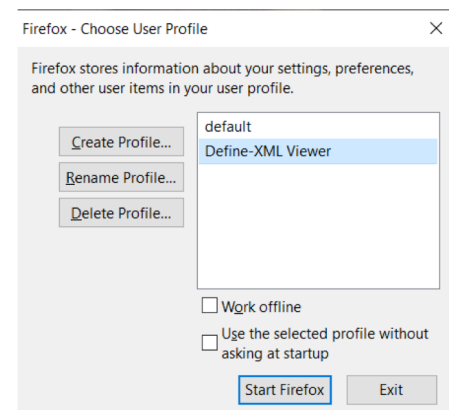
I am trying to open my Define-XML file in Chrome, but it does not open. How can I solve this?

Users may not be comfortable relaxing the security settings in a browser.

Another solution would be to use a XSLT processor to create an HTML file from the Define-XML file and the XSL stylesheet. This HTML file can then be opened in any browser with issues.

Languages like SAS, Python, Java, R, PowerShell, and many other can transform XML to HTML by using an XSL stylesheet.

When to create Value Level Metadata?

The normalized data structure used by datasets based on the SDTM, SEND, and ADaM (generally 1 record per subject per test code per visit or observation) provides an efficient method for transmitting information. However, there are cases where the dataset variable metadata does not provide sufficient detail to support data review and analysis. In these cases, value-level metadata should be provided in the Define-XML document. Value-level metadata enables the specification of the metadata of a variable under conditions involving 1 or more dataset variables. The definition of a variable for a specific condition is known as value-level metadata.

- It is most often used within SDTM Findings domains to provide definitions for Variables (e.g., --ORRES, --ORRESU, --STRES, --STRESU) that are specific to each test code (value of --TESTCD). It is not required for Findings domains where the results have the same characteristics in all records (e.g., Inclusion Exclusion domains).

- In ADaM, value-level metadata often describes AVAL or AVALC in BDS data based on values of PARAMCD, but can be applied in other cases where useful.

cdisc

When to create Value Level Metadata?
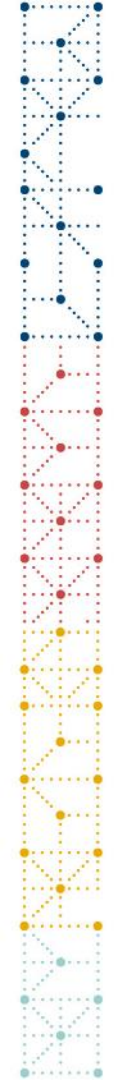
Value-level metadata can also be used on all classes of SDTM domains. For example, in a Dispositions domain, value-level metadata might define the codelists for DSTERM and DSDECOD for each value of DSCAT.

Value-level metadata for a SDTM SUPPQUAL dataset can be used to include one or more Supplemental Qualifier variables in the Domain key.

Do I need to create a codelist for all variables with a finite number of values?

- Not necessarily
    - Need to understand the meaning of the finite number of values; e.g., AGE restricted to a range in the associated Age units may have a restriction (edit check) on collection; however, it does not represent a categorical finite set of values.

- Refer to the Define-XML Completion Guidelines, PhUSE WG deliverable: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Define-XML+Version+2.0+Completion+Guidelines.pdf

cdisc

Can Origin = "Predecessor" be used for integrated SDTMs that have been created by stacking individual SDTMs?

Yes, Origin = "Predecessor" is not restricted to the use case in ADaM. If you have other use cases for using this origin it is certainly allowed to be used.

How to represent different Codelists for QSCAT in Valuelevel?

Different codelists can be defined in Value Level Metadata based on where clauses that are defined based on the value of QSCAT.

The SDTM Metadata Submission Guidelines v2.0 contain an example of different codelists based on Value Level Metadata based on QSTESTCD. The same principles can be used for conditions based on other variables.

What is the maximum length of the "Source/Derivation/Comment" column?

The "Source/Derivation/Comment" is a column in the HTML table as created by the XSL stylesheet. It combines different metadata elements in one table column. Define-XML does not restrict the length of these metadata elements.

When the length of derivations or comments becomes very long, or requires advanced formatting, the user should consider to link to a separate PDF document instead.

cdisc
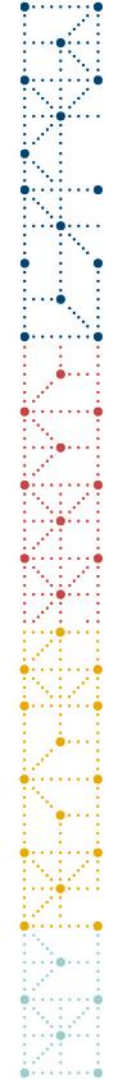
In my organisation we're wondering about BIMO Define-XML and if there is/will be an associated style sheet? Any advice on this?

Define-XML can be used to described any tabular data format. It is not restricted to ADaM, SDTM, or SEND datasets.

We are not aware of special BIMO requirements for the associated stylesheet. For this reason, we recommend using the latest stylesheet published by PHUSE. (https://wiki.cdisc.org/display/PUB/Stylesheet+Library)
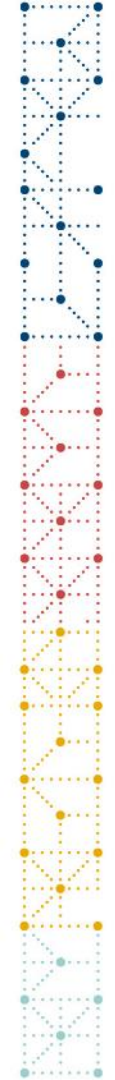
Since the Define-XML standard does not dictate how a stylesheet should display a Define-XML document, stylesheets can be altered to satisfy alternate visualization needs.

When creating define-xml for ADaMs, do I have to add to codelist those variables which are coming from SDTM as predecessor?

Yes, variables which are coming from SDTM as predecessor should have their codelists added to the ADaM Define-XML document. In this way reviewers do not have to open the separate SDTM Define-XML document and search for a specific codelist.
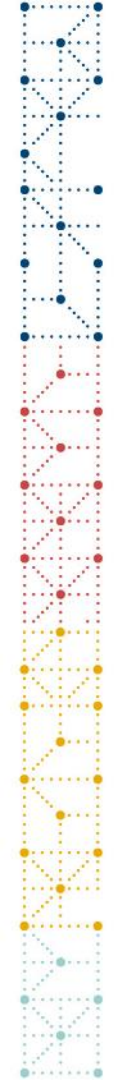
There is no defined metadata in Define-XML that links from a variable in one Define-XML file to a variable in another Define-XML file.

Do we have to mention Define-XML Terminology Version also as one <def:Standard> element with Type="CT" and PublishingSet="Define-XML" ?

Yes, in Define-XML 2.1 it is recommended to specify the Terminology version used as a <def:Standard> element with type="CT" and PublishingSet="Define-XML".

In this way it is clear which version of the Controlled Terminology has been used for certain Define-XML attributes. This version is also directly linked to the Define-XML 2.1 schema version and the Define-XML 2.1 define-enumerations.xsd schema file.

What is the value of the Decode for test parameters code ? Does it have to be the "submission value" of the parameter name, the "NCI preferred term" or the "CDICS definition" ?

In where clause,   for example in TSPARMCD EQ DIET, do I have to write   TSPARMCD EQ DIET or is it mandatory to add the submission value of the param test name between parenthesis TSPARMCD EQ DIET (Basal Diet)
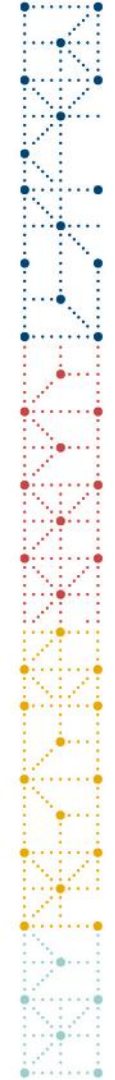
The submission value of the parameter test name is added by the XSL stylesheet between parenthesis in the where clause.

The submission value of the parameter test code and the corresponding submission value of the parameter test name have the same C-code as defined by the NCI Thesaurus.

cdisc
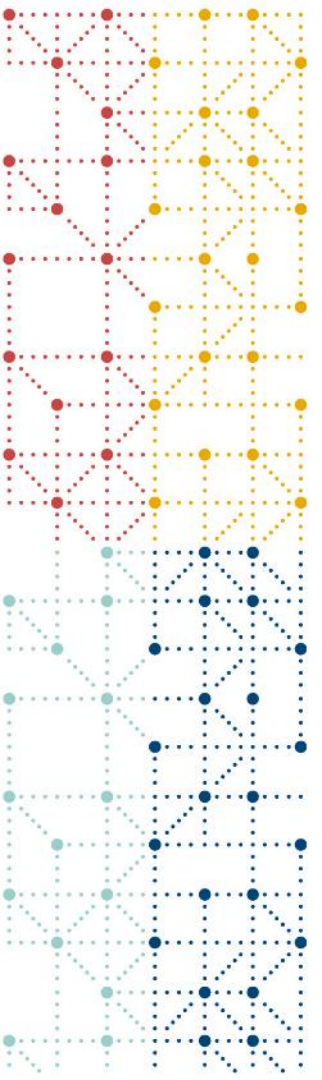
Origin and page number is required for all variables or not. And Also there is no update is required in where clause sheet while creating input excel sheet.

Origin is required for every variable. A reference to a specific location on an annotated CRF document is required when Origin Type="CRF" (Define-XML 2.0) or Origin Type="Collected" (Define-XML 2.1).

cdisc

How detailed should the methods be? I have seen algorithm/derivations very close to specifications and I have been told that it need only be very high level, general descriptions as opposed to instruction that even medical reviewers with no programming experience could understand.

The methods should be in clear English and should be detailed enough for a reviewer to exactly understand how the variable was derived.

Additional Audience Questions

Audience Questions

If a variable has Predecessor as origin and the origin variable has a codelist, should we create a codelist for this variable?

cdisc

Audience Questions

Could you provide some best practices on when to make Value Level Metadata and when explanations in the ADRG is sufficient?

Audience Questions

Which one is the best approach to create define when having multiple annotated CRFs? Should we only have 1 CRF?

cdisc

Audience Questions

If xxCAT is assigned on the CRF, not collected from Rave data, is it fine for the origin to be "CRF Page xx?"

cdisc

Audience Questions

What length should I report in define-xml for character variables? Actual length or maximum length based on documentation?

Audience Questions

Audience Questions

How detailed should the methods be?

I have seen algorithm/derivations very close to specifications and I have been told that it need only be very high level, general descriptions as opposed to instruction that even medical reviewers with no programming experience could understand.
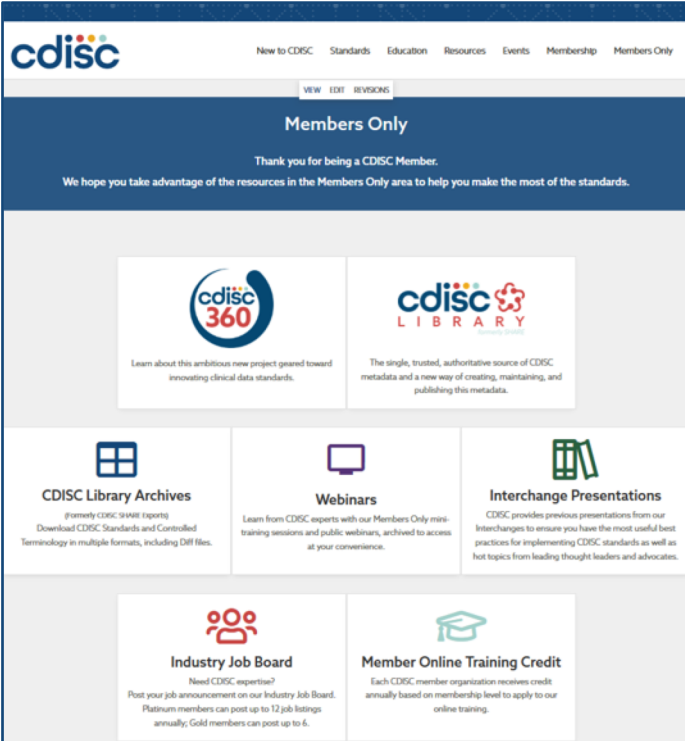
cdisc

Upcoming Learning Opportunities

Why Become a Member?

- To ensure the CDISC standards remain open and free

- To support CDISC in the development and maintenance of global standards

- To work with the CDISC community and be a voice in the development of clinical research standards

- To impact the development of regulatory requirements for submissions

- To access members only resources and benefits

- To gain visibility in the marketplace

CDISC MEMBERSHIP

Become a Member!

Join nearly 500 member organizations that contribute to bringing clarity to data.

Already a Member?

Thank you! It is our members' support which enables us to develop standards, keeping it free and accessible to all.
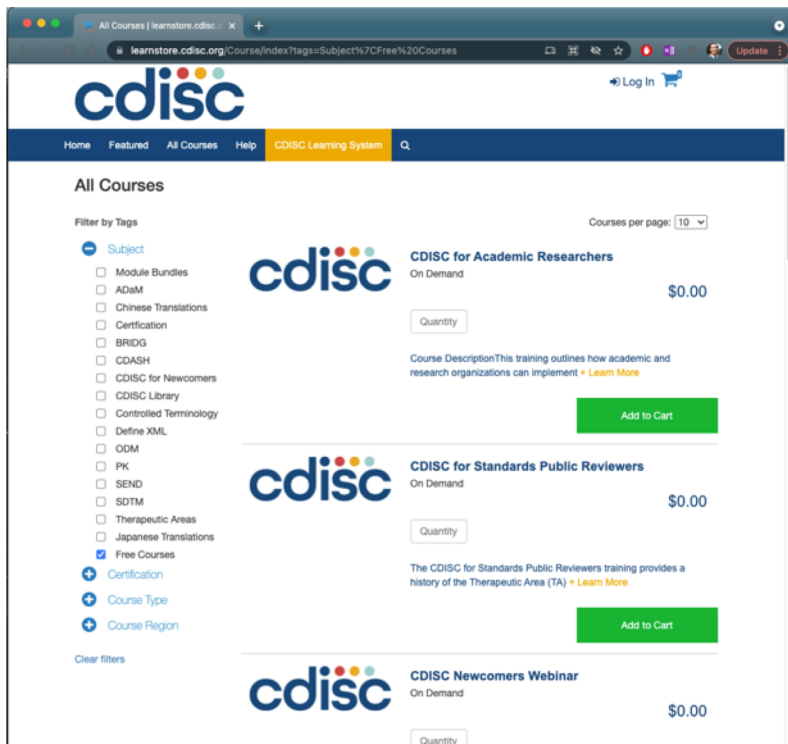
Email: membership@cdisc.org

Free CDISC Courses

Http://learnstore.cdisc.org

2022 EUROPE INTERCHANGE
CDISC VIRTUAL CONFERENCE
27-28 APRIL

2022 JAPAN INTERCHANGE
CDISC VIRTUAL CONFERENCE
13-15 JUNE

**New dates: 13-14 JUNE

2022 Japan Interchange Call for Abstracts Ends Tomorrow!

Friday, 18 FEB is your last chance to submit an abstract for the 2022 Japan Interchange – don't let the deadline pass you up!

To submit, go to CDISC.org and find the Call for Abstracts form under the Japan Interchange page.

2022
CHINA INTERCHANGE
29 - 30 JULY | BEIJING
WITH STANDARDS – UNLOCK THE POWER OF DATA

2022
US INTERCHANGE
26-27 OCTOBER | AUSTIN, TX
WITH STANDARDS – UNLOCK THE POWER OF DATA

cdisc

Upcoming Webinars

Date	Title
22 FEB	CDISC Education: What's New in 2022
15 MAR	Controlled Terminology Office Hours
22 MAR	Digital Data Flow (DDF) Project Public Review
24 MAR	Introduction to the SDTM Genomics Findings (GF) Domain
29 MAR	SDTM Office Hours
5 APR	Controlled Terminology Updates for Q1 2022 – P49 Publication + P50 Public Rev.
19 APR	QRS Office Hours

cdisc

Webinar Page Updates

New Webinar Archive

Webinar Topic Request Form

Do you have ideas for great webinar topics?

Is there something you would like us to cover?

We want to hear from you!

https://www.cdisc.org/form/webinartopicreq

cdisc

Thank you!

Contact the Events inbox:
events@cdisc.org

Contact Education inbox:
training@cdisc.org

Contact Bernard directly:
bklinke@cdisc.org

cdisc