Authors:
- Ghanshyam Jangid
- Jane Marrer

Affiliation: Merck & Co., Inc., Kenilworth, NJ, USA

Processes and Techniques for creating Integrated Summary of Safety (ISS)

Abstract

There are unique challenges when working on Integrated Summary of Safety (ISS) for compoundswith many clinical trials contributing to the investigational product's overall safety profile. Harmonizing datasets and conducting ISS analyses with from multiple sources, including vendors, and ensuring that each integrated dataset is CDISC-compliant requires careful planning and execution. Integrated analyses must align with what is reported in the CSR related to the pivotal trial in a submission. Integrated datasets must be submitted in a common version of dictionaries such as MedDRA. Variable attributes must be harmonized while ensuring no truncation occur. This paper will discuss various process and techniques for efficiently integrating ADaM datasets for ISS. We will also discuss about how to handle various challenges during development of integrated datasets.

In this paper, we will explain ways to address the unique challenges for Integrated Summary of Safety (ISS) packages for a compound with multiple trials having ongoing submissions. An example of this is an oncology compound which has submissions for multiple tumor types or indications. This type of compound can do ongoing submissions over many years. To support this type of situation, we will suggest processes and techniques which can be used to efficiently build and maintain a reusable library of reference safety datasets that reflect the compounds safety profile. We will also discuss methods to harmonize datasets across the studies to support Integrated analysis with multiple datasources and ensure harmonization of variable attributes across multiple trials; compliance of integrated ADaM datasets; alignment with pivotal trial with the Clinical Study Report (CSR) and the use of common dictionaries. Unique challenges that are addressed in this paper include the extensive resources needed to harmonize data across many trials over a long period of time; standardizing laboratory data; extensive, complex documentation for large ISS datasets; review of a large number of tables; and data truncation.

TOC

- Summary

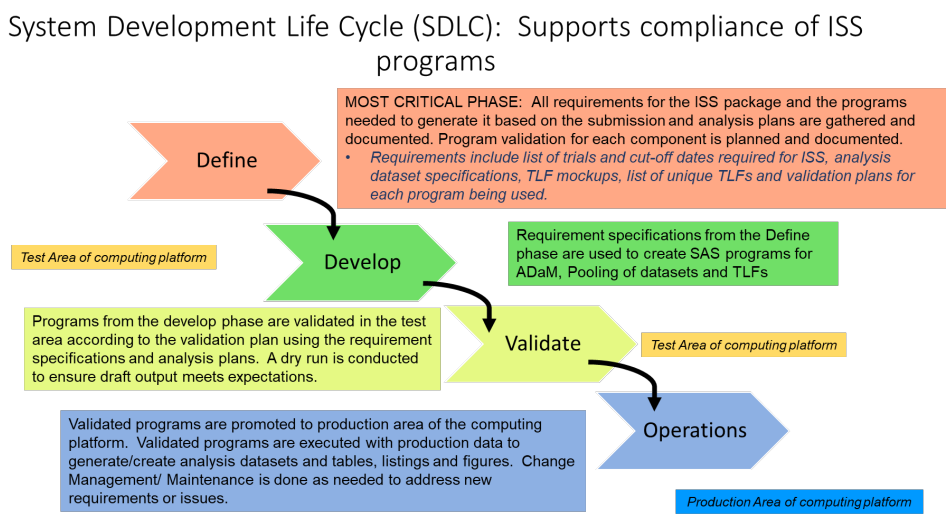
Reference Safety Dataset Library Overview and Advantages

The challenge of extensive resources to support integrated analysis and harmonize data across many trials over a long period of time uses can be addresses in a variety of ways. One major component to less resource use is to create a library of re-usable validated datasets. Having ready-to-go datasets will avoid a re-work for each individual submission. In addition to achieving consistency and quality, a Reference Safety Dataset (RSD) Library of Submission- Ready ADaM datasets is an important component for sharing the safety profile of a compound. The population in the RSD library should be defined by Regulatory. Clinical trials should be included in the library based on their submission and approval status. Once the set of trials in the RSD are defined, the library is built by creating and validating submission ready ADaM datasets for each individual trial in the RSD Library. In order to be as efficient and consistent as possible, it is important to begin the development of the RSD datasets using the analysis datasets that were used to support the Clinical Study Report (CSR) for each individual trial. Once a dataset is created, it should be stored in a central location in a sub-folder specific to the MedDRA version in place at the time the dataset is created. Once the validated RSD datasets are available, they can be pulled into the integrated datasets required for each individual submission for the compound.

The RSD library will need to be maintained and updated over time for a variety of reasons. Updates include but are not limited to appending the library with additional trials which have been added to the RSD for the compound; modifying the ADaM datasets to align with updates to industry or company standards and leveling dictionary terms to current versions.

The departmental programming SOPs should be followed when creating or updating the ADaM datasets in the RSD Library. In general, all program development should be done using a system development lifecycle (SDLC) process. In this process there should be at least 4 stages, Define, Develop, Validate and Operations. The diagram below shows how a basic SDLC process supports the compliance of the programs used to support all the integrated summary of safety (ISS) deliverables. For an ISS, the Define is the most critical phase as requirements can be complex. In addition to ADaM datasets and table, listing and figure (TLF) requirements, ISS requirements should include logic for normalizing data across studies and a list of the trials or cohorts/populations needed and the cut-off dates to be used for each trial/cohort. Dictionary versions should also be identified during requirements gathering

During the development and validation phases, you should not forget to develop and validate programs to pool the datasets to create integrated analysis datasets

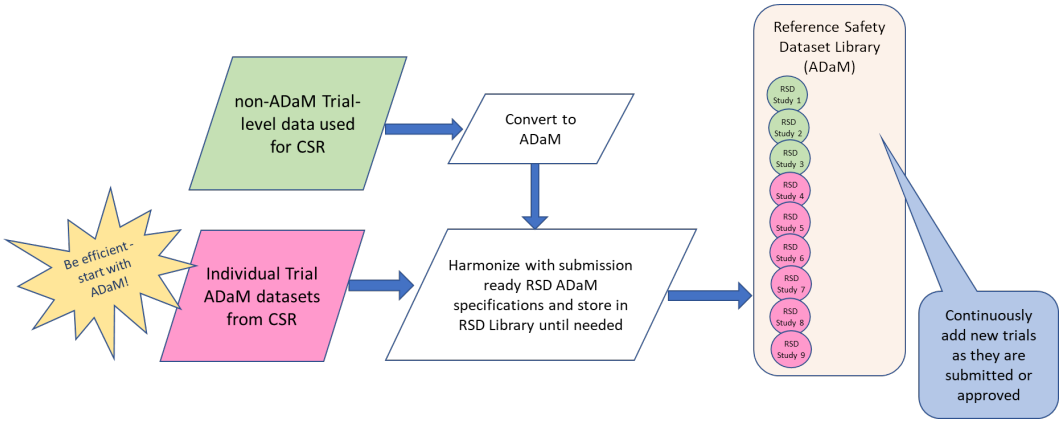


Reference Safety Dataset (RSD) library creation and maintenance

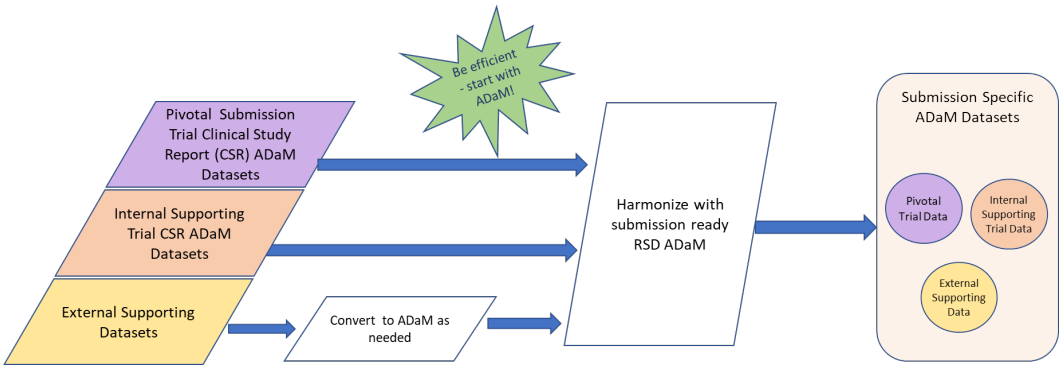
In this paper, we broke the process to create and maintain an RSD library into 4 steps

1. Create the first RSD Library of submission ready datasets containing all the trials identified for inclusion in the compounds RSD population
2. Create and harmonize submission ready ADaM datasets for additional trials identified to support a specific submission
3. Pool submission specific ADaM datasets with Reference Safety Datasets
4. Update and Maintain the RSD library

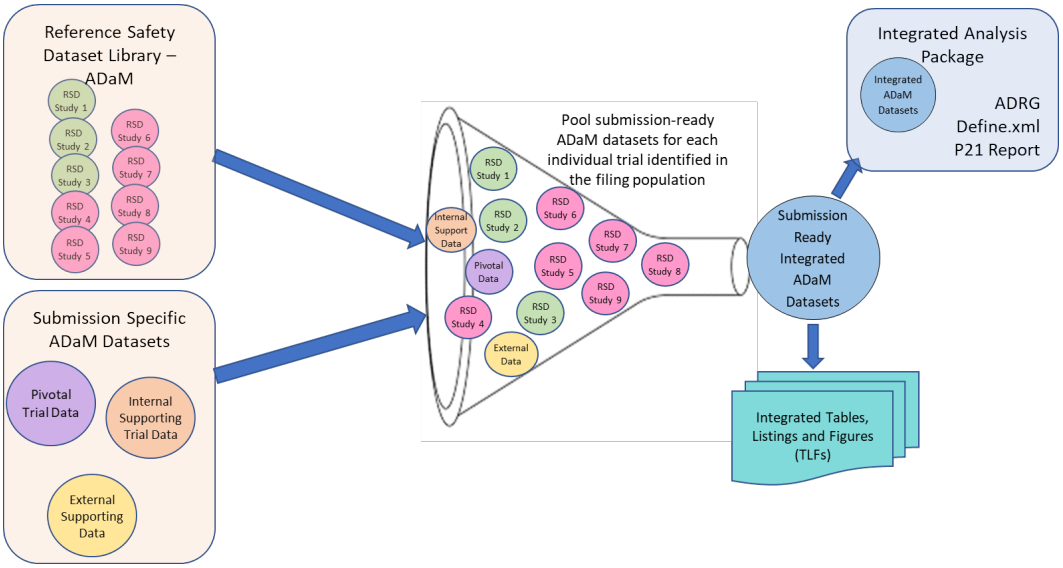
In step 1, the initial library is built by harmonizing the data in each individual trial to allow for pooling of data. Whenever possible, you should begin with ADaM from the trial CSR, however for older legacy trials, ADaM datasets may need to be created. Once each dataset is created, it should be stored in the central location identified for use by the ISS team. As new trials are identified for the RSD, they should be normalized and added to the RSD folder. For traceability purposes, create a tracker document containing details about the source datasets in the RSD library. Datasets used to support integrated safety populations generally includes ADSL, ADAE, ADEX, ADLB and ADCM.



In step 2, the datasets needed to support a specific submission are created and harmonized so that they align with the RSD format and content. This includes the pivotal trial/s for the submission as well as supporting data from internal or external trials which contain data supportive of the submission indication. Once again, ADaM should be used when possible.

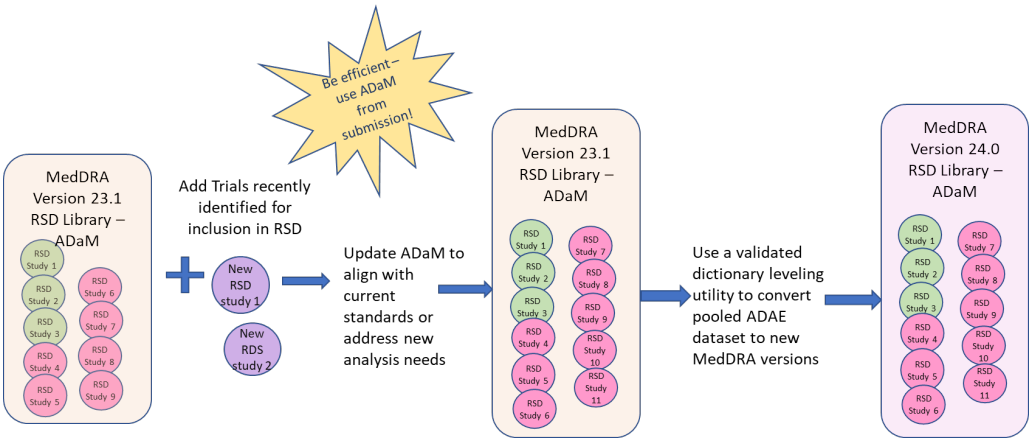


Step 3 is where the submission specific ADaM datasets are pooled with the RSD datasets. Since all the datasets have been harmonized to align, the pooling require minimal effort and result in submission ready integrated ADaM datasets which can be used to create Integrated TLFs and an Integrated Analysis Package for US submissions.



Step 4 covers on-going maintenance for the RSD Library. As new trials are added to the compound safety profile, the trials ADaM datasets should be added to the library. ADaM standards must be maintained so all the trials in the RSD will occasionally need to have the ADaM updated to align with current standards. Additionally, in order to support submissions over a period of many

years, the datasets must be leveled to the current version needed. It is recommended to perform all the necessary updates at the same time to achieve efficiency and better consistency. We have found the best timing for doing this is every 6 months as MedDRA versions are released.



Reference Safety Dataset Library Uses and Integrated Analysis Populations

A Reference Safety Dataset Library can be used to support integrated analysis for multiple purposes. Integrated analysis is needed for eCTD modules 1 and 5. The specific modules are 1.1.13 (DSUR), 1.14 (Labelling and Investigational Brochure), 1.16 (Risk Management Plan) and 5.3.5 (Reports of analyses of data from more than one study (ISS)). Additionally, the RSD can support internal Aggregate Safety Assessment and Agency Requests.

Integrated Analysis Populations can be very complex and must be carefully planned. Specifications must clearly define all Trials to include and their cut-off dates in each population. For trials supporting multiple indications or tumor types, indication specific cohorts and cut-off dates must also be defined. The RSD library should be designed so that is contains all the trials needed to support the many unique populations supported in integrated safety analysis. Once each unique population is defined, select the subset of trials identified for inclusion in a unique population.

Integrated Analysis Packages or Populations may include

- Indication-Specific Populations containing a pivotal submission trial plus other supportive trials in the same indication
- Established Reference Safety Data Populations containing trials and cohorts submitted or approved in a specific country or region (i.e. US or EU). These generally have defined cut-off dates.
- Cumulative Running Safety Populations containing all trials included in prior submissions PLUS the indication specific trials for a new submission.

A sample table header and footnote for how these populations might display in analysis TLFs is below

Protocol 005 (N=x)	Melanoma specific (protocol005+protocol 001) (N=x)	US Reference Safety Dataset for Compound (trials in defined safety profile) (N=x)	Cumulative Running Safety Dataset for Compound (all submitted trials plus columns 1 & 2) (N=x)

Data Harmonization – Best Practices

Key data points impacting the primary and secondary objectives of overall trials must be harmonized prior to pooling of the datasets. The programmer must ensure that the counts in the ISS tables (column 1 above) align with the counts in the pivotal trail's CSR. In order ensure this happens, it is important to align on input data, SOPs, and standards. Standards and templates to support the efficient, high quality and consistent RSD components are critical to this alignment. For quality harmonization, it is important to check the datasets prior to pooling and make sure they are consistent. Checks should include ensuring that:

- Variables with same name across individual trials have the same attributes.
- Variable types, length, formats, and labels.
- Consistent programming logic and categorial variables are used across trials.
    - PROC FREQ procedures can be performed when combining data in integrating datasets.
- All CODE and DECODE variables maintain a one-to-one mapping.
- Special handling that might be needed for specific variables in a given trial.
- Truncation that may occur when stacking datasets across studies for variable with different lengths in individual trials
    - To address this, create code to find the maximum real values for all variables and set the length for each variable before stacking up the datasets

If needed, additional variables can be created to handle inconsistent situations or to filter any records of interest. A table look-up approach can be used to maintain consistencies across all pooled ADaM datasets. This can help to address the challenge of standardizing lab parameters when laboratory data comes from different local and central laboratories. Creating a look-up to align lab test names and codes can help to address this. For lab data, it is also important to carefully review derivations and harmonize critical variables such as mcirt1, mcrit2, Analysis flag etc. across different trials.

Other important items are important to ensure alignment between ISS datasets includes:

- Application of trial cut off dates in requirements and in program code
- Adherence to company SOPs to develop and validate programs
- Adherence to company and agency standards and templates

Templates and Standards

Using standards will ensure compliance, quality and efficiency. Industry standards and company specific standards are equally significant to an ISS team. An ISS team must use the Technical Conformance Guidance; current versions of ADaM and ADaM Implementation Guidelines and standard dictionaries such as MedDRA and WHO Drug. Additionally, it is required that all trials

in an ISS dataset be in a consistentversion of MedDRA. *Note that it is permissible to submit in mixed WHO DD versions.*

Standard Specifications and Planning Templates are good tools to use for ensuring alignment across the multiple submissions for a compound. Examples of specifications or templates that can be used include ADaM Dataset Specifications; TLF Mockup/Shell templates, tracking tools for documentation requirements, development, and validation of TLFs and programs. Additionally, for traceability, all the datasets in the RSD Library should be tracked.

Standard re-usable validated Programs/Macros/Utilities are invaluable to support consistency. A library for ISS should contain validated programs and utilities which can be used for a variety of tasks including but not limited to harmonization of datasets; pooling datasets; generation of integrated TLFs and converting/leveling dictionaries to a common version. Departmental standards should be followed for datasets and TLFs. Teams should avoid creating project specific versions of ISS programs. For high efficiency, consistency and quality, it is recommended to make use of departmental standards. Integrated data can be pre-processed so that it can be used as input to standard programs.

Validation activities can also be supported by standard programs and utilities. Count checks and data harmonization checks are a valuable tool for an ISS team. The ISS process requires extensive documentation to explain how data is transformed and standardized. This is critical to showing data traceability for reviewers. Having template tools to track work and the data will help to address this challenge. eSubmission templates such as ADRG and define should be created for the analysis package to ensure the documentation is consistent and accurate.

QC Activities

QC to ensure quality harmonization is an important step to ensure successful integration of datasets. Count checks and TLF output review are key to quality. Combining all TLF outputs into one single RTF for internal review instead of reviewing each unique output individually can help to ensure consistency and quality across TLFs.

Count checks are critical to submitting a consistent safety profile over time for a compound. Counts in summary tables must match historic analysis. Checks should be done to make sure summary table counts for a new submission align with the same data used in previous integrated packages. If a trial appears in a historic ISS with the same date cut off, all numbers in summaries should match. The pivotal trial summary and the supportive trials summaries in the ISS must match the summaries in the submitted CSRs. This is one reason for using the CSR ADaM datasets as input for the ISS datasets. Another example for checking counts for AE summary related tables can be broken by trial/cohort as following:

- AE Summary
- AE Summary for terms of special interest
- AE Summary by baseline characteristics
- AE Summary by Demographic information (Age, Sex, Region, etc.)

Consistency checks between current ISS and previous ISS, can be supported by using excel files

with important information in conjunction with macros. The following should be checked:

- Data Cut Off against data pooling plan
- Table/Population counts against previous ISS/CSR
    - Number of subjects per trial
    - Number of AE/SAE/DEATH/AEOSI, etc.
    - Number of AE by Demo category
- Consistency check between US and EU table sets

A programmer should randomly check 5 -10 files and ensure that expected population counts changes in the respective columns. Additionally, they should ensure there are no surprises in the Controlled Terms for each trial. To do this, they can scan all related variables (i.e. AGEU, SEX, RACE, REGINE) and list all the distinct values in the pooled dataset, The programmer needs to handle the unexpected cases accordingly to avoid mistake in summary tables (i.e. Upper case/lower cases, Leading space, Missing values).

Dictionary Leveling and Conversions

Dictionary Leveling

As stated earlier a utility to level dictionary terms to a new version can save significant time for a reference safety dataset especially with a large number of trials. The validated pooled ADAE dataset can be quickly converted to a new dictionary version through the utility so that all reference safety trials are readily available latest version in the library. If there are flags for adverse events of interest, the assignment of the flags should be re-visited after each MedDRA release and related flags needs to be updated after MedDRA leveling is completed. QC of the leveled datasets is important. QC steps include:

- Conduct QC to Compare the old adae dataset with the new adae dataset
- Compare all encoded terms for values before and after (see list below)
- Ensure no blanks for previously encoded terms
- Check that new terms match and make sense (i.e., no garbled preferred terms)
- Make sure no records were added or removed

adae	AEBODSYS
adae	AEDECOD
adae	AEHLGT
adae	AEHLT
adae	AELLT

Once leveling is complete, the updated datasets should be stored in a central location in a folder indicating the new MedDRA version. Old MedDRA version datasets should be saved to ensure traceability and available for support of future requests.

Dictionary Conversions

If there are company specific dictionaries that will no longer be accepted for agency use, the ISS team must convert the company specific dictionary to a global standard. For example, legacy studies with concomitant therapy terms coded in a company specific drug dictionary instead of WHO Drug Dictionary (WHO-DD). In this situation when the pivotal trials are required to submit in WHO-DD, the RSD datasets in the company drug dictionary must be converted to WHO-DD prior to pooling to ensure compliance with regulatory agencies.

One way to convert legacy studies to a standard dictionary is to convert the RSD ADaM datasets to avoid going back to the source data for the conversion. A mapping can be created for each unique term in the RSD dataset of interest. The mapping will assign legacy terms to standard terms. To keep things efficient, terms can be limited to those being used for integrated analysis (i.e. concomitant medication of interest for integrated safety analysis). The mapping can be used to programmatically convert terms in legacy dictionary to standard dictionary. The dataset with the converted terms should have a different name than the source dataset in the legacy dictionary to avoid confusion or use of the wrong data. If additional legacy studies are added to the RSD, it is recommended to review unique terms and re-create mapping as needed to accommodate new unique terms.

Summary

In summary, there are unique challenges to conducting integrated analysis for a compound with multiple submissions occurring over a span of multiple years. Establishing a library of reference safety datasets will ensure a consistent ability to efficiently create high-quality integrated analysis which accurately reflects the safety profile of the compound. Having a defined package of deliverables and following established processes and techniques will increase efficiency, quality, and consistency. Templates and standardized tools and macros/programs are essential for harmonizing data across studies and leveling/converting dictionaries.