

Not Another Other: Bridging Data Collection with SDTM Mapping Without Open Text Fields

David Fielding, ICON plc, Brantford, ON, Canada
Susana Jimenez, ICON plc, Austin, TX, USA

ABSTRACT

Controlled Terminology is essential for ensuring consistency, accuracy, and interoperability in clinical data. The persistent use of free-text fields such as “Other, specify” undermines these goals, introducing ambiguity and complicating downstream mapping to SDTM. This paper advocates the elimination of unnecessary free-text entries in favour of structured data collection aligned with Controlled Terminology. We will demonstrate how existing Electronic Data Capture (EDC) system functionality can be leveraged to guide Investigators toward valid selections that bridge data collection with SDTM mapping. We present a practical methodology for designing case report forms (CRFs) that promote standardized input, demonstrate mapping strategies to SDTM, and explore techniques for representing extensive codelists within Define-XML. We will show how to improve data quality, streamline standardization, and enhance regulatory readiness by minimizing the use of “Other, specify” whenever possible.

INTRODUCTION

Controlled Terminology has expanded dramatically over the past decade, now encompassing more than 45,000 submission values and, when including synonyms, over 78,000 terms. Common data collection domains—such as Concomitant Medications—often include fields like Frequency, Route, and Unit, all of which rely on Controlled Terminology. Traditionally, data collection designs use a limited subset of these values based on preference or expectations, supplemented by free-text fields for entries outside the predefined list. This approach introduces inefficiencies and additional coding processes.

To improve data quality and streamline workflows, we need to leverage the full breadth of Controlled Terminology and minimize reliance on free-text fields. In this paper, we will explore potential solutions and provide recommendations spanning Data Collection, SDTM, and Define-XML.

IMPLEMENTATION

DATA COLLECTION

Codelists can be setup with the entire list from CDISC CT to leverage EDC capabilities to search using the format of “submission value [synonyms]”. Taking advantage of EDC systems that use type-ahead filtering or autocomplete functionality, a long list of entries can be narrowed down to the entry needed. You may be familiar with similar functionality with filters in Microsoft Excel. Among the synonyms, the NCI Preferred Term can also be included for a full view of all options available.

As an example, the Frequency term for the code “1 TIME PER WEEK” would have the decode “1 TIME PER WEEK [One Time Per Week; Once Weekly]”

CDISC Submission Value	CDISC Synonym(s)	NCI Preferred Term
1 TIME PER WEEK	One Time Per Week	Once Weekly

The data collection is set up in a way that satisfies data entry and passes through to SDTM, and the Define-XML is used to show all the possible values available in the study.

Let's look at an example from the CDISC eCRF Portal for Concomitant Medications
(<https://www.cdisc.org/kb/ecrf/concomitant-medications>) [1]

CONCOMITANT MEDICATIONS EXAMPLE

The “What is the Unit (for the dose of Concomitant [Medication/Treatment/Therapy])?” CMDOSU field is set up as a dropdown list of 16 common values, and then a free text “Other, specify” field. The two fields for collecting Unit need to be stored in one variable CMDOSU in SDTM. The Unit codelist in CDISC SDTM Controlled Terminology contains close to 2100 values and synonyms. These should be available to the Investigator to enter.

Codelist Name	Coded Value	Code Value Text (Investigator Visual)
LOC	ACROMION	Acromion
LOC	ACUTE MARGINAL ARTERY	Acute Marginal Artery [Acute Marginal Artery Segment(s); AMARG]

The Frequency (FREQ) codelist could also use this methodology. However, this is a codelist where the terminology permits the use of the same convention of existing terms. In the CDISC Controlled Terminology Requests Denied file [3] the request for “8 TIMES PER WEEK” was denied with the reason “The list is extensible”. Users should add new terms using the convention of existing terms, but the possible combinations are endless. The current terminology includes “1 TIME PER WEEK” to “7 TIMES PER WEEK”. Since there are endless combinations, we need an “Other, specify” but we need to put some control on allowable formats using validation rules.

DAYS PER MONTH
TIMES PER CYCLE
TIMES PER WEEK
TIMES PER YEAR
TIMES PER MONTH
TIMES PER DAY
EVERY # WEEKS
EVERY # YEARS
Q#H
Q#M
Q#D
Q#MIN

There still might be a need for an “Other, specify” field if it is important to the study to capture more complicated frequencies.

DEFINE-XML RESULT

The Define-XML is used to show all the possible values available for in the study. We can simply use the study data specification and show that all the terms in the CDISC Controlled Terminology were available. Below is a snapshot of examples above where the Display Value shows what was available for data entry.

Unit [C71620] [CDISC/NCI SDTM 2025-03-28]

Permitted Value (Code)	Display Value (Decode)
10 ¹⁰ /L [C105517]	10 ¹⁰ /L [10 ⁴ /mm ³ ; 10 ⁴ /uL; 10 ⁷ /mL]
10 ¹¹ /L [C105488]	10 ¹¹ /L [10 ⁵ /mm ³ ; 10 ⁵ /uL; 10 ⁸ /mL]
10 ¹² IU/L [C105518]	10 ¹² IU/L [Tera International Unit per Liter; TIU/L]
10 ¹² /L [C67308]	10 ¹² /L [/pL; 1/pL; 10 ⁶ /mm ³ ; 10 ⁶ /uL; M/uL; Mill/mcL; T/L; Tera/L; TI/L]
10 ³ CFU [C68895]	10 ³ CFU [Thousand CFU; Thousand Colony Forming Units]
10 ³ CFU/g [C68899]	10 ³ CFU/g [Thousand CFU/g; Thousand Colony Forming Units per Gram]
10 ⁶ /L [C67452]	10 ⁶ /L [/mm ³ ; /uL; 1/mm ³ ; 1/uL; 10 ³ /mL; M/L; Mega/L]

If we used an “Other, specify” free text field and coded this in an external process we should still be showing the entire list because everything was available for entry.

CHALLENGES

ELECTRONIC CASE REPORT FORM (eCRF)

The eCRF produced from the EDC system may have options to restrict the number of codelist items printed. Unfortunately, this doesn’t give the whole picture of what was available for entry. Some EDC systems will print all the values which lead to pages and pages of codelist values. Neither of these are very helpful. There needs to be consideration on how this can be represented by indicating the entire dictionary was used. For example, the text in red below could be included on the eCRF and the annotated case report form (acrf.pdf).

eCRF

Dose	
Dose Unit	[CDISC/NCI SDTM UNIT C71620]
Dose Form	[CDISC/NCI SDTM FRM C66726]
Frequency	[CDISC/NCI SDTM FREQ C77113]
Route	[CDISC/NCI SDTM ROUTE C66729]

Annotated CRF

Dose	CMDOSE	CMDOSTXT	
Dose Unit	CMDOSU		[CDISC/NCI SDTM UNIT C71620]
Dose Form	CMDOSFRM		[CDISC/NCI SDTM FRM C66726]
Frequency	CMDOSFRQ		[CDISC/NCI SDTM FREQ C77113]
Route	CMROUTE		[CDISC/NCI SDTM ROUTE C66729]

EDC SYSTEM LIMIT ON NUMBER OF VALUES

The set up of 2000 values is a lot more work than just 16 values. The EDC system may have restrictions on the number of values allowed in a codelist. If this is the case, split the field into multiple fields alphabetically.

EDC systems often have a process for uploading the code lists, and scripts can be written to convert the CDISC Controlled Terminology into the EDC system loader format to ease the burden. There is also the challenge of maintaining this with each new release of terminology.

Perhaps EDC systems should have built-in functionality to reference or upload CDISC Controlled Terminology.

EDC SYSTEM RECOGNITION OF CASING

Some values are case sensitive and there are EDC systems that don't have the ability to distinguish between cases in coded values so there may need to be some mapping required. Here is an example where PA is different from Pa.

Codelist Name	Coded Value	Coded Value Text
UNIT	PA	PA [/Year; Every Year; Per Annum; Per Year]
UNIT	Pa	Pa [Pascal]

DEFINE-XML FUTURE CONSIDERATIONS

If you think of how we encode medical terms using MedDRA or WHODrug, we do not include the list of all the possible lowest level terms or preferred terms that were available for use. We simply reference the external dictionary. We should be able to do the same thing with CDISC Controlled Terminology. If you did not use a subset, then just reference the codelist name.

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
CMDOSU	Dose Units	text	Variable Qualifier	200	Unit	Collected (Source: Investigator)

Reference Name	Dictionary	Dictionary Version	Codelist
Unit	CDISC/NCI SDTM 2025-03-28	2025-03-28	Unit [C71620]

DEFINE-XML REFERENCE OF EXTERNAL DICTIONARY

Define-XML only allows for the reference of an external data dictionary and not the codelist within that dictionary. For example, we usually just put MedDRA, but it would be more appropriate to reference the Lowest Level Term code list within the MedDRA dictionary. We could then set the data type properties for the list instead of for the entire dictionary reference.

BIOMEDICAL CONCEPTS

Let us consider all the great work being done with biomedical concepts and how we can leverage this information. In the CDISC SDTM Dataset Specializations it includes subset codelists and value lists. Here is an example of EGTESTCD PRAG for PR Interval Aggregate.

vlm_group_id	short_name	sdtm_variable	dec_id	codelist	codelist_submission_value	subset_codelist	value_list
PRAG	PR Interval Aggregate	EGORRESU	C42574	C71620	UNIT		ms;s

*Not all columns shown.

The value list has ms and s, so we could have a subset codelist UNIT_PRAG that could always be used for this test. An interesting code is C42574, this is a “Unit of Time” codelist which in itself is a subset of the UNIT codelist for measurements of time. (<https://evsexplore.semantics.cancer.gov/evsexplore/concept/ncit/C42574>)[4]

In the future, how do we simply load these into the EDC system and present them to reviewers in Define-XML. The CDISC 360i project is putting us on the right path. As more biomedical concepts are created there is a clear direction from data collection to SDTM.

CONCLUSION

Relying on free-text “Other, specify” fields is no longer necessary—or desirable—in a world with comprehensive Controlled Terminology and modern EDC functionality. By fully exposing CDISC Controlled Terminology to site users through intelligent search and filtering, we can capture structured data at the point of entry. This will reduce downstream coding burden, SDTM programming, and enhance SDTM compatibility. Although challenges remain, especially around large lists and system limitations, thoughtful design and automation can mitigate these issues. This shift not only improves data quality and standardization but also supports clearer, more efficient regulatory review.

REFERENCES

1. CDISC. Concomitant Medications [Internet]. CDISC eCRF Portal. [cited 2026 Jan 30]. Available from: <https://www.cdisc.org/kb/ecrf/concomitant-medications>
2. CDISC. Controlled Terminology [Internet]. [cited 2026 Jan 30]. Available from: <https://www.cdisc.org/standards/terminology/controlled-terminology>
3. CDISC. Controlled Terminology CT Requests Denied (P60) [Internet]. [cited 2026 Jan 30]. Available from: https://www.cdisc.org/system/files/members/standard/terminology/Controlled_Terminology_Requests_Denied_P60.xlsx
4. National Cancer Institute. EVS Explorer, Code: C42574 [Internet]. Enterprise Vocabulary Services. [cited 2026 Jan 30]. Available from: <https://evsexplore.semantics.cancer.gov/evsexplore/concept/ncit/C42574>

ACKNOWLEDGMENTS

We would like to acknowledge Kent Letourneau for his ongoing support and leadership in data standards, which greatly contributed to the development of this work. We also thank Maurits Meijboom and Rebecca Lakey for reviewing the paper and providing valuable feedback.

RECOMMENDED READING

CDISC Controlled Terminology FAQ: (<https://www.cdisc.org/kb/articles/controlled-terminology-faqs>)

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the authors at:

Author Name: David Fielding

Company: ICON plc

Address: Brantford, ON, Canada

Email: david.fielding@iconplc.com

Website: iconplc.com

Author Name: Susana Jimenez
Company: ICON plc
Address: Austin, TX, USA
Email: susana.jimenez@iconplc.com
Website: iconplc.com

Brand and product names are trademarks of their respective companies.