Paper SA04

The Journey of Preparing an ISS

Lieke Gijsbers, OCS Life Sciences, ‘s-Hertogenbosch, The Netherlands

ABSTRACT

The integrated summary of safety (ISS) is a critical component of a submission to the FDA regulatory authority. For the ISS, data from different studies are pooled and harmonised to conduct the integrated analyses. Integrated SDTM/ADaM Define-XML and integrated Reviewer’s Guides (icSDRG and iADRG) may accompany the ISS to provide additional context and information about the integrated SDTM and ADaM datasets.

Based on a use case, we will explain in this presentation the approach we took to create the pooled datasets. Moreover, we will share our experiences regarding the creation of an SDTM and ADaM Define-XML for integrated databases as well as the icSDRG and the iADRG.

INTRODUCTION

The Integrated Summary of Safety (ISS) is a required document to be submitted to the Food and Drug Administration (FDA) when filing a New Drug Application (NDA) and it plays a crucial role in the regulatory review process for drug approval. For the ISS, integrated datasets are created by pooling and harmonising data from different studies. The integrated safety analyses are conducted based on these integrated datasets. Different strategies can be used to create the integrated datasets each having their pros and cons. The ISS may be accompanied by integrated Define-XML(s) and integrated Reviewer's Guide(s), offering supplementary context and information regarding the integrated datasets.

One of our clients requested whether we could support them with the preparation of ISS. For this client we previously converted more than 25 legacy phase I and II studies to CDISC standards. The phase I trials were converted to SDTM only, while the phase II trials were converted to SDTM and ADaM. Besides the >25 studies converted by us, also studies converted by other CROs had to be integrated, resulting in ± 40 studies to be integrated.

In this paper, the most common integration strategies and our integration strategy will be described and our experiences regarding the creation of ISS deliverables will be highlighted. Please note that all the information presented here is based on a specific use case and may not necessarily align with the requirements of regulatory authorities in a different submission.

MOST COMMON INTEGRATION STRATEGIES

There are three main integration strategies:

- ADaM integration only using individual study ADaM/legacy datasets (Figure 1)
- ADaM integration only using individual study SDTM/legacy datasets (Figure 2)
- SDTM integration using individual study SDTM/legacy datasets and ADaM integration using SDTM integrated datasets (Figure 3)

The creation of only integration ADaM datasets has the advantage that you do not need to create integrated SDTM datasets and accompanying documentation like SDTM Define-XML and Reviewer’s Guide, which might save resources and time. However, when using individual study ADaM datasets as a source, it is imperative that the individual study ADaM datasets are available. This is not applicable when starting from the individual study SDTM/legacy datasets and hence you could initiate the creation of integrated datasets earlier in the process, i.e. when the individual SDTM datasets are available. The usage of individual study ADaM or legacy datasets also requires a consistent analysis approach (i.e. variables derived in same way) and terminology across the individual study ADaM/legacy datasets. However, when this requirement is met, the creation of the integrated ADaM datasets could be fairly straightforward and it promotes traceability to the CSR tables and listings of the individual studies. The advantage of the third strategy, both SDTM and ADaM integration, is that the integrated SDTM datasets are the sole source for the integrated ADaM datasets which facilitates the ADaM programming and review of the datasets. Moreover, listings could be directly created from the integrated SDTM datasets and the integrated SDTM datasets could also be used for the creation of analysis datasets for other purposes, e.g. the creation of the annual Investigator’s Brochure.

More information about the integration strategies and its pros and cons can be found in the white paper published by PHUSE[1].

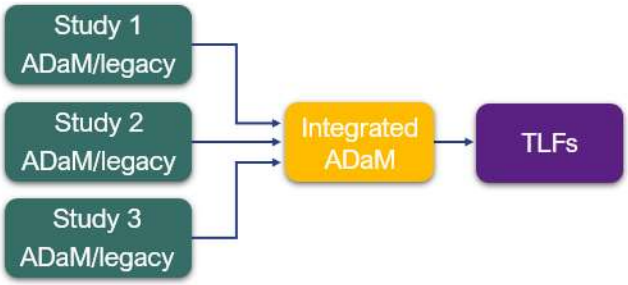

Figure 1. ADaM integration only using individual study ADaM/legacy datasets

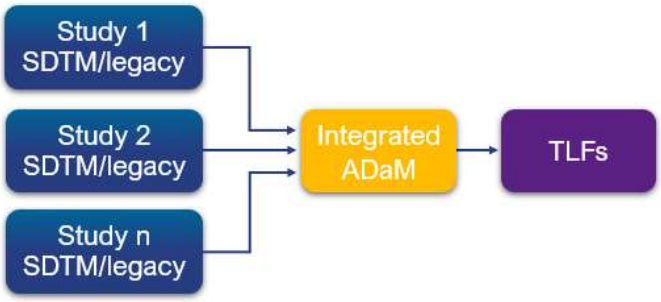

Figure 2. ADaM integration only using individual study SDTM/legacy datasets

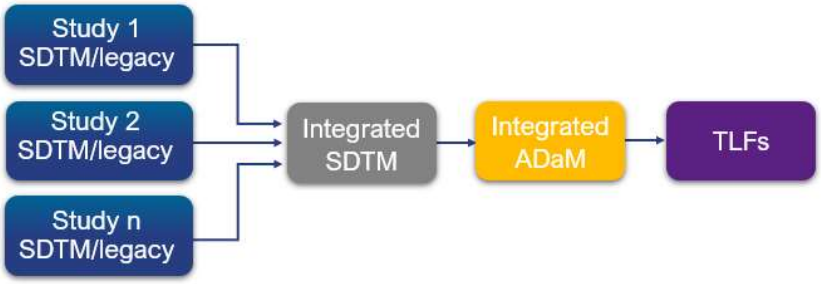

Figure 3. SDTM integration using individual study SDTM/legacy datasets and ADaM integration using SDTM integrated datasets

OUR INTEGRATION STRATEGY

Our integration strategy was designed by the client with the guidance of the regulatory agency and described in the Submission Data Management Plan (SDMP) which included the Study Data Standardization Plan. In the SDMP, amongst others, the integration strategy, the studies to be integrated, validation plan and the eCTD M5 folder structure including those for the ISS files were described.

Our integration strategy was as follows: pooled SDTM datasets AE, CM, DM, EX, MH and RP were created from the individual study SDTM datasets for ± 40 studies. From the pooled SDTM datasets containing data for ± 40 studies, subset datasets (merely a selection of studies) were created for the purpose of ISS. The subset SDTM datasets (in next sections referred to as integrated SDTM datasets) included data for two pivotal phase III trials and were only created for SDTM domains AE, DM, EX and RP.

Subsequently, pooled ADaM datasets ADSL and ADAE were created from the pooled SDTM datasets containing ± 40 studies and from individual study analysis datasets ADSL for selected variables. From the pooled ADaM datasets, subset datasets ADSL, ADAE and ADTTE (time-to-event) only containing data for the two pivotal phase III trials were created for the purpose of ISS. In the next sections, these subset datasets are being referred to as integrated ADaM datasets.

From the integrated ADaM datasets (only containing data for the two pivotal phase III trials), nearly 60 integrated safety tables and listings were created,

The integration strategy is also depicted in the data flow diagram in Figure 4. Items in the dashed border are submitted to the FDA.

In the next sections, the creation of the ISS deliverables is explained in more detail.

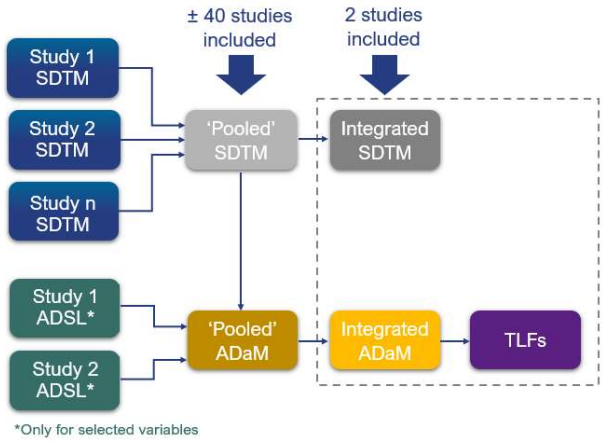

Figure 4. Applied integration strategy

POOLED SDTM DATASETS

The ± 40 studies included in the pooled datasets were a mix of phase I, II and III studies, with the first study started in 2006 and the last one in 2023. The majority of phase I and II studies were converted to CDISC standards by us, while the other studies were converted by other CROs.

In the individual study SDTM datasets, different versions of SDTM (Implementation Guide) and CDISC Controlled Terminology were used. During the creation of the pooled SDTM datasets, these versions were aligned according to the versions used in one of the pivotal phase III trials. For SDTM, this entailed Implementation Guide version 3.2. Also, the Adverse Events (AE), Medical History (MH) events and Concomitant Medications (CM) were recoded using the same MedDRA and WHODrug dictionary version because in the individual studies these were coded according to different version of MedDRA (for domains AE and MH) and WHODrug (for domain CM).

As the individual studies were converted to SDTM by different CROs, sponsor-defined terminology was also harmonised in the pooled SDTM datasets. For example, the values for variables (ACT)ARM/(ACT)ARMCD, EXTRT, AEREL were aligned, but also the values for QNAM and QLABEL for supplemental qualifier variables.

Moreover, the variables (ACT)ARM/(ACT)ARMCD were also rederived to adhere to the FDA Study Data Technical Conformance Guide[2], which states that 'Screen Failure', 'Not Assigned', and 'Not Treated' should not be specified as treatment arm. Hence these variables were set to missing in the pooled SDTM dataset DM when the individual study variables contained any of these values.

In the pooled SDTM datasets, only variables required or expected as per CDISC guideline and variables needed for the analyses were kept. Because of this approach, the pooled SDTM datasets were lean. As mentioned in the section "Our Integration Strategy", the actual integrated SDTM datasets containing data for two studies were created from the pooled SDTM datasets. The same variables were kept in the integrated SDTM datasets as in the pooled SDTM datasets.

POOLED ADAM DATASETS

When creating (pooled) ADaM datasets, you would refer to the Statistical Analysis Plan (SAP) and TLF mock shells for the specifications on how variables required for the analyses need to be derived, if and how data needs to be imputed, etc. For this ISS, no SAP was created, and hence upon the client's request we mainly followed the rules defined in the SAP of one of the pivotal phase III trials. However, not all rules applicable to the ISS were defined in the pivotal phase III trial SAP and hence quite a few conversations and iterations were needed to come to the final ADaM specifications, even though mock shells for the tables and listings (see section Integrated Safety Tables and Listings) were available.

The pooled ADSL consisted mainly of treatment variables like start and end dates and planned and actual treatments per treatment period and variables needed for subgroup analyses like race, ethnicity, and BMI. Particularly the creation of the treatment variables was rather challenging because the ± 40 studies included had different studies designs (e.g. cross-over vs. parallel, different number of treatment periods, studies in which co-administrations were given) and logically the pivotal phase III SAP did not state the rules for these different scenarios.

In the pooled ADAE dataset, besides treatment-emergent AEs (TEAEs), AEs of Special Interest (AESIs) were flagged. For the flagging of the TEAEs, the rules defined in the SAPs of the pivotal phase III trials were used. For the flagging of the AESIs, an Excel spreadsheet containing the MedDRA preferred terms pertaining to pre-identified AESIs was used to flag qualifying events. Also, more than 50 "first occurrence flags" were included in pooled ADAE

to facilitate the creation of pooled ADTTE, e.g. 1st Occurrence of Preferred Term Flag, 1st Occurrence of Maximum Severity within Preferred Term Flag and 1st Occurrence of First Serious TEAEs Related to Study Treatment Flag.

Pooled ADTTE was created to store the exposure duration until the occurrence of a certain event (e.g. TEAE, serious TEAE related to study treatment, or TEAE leading to death). This exposure duration was needed for the calculation of the exposure-adjusted incidence rate (EAIR)[3], which had to be displayed in the tables.

As mentioned in the section “Our Integration Strategy”, the actual integrated ADaM datasets containing merely data for two studies were created from the pooled ADaM datasets.

INTEGRATED SAFETY TABLES AND LISTINGS

At the start of the project mock shells for the unique tables and listings were developed. Using the mock shells, nearly 60 integrated safety tables and listings were created from the integrated ADaM datasets. In the tables, the results of the safety analysis for the two studies had to be displayed side-by-side, because an actual integrated analyses of the two studies was not deemed suitable by the regulatory agency. In Figure 5, an excerpt from a table mock shell is displayed.

Table 1: Treatment-emergent adverse events (TEAEs) by system organ class and preferred term

	Study 1					Study 2				
	Active dose 1 (N=XXX)		Placebo (N=XXX)		EAIR diff. est.	Active dose (N=XXX)		Placebo (N=XXX)		EAIR diff. est.
System Organ Class Preferred Term	n (%) [m]	EAIR	n (%) [m]	EAIR	EAIR diff. (95% CI)	n (%) [m]	EAIR	n (%) [m]	EAIR	EAIR diff. (95% CI)
TEAEs	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
System Organ Class 1	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
Preferred Term 1	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
Preferred Term 2	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
System Organ Class 2	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
Preferred Term 1	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)
Preferred Term 2	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)	xx (xx.x) [xx]	xx.x	xx (xx.x) [xx]	xx.x	x.xx (x.xx;x.xx)

EAIR = Exposure-adjusted incidence rate.

Figure 5. Mock shell for side-by-side safety table

SDTM AND ADAM DEFINE.XML AND DEFINE.PDF FOR INTEGRATED DATASETS

For the creation of the Define-XML for integrated SDTM and ADaM datasets, the general Define-XML v2.0 specifications were used. When comparing the Define-XML for the integrated datasets with an individual study Define-XML, there are a few general differences. One of the differences is the information provided in the general study information section. As our integrated datasets only contained two studies, we decided to mention both studies and to refer to the ISS in this section, see an excerpt from the general study information section from our integrated SDTM Define-XML in Figure 6. Another difference is that we did mention the STUDYID in the method or comment in the Define-XML when the method or comment was conditional to one of the studies. This is of course something that is not needed in a Define-XML based on a single study.

Standard	SDTM-IG 3.2
Study Name	<study 1> and <study 2>
Study Description	Integrated Summary of Safety for <drug> in <indication>
Protocol Name	ISS <indication>
Metadata Name	Study <study 1> and <study 2> Data Definitions

Figure 6. General study information section in ISS SDTM Define-XML

The specific difference between an integrated SDTM Define-XML and a study SDTM Define-XML is that there is no reference to an annotated CRF in any of the origins and that many origins were set to “Predecessor”. This is because many values in the integrated SDTM datasets were directly copied from the individual study SDTM datasets, including the values for study day (--DY) variables, (R)DOMAIN and STUDYID. Pinnacle 21 raised issues about this (see Figure 7), but as these could not be resolved, these were explained in the Reviewer’s Guide (see section Reviewer’s Guides for Integrated Datasets). Besides the general differences, no differences compared to a study ADaM Define-XML were found in the creation of a Define-XML based on integrated ADaM datasets.

Pinnacle 21 Validator Report

Issue Summary

Source	Pinnacle 21 ID	Message	Severity	Found
DM				
	SD1349	Inconsistent STUDYID		244
DEFINE				
	DD0102	Invalid Annotated CRF document name		1
	DD0105	Origin for Study Day variable 'AEENDY' is not set to Derived		1
	DD0105	Origin for Study Day variable 'AESTDY' is not set to Derived		1
	DD0105	Origin for Study Day variable 'DMDY' is not set to Derived		1
	DD0105	Origin for Study Day variable 'EXENDY' is not set to Derived		1
	DD0105	Origin for Study Day variable 'EXSTDY' is not set to Derived		1
	DD0105	Origin for Study Day variable 'RPDY' is not set to Derived		1
	DD0106	Origin for DOMAIN variable is not set to Assigned		4
	DD0107	Origin for RDOMAIN variable is not set to Assigned		2
	DD0108	Origin for STUDYID variable is not set to Protocol		6

Figure 7. Pinnacle 21 report for SDTM Define-XML and integrated SDTM dataset DM

For ease of reading, the FDA requested to submit a Define.pdf in addition to a Define.xml. The Define.pdf has been created following the method described by Brian Mabe[4], which makes use of the Java application Apache™ Formatting Objects Processor. This application in combination with the XSL stylesheet easily converts a Define.xml to Define.pdf and retains active hyperlinks and bookmarks, which are required in Define.xml but also in Define.pdf. In Figure 8, an excerpt of a Define.pdf for the integrated SDTM datasets is shown.

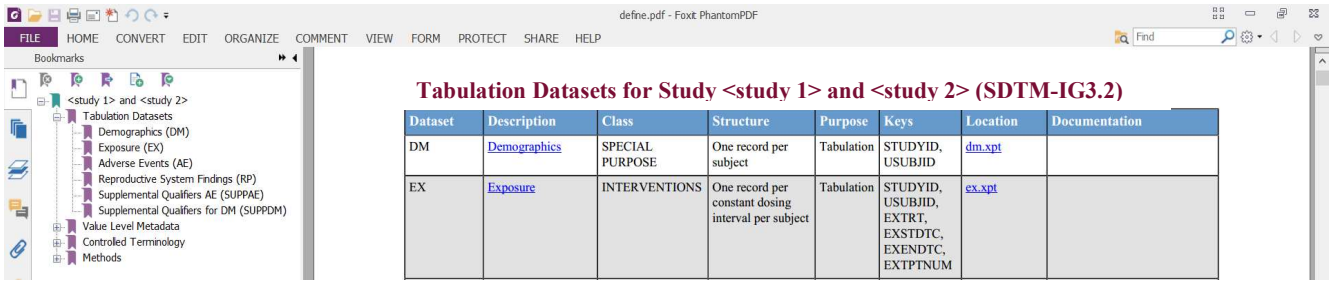

Figure 8. Define.pdf for integrated SDTM datasets

REVIEWER’S GUIDES FOR INTEGRATED DATASETS

The integrated clinical Study Data Reviewer’s Guide (icSDRG) and integrated Analysis Data Reviewer’s Guide (iADRG) are recommended as an integral part of an NDA submission to provide additional context to, respectively, the integrated SDTM and ADaM datasets. At the time of writing of the icSDRG (and also this paper) no template or guidelines are available for the icSDRG and therefore we created a proprietary version. Our icSDRG provides information on:

- Study data standards and dictionary versions
- Integrated studies, e.g. study design and treatments
- Integration strategy including traceability flow diagram
- Overview of integrated datasets and any special considerations, e.g. recoding
- Data conformance summary including explanation of unresolvable Pinnacle 21 issues

For the iADRG, we used a preliminary template created by PHUSE. In addition to the icSDRG, the iADRG provides information on the core variables and analysis considerations like the derivation of the treatment variables, imputations applied, flagging of AESIs and data dependencies. The iADRG also listed the programs including macros used for the creation of the integrated ADaM datasets and tables and listings. At the time of the writing of this paper, PHUSE published the iADRG template, completion guidelines and example[5].

KEY MESSAGES

As always, communication is the key. As no ISS SAP was available, quite some time has been spent on determining the final ADaM specifications. Having an ISS SAP would most likely have saved time and a lot of back-and-forth communication, and hence I would recommend creating an ISS SAP for future ISS projects. Even though the creation of the Define.xml/Define.pdf and Reviewer's Guides for integrated datasets was rather new to the project team, it was concluded by the team that the creation of these supporting documentation was rather straightforward and very similar to the creation of these documents for individual studies.

ACKNOWLEDGMENTS

I would like to thank Jules van der Zalm and Mitchikou Tseng from OCS Life Sciences for their support in writing this paper.

REFERENCES

1. PHUSE, Integration Strategies in Support of ISS/ISE Submissions, Version 1.0, Document ID: WP-013
2. FDA, Study Data Technical Conformance Guide - Technical Specifications Document, October 2023
3. He X, Chen L, Lei L, Xia HA, Lee MLT (2015) A Simple Method for Estimating Confidence Intervals for Exposure Adjusted Incidence Rate and Its Applications to Clinical Trials. J Biom Biostat 6: 238
4. Brian Mabe, Define.XML to Define.PDF, Just F.O.P It, PHUSE US Connect 2018
5. PHUSE, Integrated Analysis Data Reviewer's Guide Package, Version 1.0, September 2023

CONTACT INFORMATION

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

Author Name: Lieke Gijsbers
Company: OCS Life Sciences
Email: lieke.gijsbers@ocs-consulting.com