Paper SA06

Submitting eCRT SDTM Data for ISS

Aida Swierk, Takeda Development Center Americas, Inc., Cambridge, MA, USA

ABSTRACT

The creation of electronic Case Report Tabulations (eCRT) for an Integrated SDTM (iSDTM) submission is a complex, cross-functional governance and decision ownership layered process that requires rigorous planning, coordination, and metadata governance across multiple clinical studies. Unlike single-study submissions, Integrated Summary of Safety (ISS) eCRT development demands early alignment on study scope, harmonization strategies, controlled terminology, and the standards that will govern the pooled submission. Differences among contributing studies in CRFs, SDTM specifications, derivation logic, and controlled terminology present significant challenges that must be resolved to ensure a consistent, regulatory-ready package.

A central component of ISS eCRT development is the define.xml, which must ensure traceability by accurately representing pooled origins, integration-level assigned values, derivations, remapped values, and value-level metadata. Because tools such as P21E are not designed to natively support integrated eCRT creation, to address this gap, a structured process was developed to harmonize variable origins, link study-specific and pooled derivations, and clearly document variable remapping performed at the integrated level; all of which ensure traceability and helps reviewer to have a full picture on interpretability of both the data and metadata level. The goal of this process is to establish a clear, standardized methodology for using P21E to generate pooled SDTM deliverables that fully uphold core CDISC fundamentals—particularly the principle of end-to-end traceability.

INTRODUCTION

The preparation of electronic Case Report Tabulations (eCRTs) for an Integrated Summary of Safety (ISS) is inherently more complex than preparing eCRT packages for individual clinical studies. Integrated SDTM (iSDTM) submissions must unify data, metadata, derivations, and controlled terminology across multiple studies that may differ in design, level of maturity, data standards, and vendor implementation practices. Regulatory expectations for transparency, consistency, and traceability further elevate the importance of rigorous metadata management and clear documentation.

In many ISS programs, contributing studies are still ongoing as the integrated submission effort begins. As databases approach lock, updates to annotated CRFs, SDTM data, derivation logic, and controlled terminology occur frequently requiring repeated iterations of integration activities and careful version control. Inputs from each study are rarely consistent: some may provide comprehensive SDTM specifications and a complete define.xml package, while others do not. Teams need a consistent "source of truth" for derivations and metadata to reconcile differences among aCRFs, specifications, define.xml files, and delivered data.

A further challenge is that widely used tools such as P21E do not natively support integrated eCRT creation. Unlike single-study workflows, integration requires the ability to manage pooled origins, multi-study derivations, integration-level assigned values, and variable- and value-level metadata within a single cohesive metadata package. To address these limitations, a new process using the existing P21E is developed. We propose a practical framework to generate the Origin Report, Controlled Terminology Report, Derivation & Comments Report, and Remapping Report and include them in P21E and reviewer's guide. These reports together with iSDTM data, define.xml and icsdrg reviewer's guide, are foundational for demonstrating end-to-end traceability from each contributing study to the integrated datasets and the ISS-level define.xml, fulfilling regulatory expectations for transparency and completeness. This paper presents the methodology, custom workflows, and lessons learned from preparing ISS-level eCRTs, with emphasis on metadata governance, traceability, and practical solutions to the technical challenges inherent in integrated SDTM submissions.

WHY EARLY ENGAGEMENT MATTERS

ISS integrations almost always involve **studies at very different lifecycle stages**, each with its own maturity of documentation, standards adherence, annotation quality, and metadata completeness. The scope of submission can differ amongst them and the decision if they are going to have individual submission with the integrated one, plays a major role in how the integration is handled. This creates challenges for integration programmers, including:

- inconsistent variable naming for identical concepts
- mismatched controlled terminology versions
- differences in annotation quality (incorrect or missing aCRF mapping)

- different SDTM implementations or derivation approaches
- misaligned TS terminology or dictionary versions

Early engagement begins at protocol and Statistical Analysis Plan (SAP) finalization or first SDTM spec draft and continues through interim data cuts and pre-lock reviews. Experience shows that reviewing study inputs early allows the team to identify harmonization issues sooner and work with study programmers to resolve them before integration, and it reduces late-stage surprises. This leads to:

- better alignment on standards
- pooled origin assignment rules
- handling EPOCH / ARM / ACTARM / VISIT mapping across studies
- decision on value-level metadata consistency

An early package needs at the minimum these components from each study:

- aCRF,
- STMD specs/define.xml if available
- codelists,
- TS plan,
- mapping decisions

WE NEED ISDTM DEFINE.XML, NOW WHAT?

Once the need for ISS has been identified, the ISS programmers should begin **early assessment and harmonization planning**, including:

- Confirm list of contributing studies and their submission requirements
- Are all the studies included in the integration going to be submitted or not?
- Retrieve everything that available: define packages, aCRFs, SDTM versions, CT versions, and dictionaries
- Identify missing or inconsistent metadata
- Establish the ISS harmonization tracker to document discrepancies and alignment decisions

This early start is crucial because downstream define.xml and eCRT work relies heavily on cross-study consistency.

EARLY WORKSTREAMS FOR DEFINE.XML COMPONENTS

Below are the four major building blocks of define.xml — **aCRF**, **Data**, **Specifications**, and the **Reviewer's Guide** — along with what should begin **as early as possible**, given the variability across study lifecycles.

A. ACRF (ANNOTATED CRF)

We decided that the best approach for the aCRF that would be associated with integrated eCRT should be the combined aCRF. The easiest to create and maintain was the stacked approach, where every individual study aCRF is present. In addition to the already existing annotation, on the top of each page, we added the individual study identification name/number and the information of current page number in format "Page X of Y" where X is the current page and Y is the total number of pages for all combined aCRFs. This approach allowed us to automate checks with issues in the annotation and made possible the current and complete entry of page numbers in P21E.

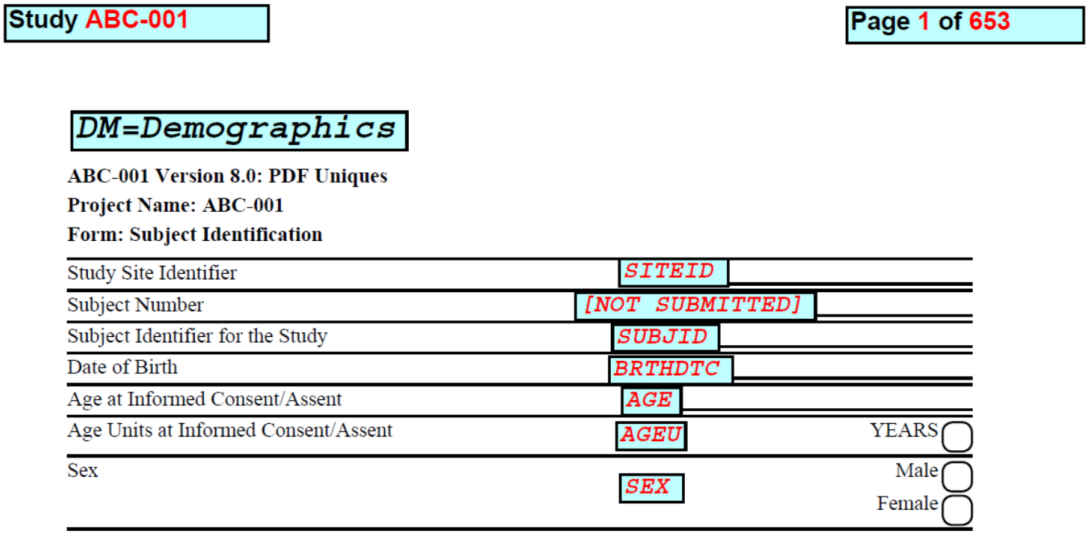


Figure 1

WHAT SHOULD BEGIN EARLY

1. STACKING OF INDIVIDUAL ACRF FILES

Stack and combine all study-level annotated CRFs early to detect:

- missing annotations
- inconsistent annotation conventions and data
- differences in variable origins or page references
- misalignment in derivation descriptions

2. RETRIEVING PAGE INFORMATION FROM THE COMBINED PDF

Extract page number references from the stacked, combined aCRF PDF.
This is essential for:

- validating aCRF page mapping
- ensuring consistency across studies with different CRF structures
- setting up the aCRF Page Number field in define.xml

3. COMPARE P21E PAGE METADATA VS. EXTRACTED PAGE INFO

Perform an early **P21E vs. aCRF comparison** to catch:

- missing page numbers
- incorrect mappings
- inconsistent CRF annotations between studies

Doing this early avoids a heavy wave of corrections right before define.xml finalization. Although P21E populates the tool with the page numbers of the combined aCRF, we found out that there were a lot of issues.

B. DATA (SDTM / POOLED ISDTM)

All data of studies involved in integration were pooled together. Any changes that were identified at this process followed a simple approach, keep the original information in the supplemental domains by renaming the original variable to "_variable_name" and implement the update to the "variable_name". All these changes were documented in separate file from which a report was created and was linked to the reviewer's guide. Here is how that looks like:

ABC ISS - Value Remapping Report
Purpose: To display the consolidated list of variable values that have been remapped at the ISS level from their original mappings in individual studies
Date Generated: 02FEB2026
Domain: CM

Domain	Study	Subset Condition	Variable Name	Variable Value in Study	Value in ISS data
CM	ABC01		CMCAT	HISTORY OF BIOLOGIC THERAPY	PRIOR BIOLOGIC THERAPIES
	ABC02		CMCAT	HISTORY OF BIOLOGIC THERAPY	PRIOR BIOLOGIC THERAPIES

Figure 2

Early data review is critical because studies deliver data at different states of completeness.

WHAT SHOULD BEGIN EARLY

1. STACKING SDTM DATA ACROSS STUDIES

Start combining datasets early to check:

- structural differences
- inconsistent variable lengths or types
- presence/absence of required domains
- mismatches against standards
- domain-level alignment challenges

2. MANAGING UPDATES WHILE PRESERVING ORIGINAL INFORMATION

When updates occur:

- retain original values in SUPPQUAL as appropriate
- document all integration-level changes
- track differences in the integration tracker

3. KEEP ONLY DOMAINS USED IN ADaM

Early identification of required SDTM domains helps avoid unnecessary work and reduces validation burden. Only non-blank datasets used in ADaM and/or that are key for other domains, such as Subject Visit (SV) should be included in the pooled data used to create eCRT.

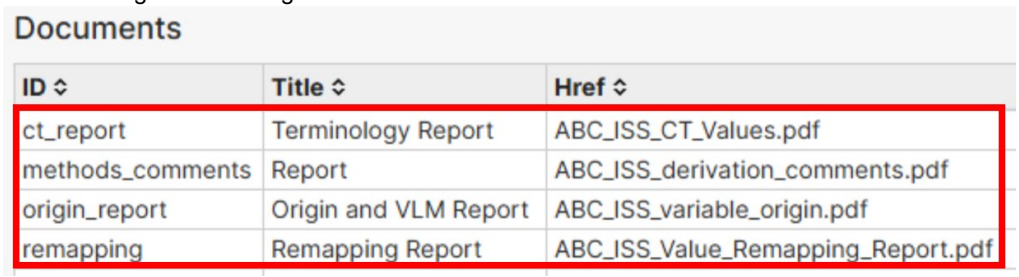
4. CREATE POOLED TS DOMAIN (TS.XPT) USING TECHNICAL CONFORMANCE GUIDELINES

TS domain is a key component in the submission of the eCRT package. Creating a pooled one requires attention to harmonize the values of TSPARM/TSPARMCD across studies, resolve TSVAL/TSVALCD inconsistencies and checks with SNOWMED registry sources. One of the key components that requires extra attention is the population of TSPARMCD="SPREFID". This parameter code should be present for each study, and its value should be "ISS".

C. SPECIFICATIONS (ORIGIN, DERIVATIONS, COMMENTS AND CONTROLLED TERMINOLOGY)

Specifications benefit significantly from early involvement because they reflect study design/maturity differences. In our experience we had studies that had complete specifications, some had partial specifications or missing. Regardless of situation you need to create the eCRT at some point, so let data guide you to understand how things are. Handle specifications with care but trust data first!

Our decision was to use the individual study define.xml as the source of our information to generate pooled define.xml. P21E expects to have information on the origin, controlled terminology and derivation for each variable that is present in define.xml. For this reason, we made some rules and created 3 reports which were used in the pooled define.xml creation: one for the origin of the variables, one for the controlled terminology and another for the derivation and comments. All of them show the details of the information from each individual study, regardless if that study is being submitted individually or not. The value remapping report, shown in Figure 2, was uploaded and linked with the define.xml as well, so reviewer can have all information handy, what is available from each project and what has changed in the integration. The figure below shows them in P21E.


Documents

ID	Title	Href
ct_report	Terminology Report	ABC_ISS_CT_Values.pdf
methods_comments	Report	ABC_ISS_derivation_comments.pdf
origin_report	Origin and VLM Report	ABC_ISS_variable_origin.pdf
remapping	Remapping Report	ABC_ISS_Value_Remapping_Report.pdf

Figure 3

Following table provides description of how we agree to set the rules for the origin of the variable in iSDTM define package, the other origin column shows the remaining distinct origins that are present in different individual studies. “ISS Origin” column presents what is entered in define.xml while “Other Origin” column is used in P21E to enter information that can be presented or linked in define.xml. To justify all the origins across studies by considering ISS and other origins methods, comments and page numbers are added accordingly in define.xml.

Origin from any study	In Combination with origins from all other studies	ISS Origin	Other Origin
Same origin across all studies		Original	
Derived	Assigned and Collected	Derived	Collected, Assigned
Derived	Collected	Derived	Collected
Derived	Assigned	Derived	Assigned
Derived	Collected and eDT	Derived	Collected, eDT
Derived	Assigned and eDT	Derived	Assigned, eDT
Derived	Assigned and Protocol	Derived	Assigned, Protocol
Derived	Collected and Assigned and Protocol	Derived	Collected, Assigned, Protocol
Derived	Collected and Assigned and eDT	Derived	Collected, Assigned, eDT
Derived	eDT	Derived	eDT
Assigned	Protocol	Assigned	Protocol
Assigned	eDT	Assigned	eDT
Assigned	eDT and Protocol	Assigned	eDT, Protocol
Collected	Assigned	Collected	Assigned
Collected	Assigned, eDT	Collected	Assigned, eDT
Collected	Assigned, Protocol	Collected	Assigned, Protocol
Collected	eDT	Collected	eDT

Table 1

Figure 4 shows how the variable origins report looks like:

ABC ISS - Variable Origin Report
Purpose: Displays the Variable origin from individual studies
Report Generated On: 02FEB2026
Domain: AE

Dataset	Variable	ABC001	ABC002	ABC003	ABC004	ABC005
AE	STUDYID	Protocol	Protocol	Protocol	Protocol	Protocol
AE	DOMAIN	Assigned	Assigned	Assigned	Assigned	Assigned
AE	USUBJID	Derived	Derived	Derived	Derived	Derived
AE	AESEQ	Derived	Derived	Derived	Derived	Derived
AE	AESPID	Collected	Assigned	eDT	Derived	Derived
AE	AETERM	Collected	Collected	Collected	Collected	Collected

Figure 4

Figures 5 and 6, show how the information is in P21E; the same variable AESPID is given as an example.

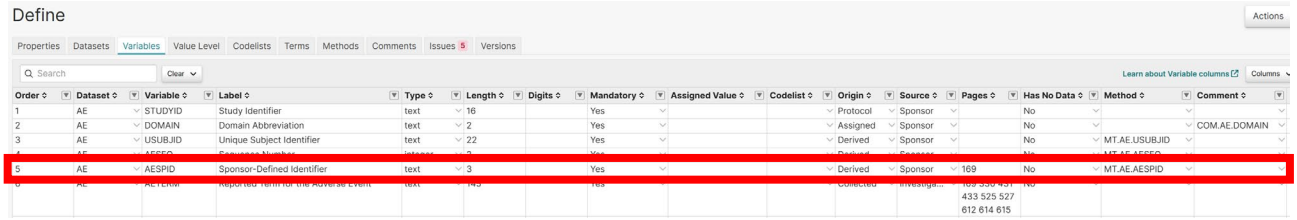


Figure 5

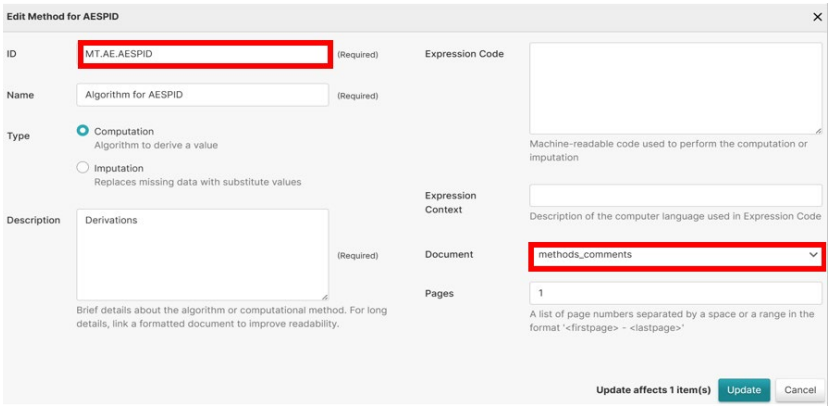
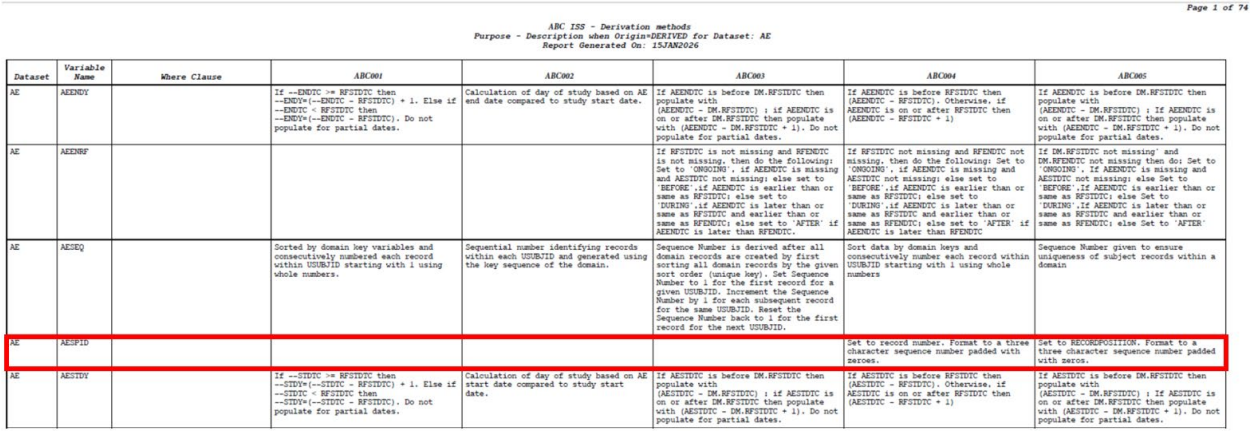


Figure 6

Figure 7 shows how the derivation and comments report looks like with same variable AESPID given as an example.



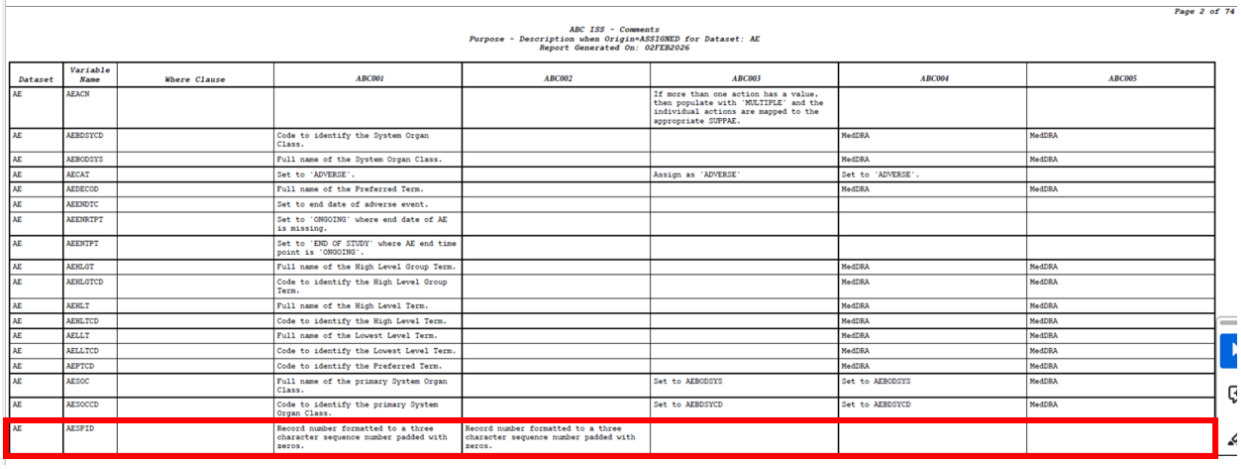
Page 1 of 74

ABC ISS - Derivation methods
Purpose - Description when Origin=DERIVED for Dataset: AE
Report Generated On: 15JAN2026

Dataset	Variable Name	Where Clause	ABC001	ABC002	ABC003	ABC004	ABC005
AE	AEENDY		If --ENDTC >= RFSTDTC then --ENDY=(--ENDTC - RFSTDTC) + 1. Else if --ENDTC < RFSTDTC then --ENDY=(--ENDTC - RFSTDTC). Do not populate for partial dates.	Calculation of day of study based on AE end date compared to study start date.	If AEENDTC is before DM.RFSTDTC then populate with (AEENDTC - DM.RFSTDTC) ; if AEENDTC is on or after DM.RFSTDTC then populate with (AEENDTC - DM.RFSTDTC + 1). Do not populate for partial dates.	If AEENDTC is before RFSTDTC then (AEENDTC - RFSTDTC). Otherwise, if AEENDTC is on or after RFSTDTC then (AEENDTC - RFSTDTC + 1)	If AEENDTC is before DM.RFSTDTC then populate with (AEENDTC - DM.RFSTDTC) ; If AEENDTC is on or after DM.RFSTDTC then populate with (AEENDTC - DM.RFSTDTC + 1). Do not populate for partial dates.
AE	AEENRF				If RFSTDTC is not missing and RFENDTC is not missing, then do the following: Set to 'ONGOING', if AEENDTC is missing and AESTDTC not missing; else set to 'BEFORE',if AEENDTC is earlier than or same as RFSTDTC; else set to 'DURING',if AEENDTC is later than or same as RFSTDTC and earlier than or same as RFENDTC; else set to 'AFTER' if AEENDTC is later than RFENDTC.	If RFSTDTC not missing and RFENDTC not missing, then do the following: Set to 'ONGOING', if AEENDTC is missing and AESTDTC not missing; else set to 'BEFORE',if AEENDTC is earlier than or same as RFSTDTC; else set to 'DURING',if AEENDTC is later than or same as RFSTDTC and earlier than or same as RFENDTC; else set to 'AFTER' if AEENDTC is later than RFENDTC	If DM.RFSTDTC not missing' and DM.RFENDTC not missing then do: Set to 'ONGOING', If AEENDTC is missing and AESTDTC not missing; else Set to 'BEFORE',If AEENDTC is earlier than or same as RFSTDTC; else Set to 'DURING',If AEENDTC is later than or same as RFSTDTC and earlier than or same as RFENDTC; else Set to 'AFTER'
AE	AESEQ		Sorted by domain key variables and consecutively numbered each record within USUBJID starting with 1 using whole numbers.	Sequential number identifying records within each USUBJID and generated using the key sequence of the domain.	Sequence Number is derived after all domain records are created by first sorting all domain records by the given sort order (unique key) . Set Sequence Number to 1 for the first record for a given USUBJID. Increment the Sequence Number by 1 for each subsequent record for the same USUBJID. Reset the Sequence Number back to 1 for the first record for the next USUBJID.	Sort data by domain keys and consecutively number each record within USUBJID starting with 1 using whole numbers	Sequence Number given to ensure uniqueness of subject records within a domain
AE	AESPID					Set to record number. Format to a three character sequence number padded with zeroes.	Set to RECORDPOSITION. Format to a three character sequence number padded with zeros.
AE	AESTDY		If --STDTC >= RFSTDTC then --STDY=(--STDTC - RFSTDTC) + 1. Else if --STDTC < RFSTDTC then --STDY=(--STDTC - RFSTDTC). Do not populate for partial dates.	Calculation of day of study based on AE start date compared to study start date.	If AESTDTC is before DM.RFSTDTC then populate with (AESTDTC - DM.RFSTDTC) ; if AESTDTC is on or after DM.RFSTDTC then populate with (AESTDTC - DM.RFSTDTC + 1). Do not populate for partial dates.	If AESTDTC is before RFSTDTC then (AESTDTC - RFSTDTC). Otherwise, if AESTDTC is on or after RFSTDTC then (AESTDTC - RFSTDTC + 1)	If AESTDTC is before DM.RFSTDTC then populate with (AESTDTC - DM.RFSTDTC) ; If AESTDTC is on or after DM.RFSTDTC then populate with (AESTDTC - DM.RFSTDTC + 1). Do not populate for partial dates.

Figure 7

One of the limitations on P21E is that there is a limitation on the number of reports that you can link to an individual variable. In cases when the same variable can be derived in one study but have a comment in another, to have a full disclosure of the specifications, we combined both derivation and comments into one report. Let's see as an example AESPID which has a derivation on page 1, while same variable has a comments on page 2 of the same report.



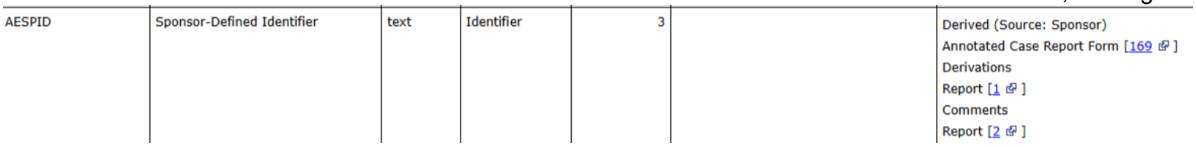
Page 2 of 74

ABC ISS - Comments
Purpose - Description when Origin=ASSIGNED for Dataset: AE
Report Generated On: 02FEB2026

Dataset	Variable Name	Where Clause	ABC001	ABC002	ABC003	ABC004	ABC005
AE	AEACN				If more than one action has a value, then populate with 'MULTIPLE' and the individual actions are mapped to the appropriate SUPPAE.		
AE	AEBODSYCD		Code to identify the System Organ Class.			MedDRA	MedDRA
AE	AEBODSYS		Full name of the System Organ Class.			MedDRA	MedDRA
AE	AECAT		Set to 'ADVERSE'.		Assign as 'ADVERSE'	Set to 'ADVERSE'.	
AE	AEDECOD		Full name of the Preferred Term.			MedDRA	MedDRA
AE	AEENDTC		Set to end date of adverse event.				
AE	AEENRTPT		Set to 'ONGOING' where end date of AE is missing.				
AE	AEENTPT		Set to 'END OF STUDY' where AE end time point is 'ONGOING'.				
AE	AEHLGT		Full name of the High Level Group Term.			MedDRA	MedDRA
AE	AEHLGTCD		Code to identify the High Level Group Term.			MedDRA	MedDRA
AE	AEHLT		Full name of the High Level Term.			MedDRA	MedDRA
AE	AEHLTCD		Code to identify the High Level Term.			MedDRA	MedDRA
AE	AELLT		Full name of the Lowest Level Term.			MedDRA	MedDRA
AE	AELLTCD		Code to identify the Lowest Level Term.			MedDRA	MedDRA
AE	AEPTCD		Code to identify the Preferred Term.			MedDRA	MedDRA
AE	AESOC		Full name of the primary System Organ Class.		Set to AEBODSYS	Set to AEBODSYS	MedDRA
AE	AESOCCD		Code to identify the primary System Organ Class.		Set to AEBODSYCD	Set to AEBODSYCD	MedDRA
AE	AESPID		Record number formatted to a three character sequence number padded with zeros.	Record number formatted to a three character sequence number padded with zeros.			

Figure 8

This is how the define will show the information for the above variables for derivation and comments, see Figure 9.



AESPID	Sponsor-Defined Identifier	text	Identifier	3		Derived (Source: Sponsor) Annotated Case Report Form [169] Derivations Report [1] Comments Report [2]

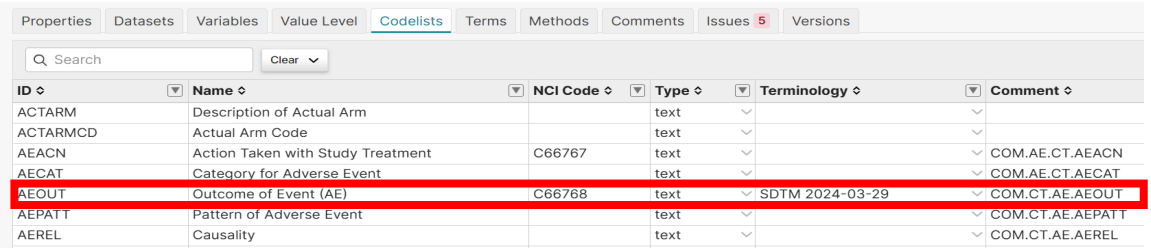
Figure 9

Controlled terminology information was followed in the same approach. Here we used each individual dataset and their content to generate the report. Figure 10 shows how the controlled terminology report looks like with variable AEOUT given as an example. Figures 11 and 12 show how that report is linked.

ABC ISS - Unique Controlled Terminology Values
Purpose: Displays the unique values for all variables which have a controlled terminology
Report Generated On: 15JAN2026
Adverse Events (AE)

Variable	ABC001	ABC002	ABC003	ABC004	ABC005
AEACN	DOSE NOT CHANGED | | DRUG INTERRUPTED | | DRUG WITHDRAWN | | NOT APPLICABLE | | UNKNOWN	DOSE INCREASED | | DOSE NOT CHANGED | | DRUG INTERRUPTED | | DRUG WITHDRAWN | | NOT APPLICABLE	DOSE DELAYED | | DOSE NOT CHANGED | | DRUG INFUSION INTERRUPTED | | DRUG WITHDRAWN | | NOT APPLICABLE	DOSE DELAYED | | DOSE NOT CHANGED | | DRUG INFUSION INTERRUPTED | | DRUG WITHDRAWN | | NOT APPLICABLE | | UNKNOWN	DOSE DELAYED | | DOSE NOT CHANGED | | NOT APPLICABLE
AECAT	ADVERSE	ADVERSE	ADVERSE	ADVERSE	ADVERSE
AEENRF			AFTER | | BEFORE | | DURING | | ONGOING	AFTER | | BEFORE | | DURING | | ONGOING	AFTER | | DURING | | ONGOING
AEENRTPT	ONGOING				
AEOUT	NOT RECOVERED/NOT RESOLVED | | RECOVERED/ RESOLVED | | RECOVERED/ RESOLVED WITH SEQUELAE | | RECOVERING/RESOLVING | | UNKNOWN	NOT RECOVERED/NOT RESOLVED | | RECOVERED/ RESOLVED | | RECOVERED/ RESOLVED WITH SEQUELAE | | RECOVERING/RESOLVING	NOT RECOVERED/NOT RESOLVED | | RECOVERED/ RESOLVED | | RECOVERED/ RESOLVED WITH SEQUELAE | | RECOVERING/RESOLVING | | UNKNOWN	NOT RECOVERED/NOT RESOLVED | | RECOVERED/ RESOLVED | | RECOVERED/ RESOLVED WITH SEQUELAE | | RECOVERING/ RESOLVING	NOT RECOVERED/NOT RESOLVED | | RECOVERED/ RESOLVED | | RECOVERING/RESOLVING
AEPATT	CONTINUOUS | | INTERMITTENT | | ONCE	CONTINUOUS | | INTERMITTENT | | ONCE	CONTINUOUS | | INTERMITTENT | | ONCE	CONTINUOUS | | INTERMITTENT | | ONCE	CONTINUOUS | | INTERMITTENT | | ONCE

Figure 10



ID	Name	NCI Code	Type	Terminology	Comment
ACTARM	Description of Actual Arm		text		
ACTARMCD	Actual Arm Code		text		
AEACN	Action Taken with Study Treatment	C66767	text		COM.AE.CT.AEACN
AECAT	Category for Adverse Event		text		COM.AE.CT.AECAT
AEOUT	Outcome of Event (AE)	C66768	text	SDTM 2024-03-29	COM.CT.AE.AEOUT
AEPATT	Pattern of Adverse Event		text		COM.CT.AE.AEPATT
AEREL	Causality		text		COM.CT.AE.AEREL

Figure 11

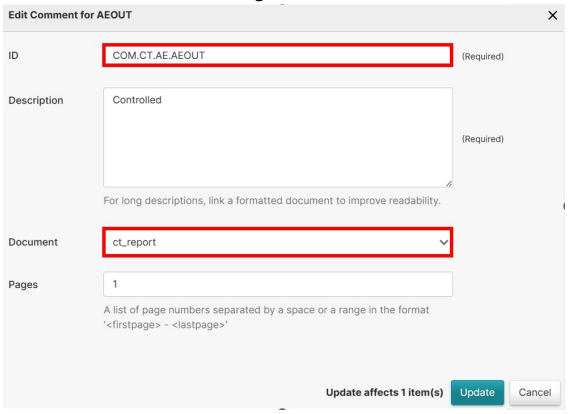
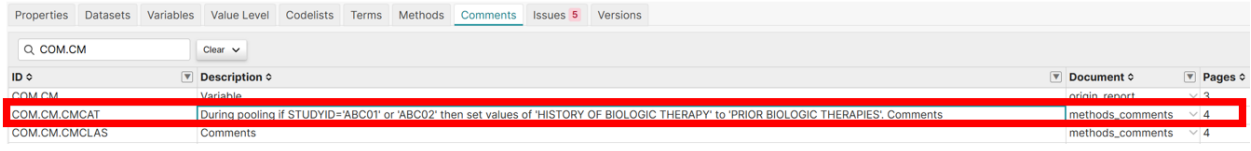


Figure 12

For the variables that have been changed during the pooling information, the details are saved in the comments as is shown below in Figure 13 for CMCAT variable. Beyond the description of change at pooling, there is the link for the derivation and comments report for individual studies in the define.xml.



ID	Description	Document	Pages
COM.CM	Variable	origin_report	3
COM.CM.CMCAT	During pooling if STUDYID='ABC01' or 'ABC02' then set values of 'HISTORY OF BIOLOGIC THERAPY' to 'PRIOR BIOLOGIC THERAPIES'; Comments	methods_comments	4
COM.CM.CMCLAS	Comments	methods_comments	4

Figure 13

REVIEWER'S GUIDE

There were some changes we made to the template of reviewer's guide that is created by P21E. In addition to the information of standards metadata for each study, we added in a table the status of study (ongoing or locked), date of extraction used for this submission, and if the data were blinded or not.

We also created a table with what datasets were provided for each study that participated in the pooling and explained through a flow chart the patient rollover information from one study to the next.

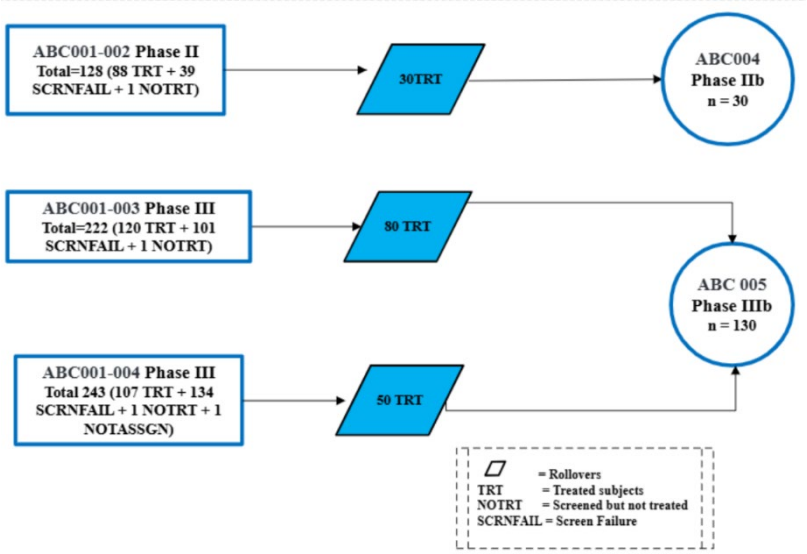


Figure 14

We spent a significant amount of time with the reporting of P21E. For this we decided on the following rule:

In the pooled eCRT Reviewer's Guide, only issues relevant to the integration should be included. Data issues from individual studies are documented only when those studies are not being submitted as standalone packages. For studies submitted independently, their study-specific issues are excluded from the pooled Reviewer's Guide.

CHALLENGES & LESSONS LEARNED

Developing an ISS-level eCRT package at the SDTM layer introduces a set of challenges that go well beyond those typically encountered in single-study submissions. Integrating multiple studies—often produced by different vendors and at different stages of completeness—requires not only strong technical capability but also a disciplined metadata governance framework. The experiences summarized in this work highlight several key challenges and lessons learned that are essential for teams undertaking similar efforts.

1. VARIABILITY ACROSS STUDIES CREATES EARLY INTEGRATION BURDEN

The most significant challenge arises from inconsistencies in inputs. Studies may provide different levels of documentation, some with complete SDTM specifications and define.xml files, others with only datasets. These discrepancies make it difficult to establish a consistent metadata foundation and often require extensive comparison across CRFs, specifications, define.xml files, and actual data.

Lesson Learned: Begin integration work with a clear plan for establishing a "single source of truth" for each variable and derivation, even if that source differs across studies.

2. ONGOING STUDIES REQUIRE ITERATIVE RECONCILIATION

When integration begins before all studies are locked, programming teams must accommodate numerous rounds of updates to CRFs, SDTM datasets, controlled terminology, and derivations.

Lesson Learned: Implement strong version-tracking processes and clearly assigned ownership while anticipating multiple reconciliation cycles to avoid metadata drift.

3. TOOL LIMITATIONS REQUIRE CREATIVE WORKFLOWS

Since P21E is designed for single-study workflows, the integration team relied on custom frameworks and workflows—such as harmonization trackers, pooled-level origin rules, and cross-study controlled terminology reconciliation—to bridge gaps. P21E supported implementation, but the methodology drove the outcome.

Lesson Learned: Understanding the strengths and limitations of P21E is essential for using it effectively. With thoughtful design, the tool can still significantly accelerate integrated eCRT development.

4. TRACEABILITY REQUIRES PURPOSE-BUILT REPORTING

Regulators expect clear traceability from each study into pooled SDTM. Achieving this required custom-built integration reports (Origin, CT, Derivation/Comments, Remapping) and intentional linkages in define.xml and the Reviewer's Guide. P21E served as a delivery mechanism, but the traceability model itself was a governance decision, not a tool feature.
Lesson Learned: Create custom integration-focused reports—such as Origin, Controlled Terminology, Derivation Comments, and Remapping Reports—to provide clear, study-level traceability. These should be linked proactively in both the Reviewer's Guide and the define.xml.

5. VALIDATION FINDINGS MUST BE INTERPRETED CAREFULLY

Automated validation findings tend to flag study-level discrepancies that are no longer relevant or meaningful in an integrated environment. Including such findings in the Reviewer's Guide may obscure the true purpose of the ISS eCRT and mislead reviewers.
Lesson Learned: Only integration-relevant findings should be carried into reviewer documentation. Thoughtful evaluation of validation messages is critical.

Overall, these challenges underscore a central lesson: **successful ISS-level SDTM integration relies as much on process, governance, and judgment as on tools.** With a well-designed workflow, a clear understanding of expectations, and a disciplined approach to metadata management, even tools not explicitly designed for integration—such as P21E—can be leveraged effectively to streamline the creation of transparent, traceable, regulatory-ready eCRTs.

CONCLUSION

Although integration activities are most performed at the ADaM level, teams find benefits integrating data in SDTM. When SDTM-level integration is necessary, the efficiency and quality of the overall eCRT development process can be significantly improved through the strategic use of P21E—even though the tool does not natively support the creation of integrated Define.xml files. The absence of built-in integration functionality does not diminish the value P21E provides; rather, it underscores the importance of understanding both the capabilities and limitations of the tool.

Success in using P21E for ISS-level eCRTs depends on the programmer's ability to clearly interpret the true scope of the integrated deliverables, including what must be represented in the eCRT and what level of traceability regulators expect to see. With a thorough understanding of the tool's metadata governance features—such as its handling of origins, controlled terminology, value-level metadata, comments, and merge-versus-overwrite behavior—P21E can meaningfully streamline the development of integrated metadata and accelerate the assembly of the ISS define.xml.

Equally important is the establishment of a well-structured process and a defined set of integration rules. By designing workflows that bridge P21E outputs with customized integration-level reports—such as Origin Reports, Controlled Terminology Reports, Derivation Comments Reports, and Remapping Reports—teams can provide true transparency across all contributing studies. These reports, when incorporated into the Reviewer's Guide and appropriately linked within the define.xml, offer reviewers a clear and traceable path from study-level sources to integrated SDTM outcomes.

However, validation findings from P21E must be interpreted judiciously. Not all automated issues are relevant in an integrated context, and only findings that legitimately pertain to pooled data or impact interpretability at the ISS level should be included in the Reviewer's Guide. Over-reporting study-level discrepancies can dilute the clarity of the integrated submission and obscure the essential context needed for regulatory review.

In conclusion, while P21E does not provide a dedicated solution for integrated define.xml creation, it remains a powerful and appropriate tool for ISS-level eCRT development when applied with expertise and supported by well-designed integration processes. With careful governance, thoughtful interpretation of outputs, and robust traceability workflows, P21E can and should be leveraged to produce high-quality, transparent, and regulator-ready integrated SDTM submissions.

REFERENCES

FDA U.S. Food & Drug Administration "Study Data Technical Conformance Guide" Study Data Technical Conformance Guide_v6.1_December 2025.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Mekhala Acharya for her unwavering support, guidance, and trust throughout the development of this process and the preparation of this paper. Her insight and partnership were instrumental in bringing this work to completion.

CONTACT INFORMATION

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

Author Name: Aida Swierk
Company: Takeda Development Center Americas Inc.,
Address: 650 E Kendall St, Cambridge, MA, 02142
Email: aida.swierk@takeda.com
Website: https://takeda.com/