

CUSTOM DOMAINS...THE GAPS IN STANDARDIZED DATA

Author: Michael Beers, Pinnacle 21

PHUSE US Connect 2023

Paper #DS07

The logo consists of a cluster of hexagons in various shades of orange and yellow. One hexagon is dark and contains the text 'P21' in white. The background of the slide features a teal wavy line pattern on the right and a yellow gradient on the left.

P21

DISCLAIMER

- ▶ The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the regulatory agencies or standards development organizations.



AGENDA

- ▶ Introduction
 - ▶ Guidance for using custom domains
- ▶ Custom domains used per version of SDTMIG
- ▶ Patterns of use
- ▶ Challenges in knowing where data should be mapped
- ▶ Impact of using custom domains
- ▶ Conclusion



INTRODUCTION

REGULATORY GUIDANCE - CUSTOM DOMAINS

4.1.1.3 SDTM Domain Specifications

Custom Domains

The SDTMIG permits the creation of custom domains if the data do not fit into an existing domain. Prior to creating a custom domain, sponsors should confirm that the data do not fit into an existing domain. If it is necessary to create custom domains, sponsors should follow the recommendations in the SDTMIG. In addition, sponsors should present their implementation approach in the cSDRG. To provide study data that do not fit into an existing SDTM domain or draft SDTM domain, consider creating a custom dataset aligned with the Study Data Tabulation Model (SDTM). Questions about custom domains should be addressed in pre-submission meetings and documented in the SDSP.

*FDA's Study Data Technical Conformance Guide, October 2022

4.1.1.2 SDTM datasets

Depending on the characteristics of the collected data, it may not fit into an existing domain of SDTM. In such a case, it is acceptable for the applicant to create a custom domain. To perform this, the applicant must confirm that the data does not fit into existing domains, then create a custom domain according to the SDTM IG, and store the data under this domain. Explanation of the custom domain, together with the reason why it was necessary, should be described in the reviewer's guide.

*PMDA's Technical Conformance Guide on Electronic Study Data Submissions, April 2022

CDISC GUIDANCE FOR USING CUSTOM DOMAINS

2.6 Creating a New Domain

This section describes the overall process for creating a custom domain, which must be based on 1 of the 3 SDTM general observation classes. The number of domains submitted should be based on the specific requirements of the study. To create a custom domain,

1. Confirm that none of the existing published domains will fit the need. A custom domain may only be created if the data are different in nature and do not fit into an existing published domain.
- ▶ Establish a domain of a common topic...
 - ▶ Do not create separate domains based on time...
 - ▶ How collected data are used (e.g., to support analyses and/or efficacy endpoints) must not result in the creation of a custom domain...
 - ▶ Data that were collected on separate CRF modules or pages may fit into an existing domain...
 - ▶ If it is necessary to represent relationships between data that are hierarchical in nature (e.g., a parent record must be observed before child records), then establish a domain pair...

CDISC GUIDANCE FOR USING CUSTOM DOMAINS

2.6 Creating a New Domain

This section describes the overall process for creating a custom domain, which must be based on 1 of the 3 SDTM general observation classes. The number of domains submitted should be based on the specific requirements of the

2. Check the SDTM Draft Domains area of the CDISC wiki (<https://wiki.cdisc.org/display/SDD/SDTM+Draft+Domains+Home>) for proposed domains developed since the last published version of the SDTMIG. These proposed domains may be used as custom domains in a submission.
3. Look for an existing, relevant domain model to serve as a prototype. If no existing model seems appropriate, choose the general observation class (Interventions, Events, or Findings) that best fits the data by considering the topic of the observation. As illustrated in the following figure, the general approach for selecting variables for a custom domain is:

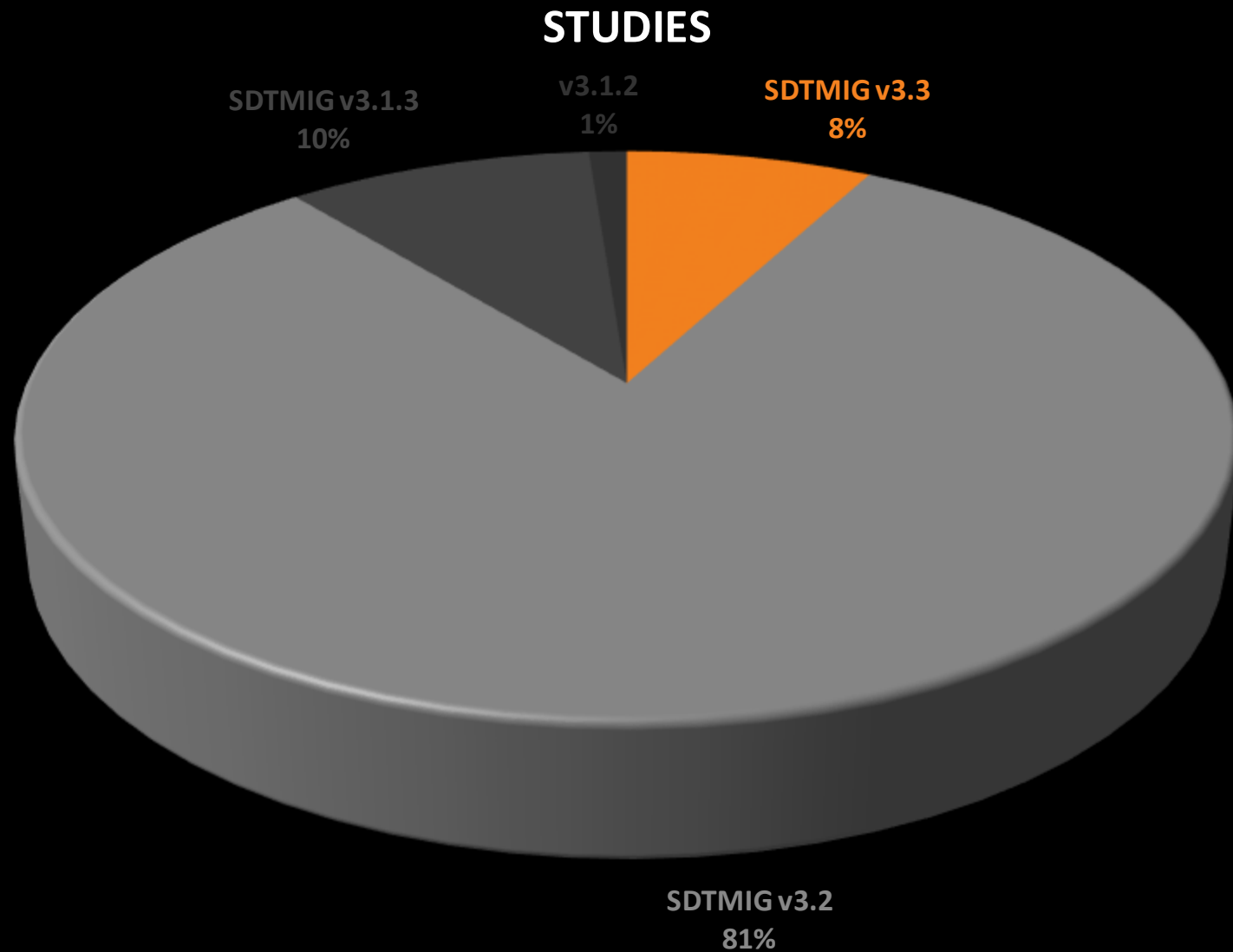
4.1.6 Additional Guidance on Dataset Naming

In some cases, sponsors may need to define new custom domains and may be concerned that CDISC domain codes defined in the future will conflict with those they choose to use. To eliminate any risk of a sponsor using a name that CDISC later determines to have a different meaning, domain codes beginning with the letters X, Y, and Z have been reserved for the creation of custom domains. Any letter or number may be used in the second position. Note the use of codes beginning with X, Y, or Z is optional, and not required for custom domains.



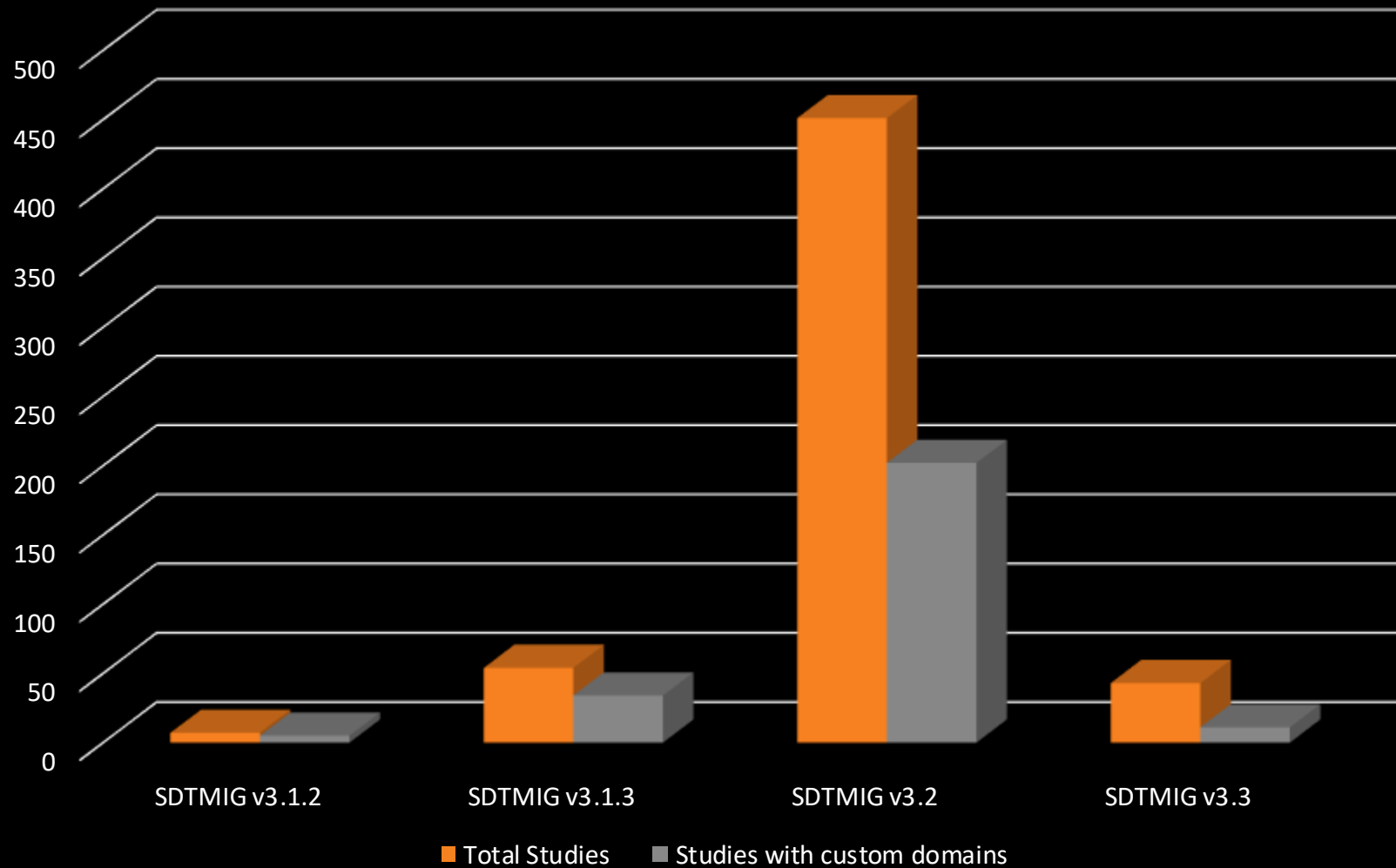
INDUSTRY USAGE OF CUSTOM DOMAINS

VERSIONS OF SDTMIG CURRENTLY BEING USED

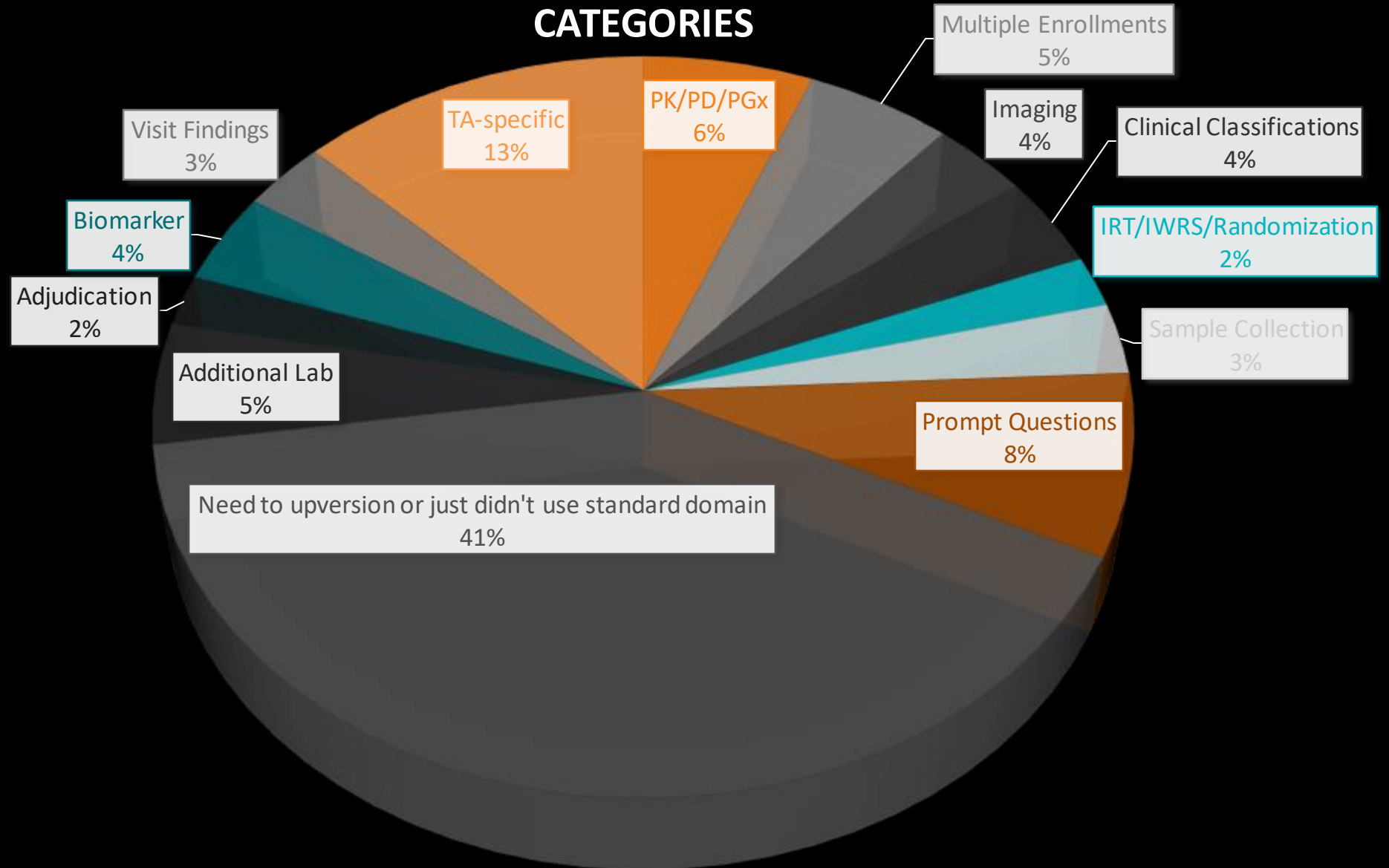


CUSTOM DOMAINS USED PER VERSION OF SDTMIG

Studies and Custom Domains per Standard



CUSTOM DOMAINS USED PER VERSION OF SDTMIG



CUSTOM DOMAINS – MULTIPLE ENROLLMENTS

Custom Domains - Demographics for Multiple Enrollments	
Dataset	SDTMIG
XD - Demographics Support	3.3
DC - Demographics as Collected	3.2
XM - Multiple Participation	3.2
XM - Demographics (Multiple Screenings)	3.2
XD - Demographics for Re-Screened Subjects	3.2
XD - Demographics	3.2
XM - Additional Demographics	3.2
XD - Demographics for Re-Screened Subjects	3.2
XM - Demographics Re Screened Subjects	3.2
DC - Demographics as Collected	3.2
XI - Previous Subject Identifiers	3.2
XI - Previous Subject Identifiers	3.2

Regulatory Guidance:

DM Domain (Demographics)

For subjects with multiple enrollments within a single study, the primary enrollment should be submitted in DM. Additional enrollments should be included in a custom domain with a similar structure to DM. Clarifying statements in the RG would be helpful.

For subjects with multiple screenings and no subsequent enrollment, include the primary screening in DM with additional screenings in a custom domain with a structure similar to DM.

For subjects with multiple screenings and subsequent enrollment, include the enrollment in DM with screenings in a custom domain with a structure similar to DM.

*FDA's Study Data Technical Conformance Guide, October 2022

CUSTOM DOMAINS – VISIT FINDINGS DATA

Custom Domains - Visit Data	
Dataset	SDTMIG
ZO - Office Visit Log	3.2
ZV - Missed Visits	3.2
XC - COVID-19 Missing Visits	3.2
XV - Subject Visit Findings	3.2
XV - Subject Visit Findings	3.2
EE - Epidemic Events	3.2
XU - Unforeseen Impact	3.2

CDISC Guidance:

CDISC Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic Version 1.0

6 Missed Visits

Known Issue: The Subject Visits (SV) domain is a standard domain that includes data about visits. However, the SV domain includes only visits that occurred and there is no way within the SDTM model to add variables to the SV domain as either standard variables or supplemental qualifiers. The custom Visit Events (VE) domain shown in the following example is an interim solution. In the future, it is possible that the solution for recording data about visits that did and did not occur may mean using a modified SV domain or using a new domain (e.g., the VE domain).

Subject Visits and COVID-19

VE or SV?

The Guidance used the domain abbreviation VE for a custom domain to represent subject visits. However, feedback from FDA led to keeping the domain code SV, and feedback from public review led to keeping the classification of SV as a special purpose domain.

What do I submit now, under the current standard?

The FDA has stated their requirements in the TCG. An implementer submitting SV to the FDA would expect to follow their requirements and to explain validation findings in the cSDRG. PMDA has indicated they would accept SV datasets prepared to meet FDA requirements. Specific questions about a particular submission can, as always, be discussed with the regulatory authority.

*<https://www.cdisc.org/kb/articles/subject-visits-and-covid-19>

Regulatory Guidance:

Appendix D: Additional Documents Evaluated By FDA

2. CDISC Document: Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic

It is the current preference of the Agency that for all clinical studies, not limited to those impacted by COVID-19, subject visit data for scheduled (whether or not they occurred), and unscheduled visits be submitted in one single dataset structured as the current CDISC Subject Visits (SV) domain. It is also Agency preference that three non-standard variables (NSVs) for missed visits, --REASOC (Reason for Occur Value), --EPCHGI (Epi/Pandemic Related Change Indicator), and --CNTMOD (Contact Mode), outlined in the CDISC document “Guidance for Ongoing Studies Disrupted by COVID-19 Pandemic” be included within the SV domain and not within the supplemental SUPPSV domain or in other SDTM datasets. Submitting subject visits information in one single structured dataset allows both the human and technology consumer of this information to operate efficiently and with confidence that all visit data are considered during regulatory review.

*FDA’s Study Data Technical Conformance Guide, October 2022

CUSTOM DOMAINS – PROMPT QUESTIONS

Custom Domains - Prompt Questions	
Dataset	SDTMIG
XA - Reference Information	3.3
XM - Clinical Monitoring	3.3
XA - Reference Information	3.2
XD - Debriefing Questions	3.2
XI - Screening Questions	3.2
XQ - Questions Dataset	3.2
QU - Questions Dataset	3.2
PQ - Proxy Question	3.2
ZZ - Leading Questions	3.2
XX - CRF Findings	3.2
XQ - Findings	3.2
ZA - PROTOCOL REQUIREMENTS	3.2
ZN - External Data	3.2
PQ - Proxy Question	3.1.3
XX - Study Maintenance	3.1.3
MQ - Monitoring Questions	3.1.2

CDISC Guidance:

4.5.7 Presence or Absence of Prespecified Interventions and Events

Interventions (e.g., concomitant medications) and events (e.g., medical history) can generally be collected in 2 different ways, by recording either verbatim free text or the responses to a prespecified list of treatments or terms. Because the method of solicitation for information on treatments and terms may affect the frequency at which they are reported, whether they were prespecified may be of interest to reviewers. The --PRES variable is used to indicate whether a specific intervention (--TRT) or event (--TERM) was solicited. The --PRES variable has controlled terminology of "Y" (for "Yes") or a null value. It is a permissible variable, and should only be used when the topic variable values come from a prespecified list. Questions such as "Did the subject have any concomitant medications?" or "Did the subject have any medical history?" should not have records in an SDTM domain because (1) these are not valid values for the respective topic variables of CMTRT and MHTERM, and (2) records whose sole purpose is to indicate whether or not a subject had records are not meaningful.

*Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 (Final), CDISC, 2021-11-29

TU – Assumptions

9. If indicator questions for specific types of tumor or lesions are collected (e.g., Does the subject have target tumors? Does the subject have any non-targets? Did the subject have metastatic disease at screening?), then these TUTESTs will be included in TU. If indicator questions are not collected, do not introduce them into TU.

*Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 (Final), CDISC, 2021-11-29

CUSTOM DOMAINS – ADDITIONAL LAB DATA

Custom Domains - Additional Lab Data	
Dataset	SDTMIG
XB - Laboratory Test Results (Conventional)	3.3
ZT - Other Test	3.3
XB - Laboratory CRF Data/Local Laboratory Data	3.2
XB - Laboratory Test Results (Conventional)	3.2
XL - Non-Safety Laboratory Results	3.2
ZL - Laboratory Test Results - US Units	3.2
XL - Laboratory Test Results XL	3.2
XL - Laboratory Test Results XL	3.2
XD - Other Tests Results	3.2
XF - Hormones	3.2
ZL - Laboratory Test Results - US Units	3.1.3
ZM - Lab Miscellaneous	3.1.3

Regulatory Guidance:

4.1.1.3 SDTM Domain Specifications

LB Domain (Laboratory)

The size of the LB domain dataset submitted by sponsors is often too large to process (See section 3.3.2). This issue can be addressed by splitting a large LB dataset into smaller datasets according to LBCAT and LBSCAT, using LBCAT for initial splitting. If the size is still too large, then use LBSCAT for further splitting. For example, use the dataset name lb1 (file name 'lb1.xpt') for chemistry, dataset name lb2 (file name 'lb2.xpt') for hematology, and dataset name lb3 (file name 'lb3.xpt') for urinalysis. Splitting the dataset in other ways (e.g., by subject or file size) makes the data less useable. Sponsors should submit these smaller files in addition to the larger non-split standard LB domain file. Sponsors should submit the split files in a separate sub-directory/split that is clearly documented in addition to the non-split standard LB domain file in the SDTM datasets directory (See section 7). FDA may require laboratory data using conventional units for reviewing submissions and labeling. Sponsors should discuss with the review divisions what laboratory data should utilize conventional units prior to submission.

*FDA's Study Data Technical Conformance Guide, October 2022

FAQs on Electronic Study Data Submission

4. Questions on CDISC-conformant electronic study data

A: The use of SI units is required for all variables and parameters of test results to be stored in the Findings class domain of the SDTM dataset, as long as SI units are applicable. However, it is acceptable to store only the data in mmHg in the SDTM dataset without storing the converted data in SI unit for test results (e.g. blood pressure) collected in mmHg as conventional unit.

When the original data and the converted data in SI units are stored together in the SDTM dataset, store the converted data into "--STRESN" (or "--STRESN" if necessary) and the original data into "--ORRES". Also, in the reviewer's guide or dataset definition document, describe how each datum (original and converted) is stored, as well as the conversion equation.

In the case of multicenter or multiregional studies, etc., where different measurement units are used for capturing data by center or region, data in multiple units could be obtained for a single parameter (laboratory test). In such cases, it is possible to store data in uniform units other than SI units into "SUPP--" if necessary.

*PMDA's FAQs on Electronic Study Data Submission, June 2022

CDISC Guidance:

Standardized Lab Units

Regulatory Strategy

CDISC does not provide a specific solution when a sponsor is asked to provide both. If a Sponsor anticipates an FDA request for conventional units, a common solution is to produce two versions of the affected datasets (e.g., LB): One with standardized results in SI units (i.e., value in LBSTRESU is an SI unit) and one with standardized results in US conventional units (the value in LBSTRESU is a conventional unit).

*<https://www.cdisc.org/kb/articles/standardized-lab-units>

CUSTOM DOMAINS – ADJUDICATION DATA

Custom Domains - Adjudication	
Dataset	SDTMIG
XA - Adjudication	3.3
ZA - Adjudication	3.3
ZA - Adjudicated Results	3.2
EA - Event Adjudication	3.2
XA - Adjudication Results	3.1.3

PHUSE Guidance:

Recommendation

Best practices in data handling for adjudicated outcomes can be summarized as follows:

Can the same data structure be generalized for use across all therapeutic areas?

The therapeutic area should, on a general level, not affect the structure of the data. The same basic information should be present irrespective of the therapeutic area, with the possibility of adding disease-specific detailed assessments. If the assessments are to track the development of the investigated disease in the trial, then it is not in this concept of adjudication; hence, other specific disease domains, e.g. TU, TR, RS, should be used.

What level of detailed data should be submitted? Is it sufficient to submit only data points for the final outcome or should individual adjudicator assessments be submitted as well?

There are different approaches observed in the industry for the submission of results from the adjudication process; some submit only final agreed assessments from the adjudicators and some submit individual adjudicator assessments from primary adjudication as well. However, based on the experiences of the team, **a new Event Adjudication (EA) domain is proposed** and the suggestions below are made for submission of event adjudication data.

*Best Practices for Submission of Event Adjudication Data, PHUSE, October 2019

Regulatory Guidance:

4.1.1.2 SDTM General Considerations

Adjudication Data

There are no existing standards or best practices for the representation of adjudication data as part of a standard data submission. Until standards for adjudication data are developed, it is advised that sponsors discuss their proposed approach with the review division and also include details about the presence, implementation approach, and location of adjudication data in the SDRG.

Whenever adjudication data are provided, they should be clearly identified so that the reviewer can distinguish the results of adjudication from data as originally collected.

*FDA's Study Data Technical Conformance Guide, October 2022

CDISC Guidance:

4.5.4 Evaluators in the Interventions and Events Observation Classes

This example illustrates a case where an adjudication committee evaluated an adverse event. The evaluations of the adverse event by the primary investigator were represented in the standard AE dataset. **The evaluations of the adjudication committee were represented in SUPPAE.** See Section 8.4, [Relating Non-standard Variable Values to a](#)

*Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 (Final), CDISC, 2021-11-29

TR – Assumptions

5. The acceptance flag variable (TRACPTFL) identifies those records that have been determined to be the accepted assessments/measurements by an independent assessor. This flag would be provided by an independent assessor and when multiple assessors (e.g., "RADIOLOGIST 1", "RADIOLOGIST 2", "ADJUDICATOR") provide assessments or evaluations at the same time point or an overall evaluation. This flag should not be used by a sponsor for any other purpose. It is not expected that the TRACPTFL flag would be populated by the sponsor; instead, that type of record selection should be handled in the analysis dataset (ADaM).

*Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 (Final), CDISC, 2021-11-29



IMPLEMENTATION CHALLENGES AND IMPACT

CHALLENGES - KNOWING WHERE DATA GETS MAPPED

Example 1:

▶ Apgar score

- ▶ Find something related to Apgar in CT...CCCAT value
- ▶ Find out where CCCAT is mapped to (RS domain)

Example 2:

▶ ZO - Clinical Outcome dataset

- ▶ Uses values from Disease Outcome SDTM codelist (C102578)
 - ▶ Values such as: CLINICAL CURE, CLINICAL FAILURE
 - ▶ This codelist is not listed in any IG or anything. Where does the data get mapped to using this codelist?

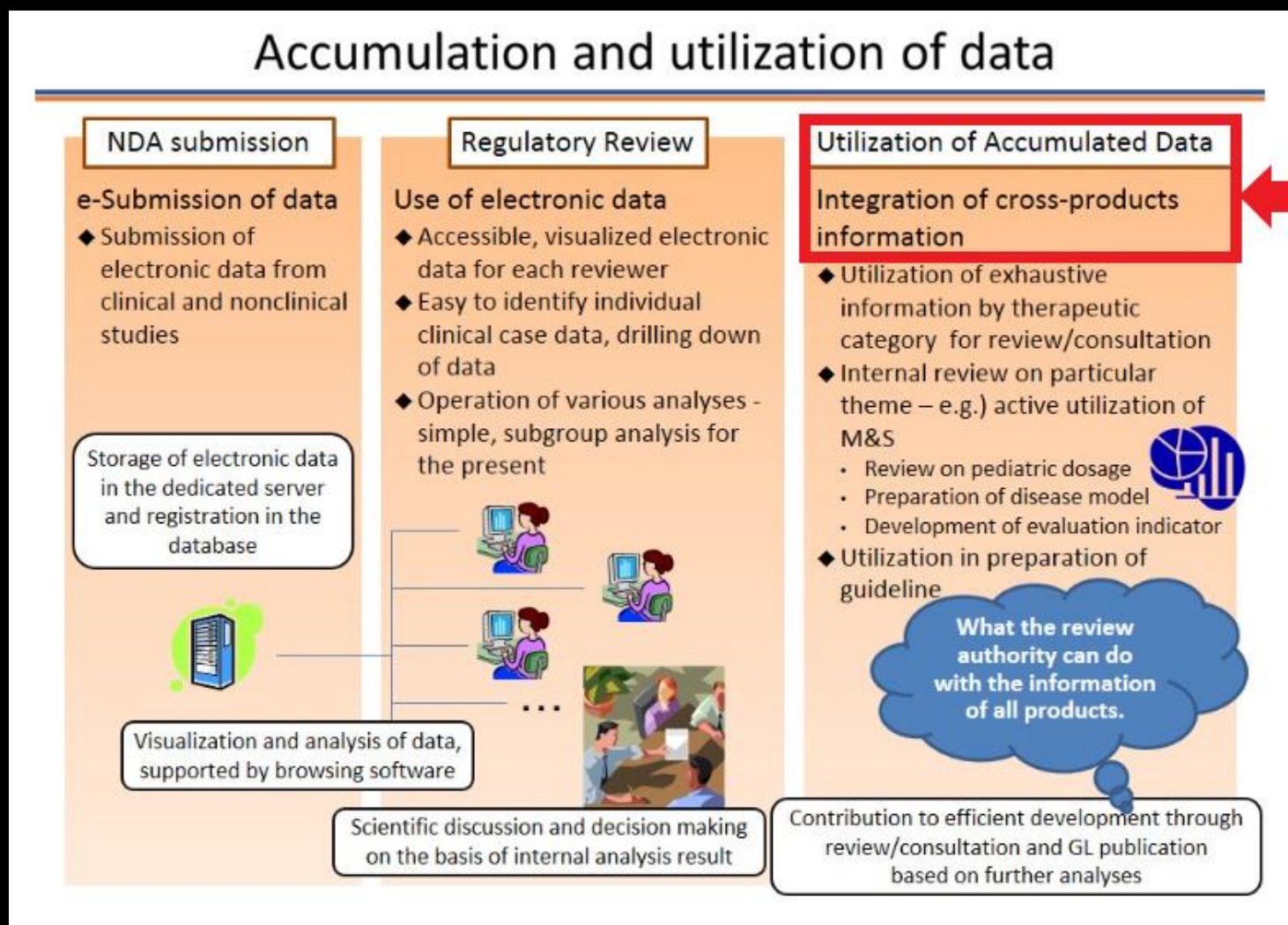
IMPACT OF NOT USING STANDARD DOMAINS

- ▶ Goal is to make drugs available ASAP
- ▶ Increased automation is needed for this goal
- ▶ Manual review process is inefficient
 - ▶ Data review time is limited
 - ▶ Data quality evaluation is difficult
- ▶ Automation relies on high quality data
 - ▶ Value of standard reporting and analysis for Reviewers increases demand for data quality
- ▶ Standardized data allows early evaluation and enforcement of data quality
 - ▶ Customization is not standardization

IMPACT OF NOT USING STANDARD DOMAINS

- ▶ Validation is too general on custom domains
 - ▶ Domain-specific checks not run and domain-specific codelists not checked properly
 - ▶ Example: Additional Lab Data
 - ▶ In SDTMIG v3.3, there are 155 validation rules run on the LB domain, but there are only 144 validation rules run on a custom findings domain.
 - ▶ Variables such as –TESTCD and –TEST would not be checked against the LBTESTCD and LBTEST codelists, even though the controlled terminology should have been used.
- ▶ Validation might not even occur
 - ▶ Custom special-purpose domains, such as for Multiple Enrollments
 - ▶ This results in an unvalidated dataset being submitted to the regulatory agency

IMPACT OF NOT USING STANDARD DOMAINS



*<https://www.pmda.go.jp/english/review-services/reviews/0002.html>

CONCLUSION

- ▶ Use of custom domains is going down with implementation of each new SDTMIG
 - ▶ There is always room for improvement however
- ▶ More precise guidance should be created for some of these patterns of custom domain use
- ▶ An easier way to determine where data should be mapped would be very useful
- ▶ Reducing custom domain usage = more standardization = more automation

A blurred background image of a modern office hallway with large glass walls. Several people are walking through the hallway, their figures out of focus, creating a sense of motion. The floor is a light-colored, polished surface.

QUESTIONS?

KEEP IN TOUCH!

MICHAEL BEERS

MBEERS@PINNACLE21.COM

You will receive an email when the
slides are are posted on our blog:

www.pinnacle21.com/blog

