Paper SA04

Making SDTM/ADaM CRT Packages Submission Ready

Phillip Spowart, Fortrea, Leeds, United Kingdom

ABSTRACT

When submitting a compound, Regulatory bodies require SDTM and ADaM datasets to be an IG version that is accepted as per their standards catalogue. With big compounds we have seen studies where datasets are not in an SDTM/ADaM format or the SDTM/ADaM IG is in a version no longer accepted. This paper aims to cover how to convert Legacy datasets and datasets in an older version of IG into an accepted version of IG. This will cover how the SDTM and ADaM datasets become submission ready CRT packages, which can then be submitted and used to create an integrated database.

INTRODUCTION

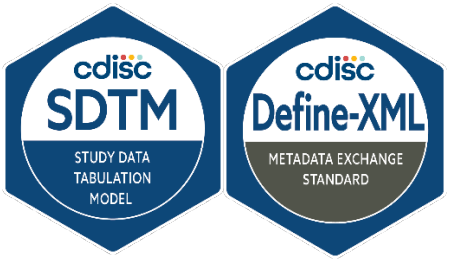
Study Data Tabulation Model (SDTM) Case Report Tabulation (CRT) packages and Analysis Data Model (ADaM) Case Report Tabulation packages, go hand in hand with any clinical study. Since 2016 regulatory bodies have required data to be submitted in these formats.

Most studies, over the last 10 years, have presented data in SDTM and ADaM. However, there are compounds, yet to be submitted, with data not currently in SDTM and/or ADaM. The data in these studies are referred to as legacy data. Legacy conversion is the name of the process used to convert the legacy data into SDTM and ADaM, allowing these studies to become compliant and submission ready.

SDTM and ADaM have both been adopted as the standard within the last 20 years. The clinical world is a very proactive industry, this has led to multiple versions of implementation guide (IG) for SDTM and ADaM. SDTMs are currently on v3.3, which will become v3.4 15MAR2025, and ADaM v1.3. With the new versions of IG being released, regulatory bodies will eventually stop supporting prior versions of the IG. Due to this, the initial studies for a compound may be on a version of IG that is no longer supported. The process to convert these studies to a supported IG version is called up-version.

A define.xml is always required for SDTM CRT and ADaM CRT packages. There are now 3 versions of define.xml, 1.0, 2.0 and 2.1. The IG version for the SDTM and ADaM datasets may be supported, but the define version may not. When this occurs, the define.xml would need an up-version to make it submission ready.

WHAT ARE SDTM CRT AND ADAM CRT PACKAGES?



A submission ready SDTM CRT package contains:

1. Annotated CRF (aCRF) – The aCRF should be compliant with the SDTM Metadata Submission Guidelines (SDTM-MSG). The annotations within the aCRF should then link with variables presented in the define.xml.

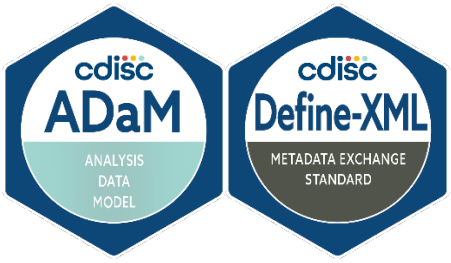
2. SDTM datasets – The SDTM datasets used for creating ADaM and TFLs may be in a sas7bdat format. However, the SDTM datasets need to be xpt files when provided in the CRT package. The SDTM datasets also need to be provided in an SDTM IG version that is supported by the regulatory body.

3. Define.xml and stylesheets – the define.xml should be provided in a define.xml version supported by the regulatory body. The supporting files that the define.xml links to are also required to be submitted.

4. Pinnacle21 reports – Pinnacle21 community or enterprise should be used to validate the SDTM datasets and the define.xml. The engine used should be supported by the regulatory body.

5. Clinical Study Data Reviewer's Guide (cSDRG) – It is advisable to use the cSDRG template[1] on the PHUSE Advance Hub. The five sections should be appropriately updated, ensuring it accurately reflects everything that occurred on the study.

The SDTM CRT package can also include the SDTM programs and the define.pdf. It is best practice to have these stored with the SDTM CRT package, so they are easily accessible if requested.



A submission ready ADaM CRT package contains:

1. ADaM datasets – The ADaM datasets used for creating TFLs may be in a sas7bdat format. However, the ADaM datasets need to be xpt files when provided in the CRT package. The ADaM datasets also need to be provided in an ADaM IG version that is supported by the regulatory body.

2. Define.xml and stylesheets – the define.xml should be provided in a define.xml version supported by the regulatory body being submitted to. The supporting files that the define.xml links to are also required to be submitted.

3. Pinnacle21 reports – Pinnacle21 community or enterprise should be used to validate the ADaM datasets and define.xml. The engine used should be supported by the regulatory body.

4. Analysis Data Reviewer's Guide (ADRG) – It is advisable to use the ADRG template[2] on the PHUSE Advance Hub. The eight sections should be appropriately updated, ensuring they accurately reflect everything that occurred on the study.

5. ADaM programs – The programs used to create the ADaM datasets should be included, these should also be mentioned in section 7 of the ADRG.

6. Other data sources – Any non-SDTM datasets used as input to an ADaM dataset should be included. An example of this could be the non-compartmental analysis (NCA). Note the NCA can be included in an ADaM per the Analysis Data Model Implementation Guide for Non-compartmental Analysis Input Data[3].

7. Analysis Results Metadata (ARM) – ARM is only required to be included if performed on the study.

The ADaM CRT package can also include the define.pdf. It is best practice to have this stored with the ADaM CRT package, so it is easily accessible if requested.

REGULATORY REQUIREMENTS

When a compound is submitted a regulatory body:

- The data for each study being submitted should be included in an SDTM CRT package or an ADaM CRT package.
- The SDTM and ADaM datasets being submitted, should adhere to an IG version that is supported by the regulatory body, unless an exemption/waiver is granted.
    - Standards catalogues can be found on the regulatory body's website[4] [5].
- If any SDTM or ADaM datasets are greater than 5 gigabytes (GB) in size, these should be split into smaller datasets.
- The define.xml version, included in the CRT package, should be a version that is supported by the regulatory body, unless an exemption/waiver is granted.
- The SDTM and ADaM CRT packages should be validated using the correct engine within pinnacle21.
    - Pinnacle21 Community or Enterprise can be used.
    - Unresolvable findings should be clearly discussed in either the cSDRG or ADRG.
- A legacy conversion will be required for any study where the data collected isn't provided in SDTM or ADaM format.
- An up-version will be required when an SDTM or ADaM IG version or a define.xml version used is no longer supported and an exemption/waiver has not been approved.
    - If an exemption/waiver is approved, then an up-version will not be required.

LEGACY CONVERSION

There are two types of legacy conversion, 1) an SDTM legacy conversion, and 2) an ADaM legacy conversion.

SDTM legacy conversion

SDTM was selected as the standard specification for submitting tabulation data for clinical studies on 21JUL2004 and for non-clinical studies 05JUL2011. This led to a lot of studies adopting SDTM. However, SDTM wasn't a requirement until 2016. This has led to some long running compounds containing studies that have data not presented in SDTM, the data on these studies is referred to as legacy data. An SDTM legacy conversion is the process of converting the legacy data into SDTM.

As the years progress the number of legacy conversions required is dropping. However, this also aids in it becoming more difficult to track down the required documentation and data required to perform the legacy conversion.

The process for an SDTM conversion involves the steps below.

Studies

The sponsor will review the studies that have been performed on a compound and are being submitted. Once selected, they will approach the vendor and request for a legacy conversion to be performed on the studies not in SDTM format.

For the legacy conversion to be done correctly the vendor will need to get confirmation from the sponsor on the following items:

- SDTM IG version – Often the latest version is selected. However, it may not be used if the regulatory body the compound is being submitted to doesn't support it yet or if the other studies are in an earlier version of IG that is still supported, and the sponsor would like consistency.
- Coding dictionary versions – The sponsor may request that the versions used in the study are updated to use a later version. If this happens then this needs to be documented within the cSDRG.
- Regulatory bodies – Knowing the regulatory bodies the compound is being submitted to, allows for the SDTM CRT package to be set-up appropriately. If multiple regulatory bodies are being used this may require the SDTM IG version to be adjusted to ensure it can be submitted to them all. It will also impact the pinnacle21 reports that are produced.

Request All Source Data

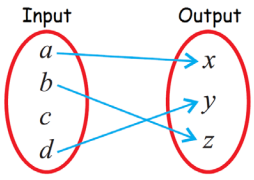
Performing a legacy conversion requires a good proportion of the study data and documentation to be available. Given that SDTM has been a requirement for 8 years, studies requiring a legacy conversion are older than 8 years and a lot of issues can crop up when requesting/retrieving the required data and documentation. These issues arise due to:

- Study data and documentation has been archived – having to restore everything takes time and can involve third parties having to provide this information.
- Sponsor has bought the compound from a different company and needs to request that company to provide everything from their archive and potentially their third-party vendors.
- Programming was performed by a different vendor and requesting their data and documentation can take time.
- Documents are only available on paper so need to be scanned.

Once all the issues have been navigated, a successful legacy conversion can take place once you have and know the following:

- Case Report Form (CRF)
- Clinical Study Report (CSR)
- Protocol
- Raw data
- SAP and TFL Shells
- SDTM IG version
- Define.xml version
- Coding dictionary versions
- Controlled Terminology (CT) version
- Regulatory body's supported standards.
- Pinnacle21 engine version

Mapping Review



The mapping review is the critical step of the process. The reason for this is that it's the step where you need to review all the study data and documentation and determine how to map it to the SDTM datasets. The mapping review should involve the creation of a mapping document. This mapping document is created by reviewing all the source data and documentation and determining, which SDTMs will be required and how the variables will be populated. This mapping document can then be used to:

- Align with the sponsor expectations
- Create the SDTM specification
- Create the aCRF
- Determine how to map new versions of coding dictionaries (if required).

Once the mapping review is complete, a plan of how to perform the legacy conversion should be in place.

Legacy Conversion

Having come up with a plan from the mapping review, the mapping document can then be used to begin the production of the SDTM specification and aCRF. The process for the legacy conversion is then like any other SDTM conversion. The source data is mapped to SDTM following the guidance from the IG, SDTM specification and utilising the aCRF to ensure all variables required for the SDTM are retained.

It is good practice to validate the SDTMs in pinnacle21 prior to creating the define.xml, ensuring resolvable issues are resolved. The define.xml can then be created using either pinnacle21 or any other tools you may use. The define.xml should use the correct version of the style sheet. Following completion of the define.xml the full package should then be validated in pinnacle21.

The production of the cSDRG should then take place. It is advisable to utilise the Legacy Data Conversion Plan and Report Appendix within the cSDRG template.

The sponsor should then review the full package.

SDTM CRT Package

Once the SDTM CRT Package has been fully reviewed to ensure that it is compliant with the regulatory body and the sponsor requirements, it can then be considered submission ready and delivered to the sponsor, who can then combine it with the rest of the studies on the compound and hand over to the regulatory body for submission.

ADaM Legacy Conversion

ADaM datasets were introduced in 2005. ADaM IG v1.0 was released as draft in 2008 and final in 2009. However, ADaM datasets did not become a requirement until 2016. As with SDTMs, not all sponsors adopted ADaM immediately and not all studies had ADaM datasets. For studies without ADaM datasets they may have used intermediate datasets to produce the TFLs, in addition to SDTM datasets. An ADaM legacy conversion, is the process of converting the intermediate datasets into ADaM datasets.

The process for an ADaM legacy conversion involves the steps below.

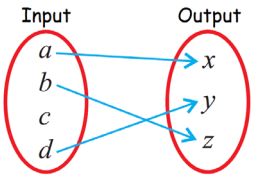
SDTM CRT Package

The ADaM legacy conversion should use the SDTM CRT package as the source. The SDTM CRT package doesn't need to be final to begin production of the ADaM CRT package. However, the final ADaM CRT package should use the final SDTM CRT package.

Request Source Data

All study documentation and data used for the SDTM CRT package should be available for the ADaM CRT package. In addition, the TFL programs, TFL outputs and intermediate datasets used are required. The ADaM CRT package may also contain the same issues that were present for the SDTM CRT package when retrieving the source data and documentation.

Mapping Review



The mapping review is the critical step of the process. The mapping review should involve a thorough review of the intermediate datasets, TFL programs and TFL outputs. As well as the other source documents and data used in the mapping review of the SDTM CRT package. If the SDTMs didn't require a legacy conversion, then all documents specified in the SDTM legacy conversion section of this paper should also be reviewed. A mapping document can be created, however the ADaM specification is able to be used to provide the methods used for all derivations. The result of the mapping review should lead to:

- List of ADaM datasets to be created. This should include any additional ADaM datasets required per regulatory body standards.
- Methods required to match the derivations performed initially on the study.
- Any participant specific exclusions.

Once the mapping review is complete, a plan of how to perform the legacy conversion should be in place.

Legacy Conversion

Having come up with a plan from the mapping review. The process for the legacy conversion is then like any other ADaM conversion. The ADaM specification is created. The ADaM specification and SDTM datasets are then used to program the ADaM datasets.

It is good practice to validate the ADaM datasets in pinnacle21 prior to creating the define.xml, ensuring resolvable issues are resolved. The define.xml can then be created using either pinnacle21 or any other tools you may use. The define.xml should use the correct version of the style sheet. Following completion of the define.xml the full package should then be validated in pinnacle21.

The production of the ADRG should then take place. It is advisable to utilise the Legacy Data Conversion Plan and Report Appendix within the ADRG template.

The sponsor should then review the full package.

ADaM CRT Package

Once the ADaM CRT Package has been fully reviewed to ensure that it is compliant with the regulatory body and the sponsor requirements, it can then be considered submission ready and delivered to the sponsor, who can then combine it with the rest of the studies on the compound and hand over to the regulatory body for submission.

UP-VERSION

The development of a compound can take many years, with multiple studies completed each year. Sponsors often tend to prefer the latest version of IG and define.xml to be used. Therefore, multiple IG versions and define.xml versions can be present on a compound. This is perfectly acceptable, provided the regulatory body still supports the versions that have been used.

When a new version of IG or define.xml are created, regulatory bodies will set a date when the version becomes a requirement and set a date for when a prior version is no longer supported. There is usually a sensible overlap. An example of this would be in the FDA Data Standards Catalog[4], where SDTM IG v3.3 was supported from 07JUL2020, became a requirement 15MAR2023 and SDTM IG v3.2 was no longer supported from 13DEC2023.

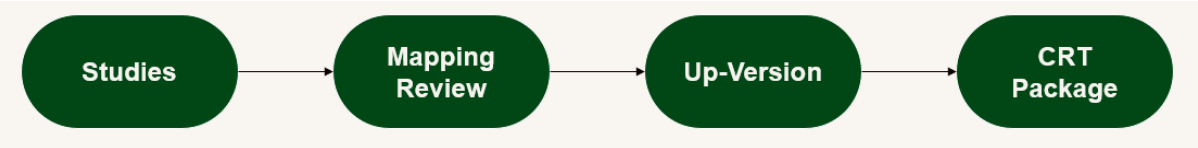
The problem most compounds have is that they have not gone to plan, various delays have occurred, or further studies have been requested. This can lead to studies no longer being in a version of IG or define.xml that is supported. A sponsor can submit an exemption/waiver form. If this is not accepted, then an up-version will be required. An up-version is the process of converting a version of IG or define.xml no longer supported into a version that is supported.

There are two forms of up-version.

1. SDTM or ADaM IG up-version.
2. Define.xml up-version.

Up-Version of SDTM and/or ADaM IG

The process of the SDTM/ADaM IG up-version should follow the steps below.



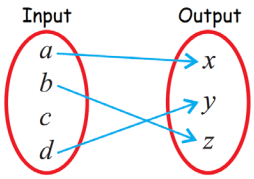
Studies

The sponsor will review the studies that have been performed on a compound and are being submitted. Once selected, they will submit an exemption/waiver form to the regulatory body. If not approved, they will approach the vendor and request for an up-version to be performed on the studies in an SDTM and/or ADaM IG version no longer supported.

For the up-version process to be performed correctly, the vendor will need to get confirmation from the sponsor on the following items:

- Studies that require an SDTM and/or ADaM IG up-version.
- SDTM and ADaM IG version – Often the latest version is selected. However, it may not be used if the regulatory body the compound is being submitted to doesn't support it or if other studies are in an earlier version of IG that are still supported, and the sponsor would like consistency.
- Coding dictionary versions – The sponsor may request that the versions used in the study are updated to use a later version. If this happens then this needs to be documented within the cSDRG.
- Controlled Terminology (CT) version.
- Regulatory bodies – Knowing the regulatory bodies the compound is being submitted to, allows for the up-version to be set-up appropriately. If multiple regulatory bodies are being used this may require the IG version to be adjusted to ensure it can be submitted to them all. It will also impact the pinnacle21 reports that are produced.

Mapping Review



The mapping review is the critical step of the process. It is less work than the legacy conversion equivalent. The main part of the review is the revision histories provided within the IG. The review of these revision histories ensure that an impact of the changes required can be assessed and the data previously provided is mapped accurately and not lost within the up-version of the study. In addition, upon review it could be determined that the current SDTM or ADaM datasets are already in an acceptable format for the new IG version.

On top of the IG review all existing documentation that was included in the previous CRT package needs to be reviewed to ensure any impacts from using a new IG version are planned and discussed. The up-version can lead to changes in the aCRF, specifications and reviewer's guides.

The expectation of this up-version is that the define.xml produced will be to the original version if still supported. However, if no longer supported then the mapping review will also assess the impacts on a newer version of the define.xml.

If only the SDTM datasets are being up-versioned this does not mean the ADaM datasets will not be affected. If the ADaM datasets are affected, it is best to check with the sponsor if they would require the ADaM datasets to be updated.

Up-Version

Following the mapping review, the plan created should then be implemented. The up-version can lead to updates to the following deliverables:

- aCRF – aCRF does not always need to be updated. If the original SDTM datasets have no changes to the destination of data specified in the original aCRF, then the original version can be used.
- SDTM Specification – Changes would often be minimal and only to a few SDTM datasets.
- ADaM Specification – Currently differences between IG versions are minimal and changes required are seen in ADAE where additional variables are required and then methodology of a couple of variables have changed.
- SDTM and ADaM datasets – The changes in IG would directly impact these deliverables. In addition, if updated coding dictionaries are requested to be used this could affect some of the datasets. If the CT version is newer this could also lead to code lists being updated.
- Define.xml – given the SDTM and ADaM datasets are expected to change the define.xml would need to be created using the latest version of datasets. The define.xml version may also need to be up-versioned.
- cSDRG and ADRG – the IG versions, define.xml, coding dictionary versions and CT versions. In addition, the pinnacle21 validation may produce new findings and resolve previous findings.
- Pinnacle21 validation should be performed on the latest version of the CRT packages.

Once all work is complete the sponsor should review.

Final CRT Package

Once the CRT Packages have been fully reviewed to ensure they are compliant with the regulatory body and the sponsor requirements, it can then be considered submission ready and delivered to the sponsor. Who can then combine it with the rest of the studies on the compound and hand over to the regulatory body for submission.

Up-Version of Define.xml Version

The process of the Define.xml up-version should follow the steps below.

Studies

The sponsor will review the studies that have been performed on a compound and are being submitted. Once selected, they will submit an exemption/waiver form to the regulatory body. If not approved, they will approach the vendor and request for an up-version to be performed on the studies with a define.xml version no longer supported.

For the up-version process to be performed correctly, the vendor will need to get confirmation from the sponsor on the following items:

- Define.xml version – Often the latest version is selected. However, it may not be used if the regulatory body the compound is being submitted to doesn't support it or if other studies are in an earlier version of define.xml that is still supported, and the sponsor would like consistency.
- Regulatory bodies – Knowing the regulatory bodies the compound is being submitted to, allows for the up-version to be set-up appropriately. If multiple regulatory bodies are being used this may require the define.xml version to be adjusted to ensure it can be submitted to them all. It will also impact the pinnacle21 reports that are produced.

Up-Version

There isn't a mapping review required for the define.xml up-version. The SDTM/ADaM datasets and specifications should be in a usable format and either pinnacle21 or the tool used by the vendor to create the define.xml to the correct version can be used.

The up-version of the define.xml should only lead to changes to the define.xml, stylesheets and reviewer's guides. The process for creating the define should be the same as the define process used for any other study. Once the define.xml is created, it should be validated in pinnacle 21. Issues should be resolved or discussed in the reviewer's guides.

Once all work is complete the sponsor should review.

Final CRT Package

Once the define.xml has been up-versioned and fully reviewed to ensure they are compliant with the regulatory body and the sponsor requirements, it can be added to the CRT package and considered submission ready and delivered to the sponsor, who can then combine it with the rest of the studies on the compound and hand over to the regulatory body for submission.

CONCLUSION

SDTM and ADaM CRT packages are the required format for data being submitted. The SDTM and ADaM datasets should be in a supported IG version. The define.xml should be in a supported define version.

If the data is not in SDTM or ADaM format, then the study will require a legacy conversion. If the SDTM and ADaM IG version is no longer supported, the datasets should be up-versioned to a supported IG version. If the define.xml is no longer in a supported version, this should be up-versioned to a supported version.

For the legacy conversion and up-version, the mapping review is critical. The mapping review allows for a clear plan to be defined and steps to be put in place to follow. This can give a clear oversight of the process to the sponsor and regulatory body, which may also answer any questions that arise during the submission process.

The legacy conversion is more complicated than the up-version, due to the age of the studies involved and the status of all the required data and documentation.

All outputs from a legacy conversion and up-version should utilise pinnacle21 for validation and findings should be clearly discussed in the respective reviewer's guides.

REFERENCES

[1] Clinical Study Data Reviewer's Guide (cSDRG) Package: Clinical Study Data Reviewer's Guide (cSDRG) Package - WORKING GROUPS - PHUSE Advance Hub

[2] Analysis Data Reviewer's Guide (ADRG) Package: Analysis Data Reviewer's Guide (ADRG) Package - WORKING GROUPS - PHUSE Advance Hub

[3] Analysis Data Model Implementation Guide for Non-compartmental Analysis Input Data: ADaMIG for Non-compartmental Analysis Input Data v1.0 | CDISC

[4] FDA Data Standards Catalog: Data Standards Catalog | FDA

[5] PMDA Data Standards Catalog: New Drug Review with Electronic Data | Pharmaceuticals and Medical Devices Agency (pmda.go.jp)

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:
Phillip Spowart
Fortrea Clinical Research Unit Ltd.

Drapers Yard
Marshall St, Holbeck, Leeds, LS11 9EH
Email: phillip.spowart@fortrea.com
Web: www.fortrea.com