

Transforming Legacy Data into CDISC Standards: Practical Approaches and Regulatory Insights

Veena Palle and Khalif Halani

**PHUSE Connect,
Austin, TX
March 26, 2026**

**Data Standards, Governance
& Implementation**

DS20

On the Menu for Today's Talk

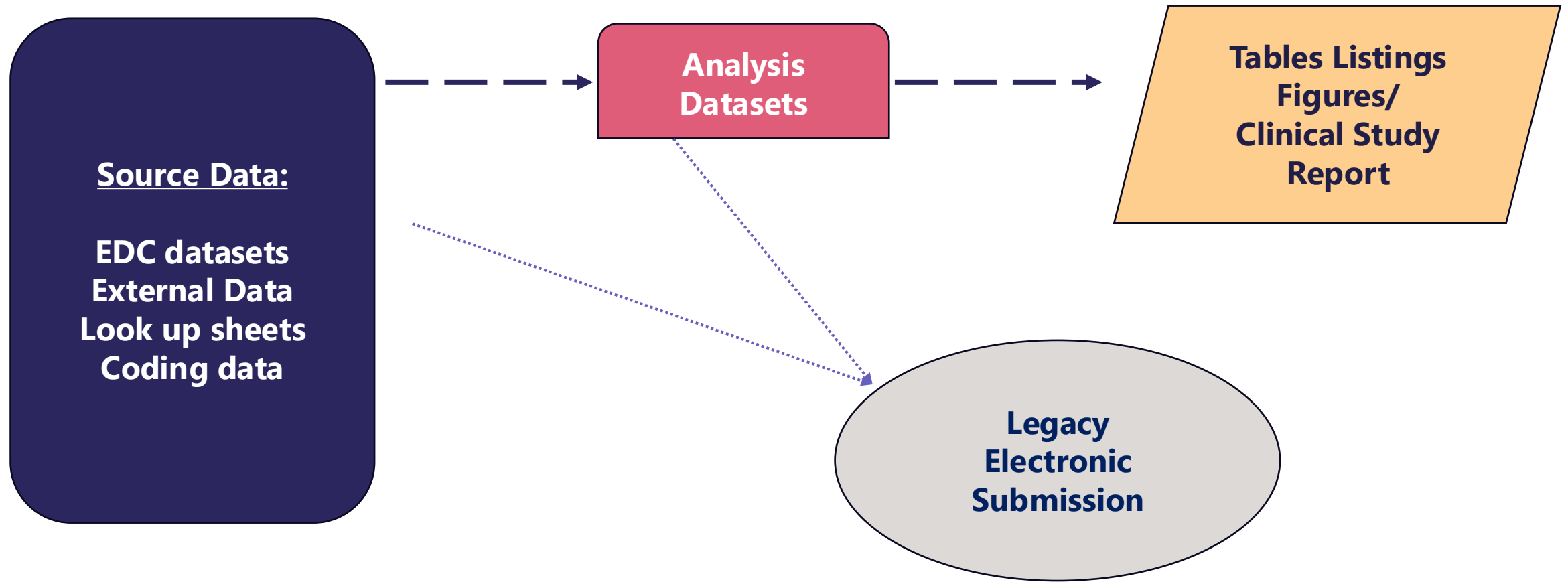
- 01** What is Legacy Data and why does it need converting?
- 02** Why would these situations happen and what do the regulators say?
- 03** How is it different?
- 04** Approaches – what must be done and what to expect
- 05** Challenges Encountered
- 06** Tips and Takeaways

01

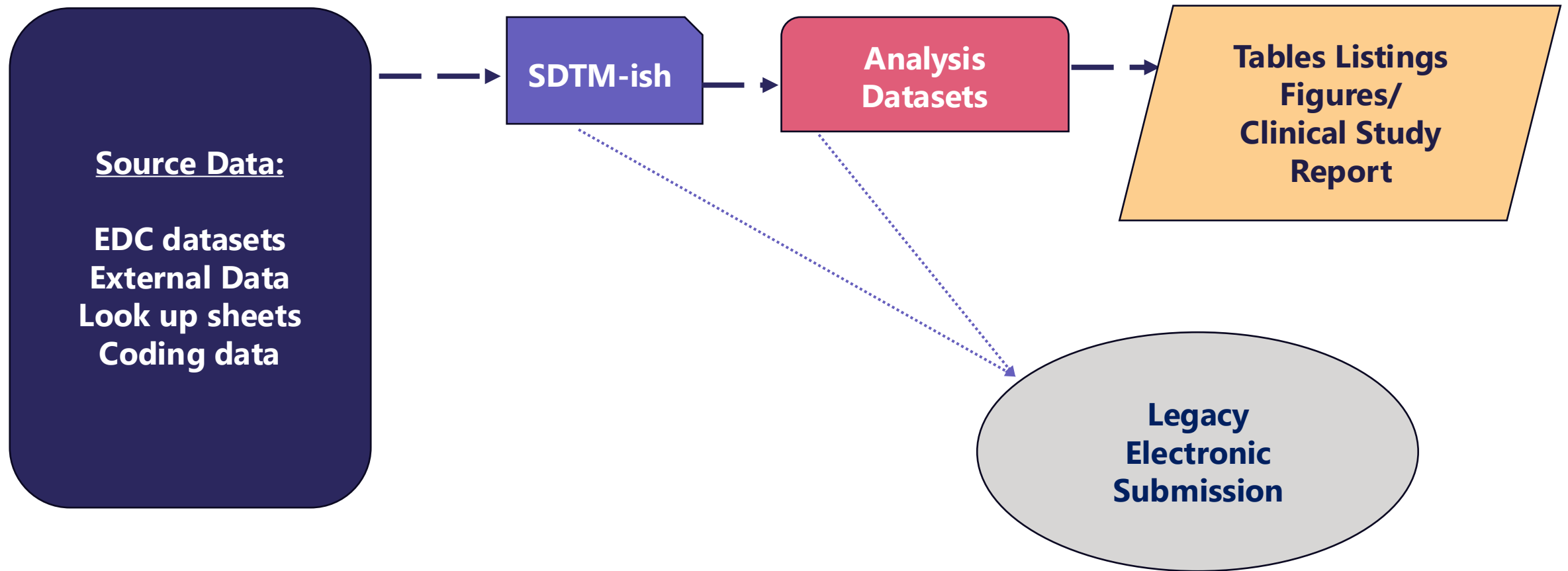
Legacy Data Conversion



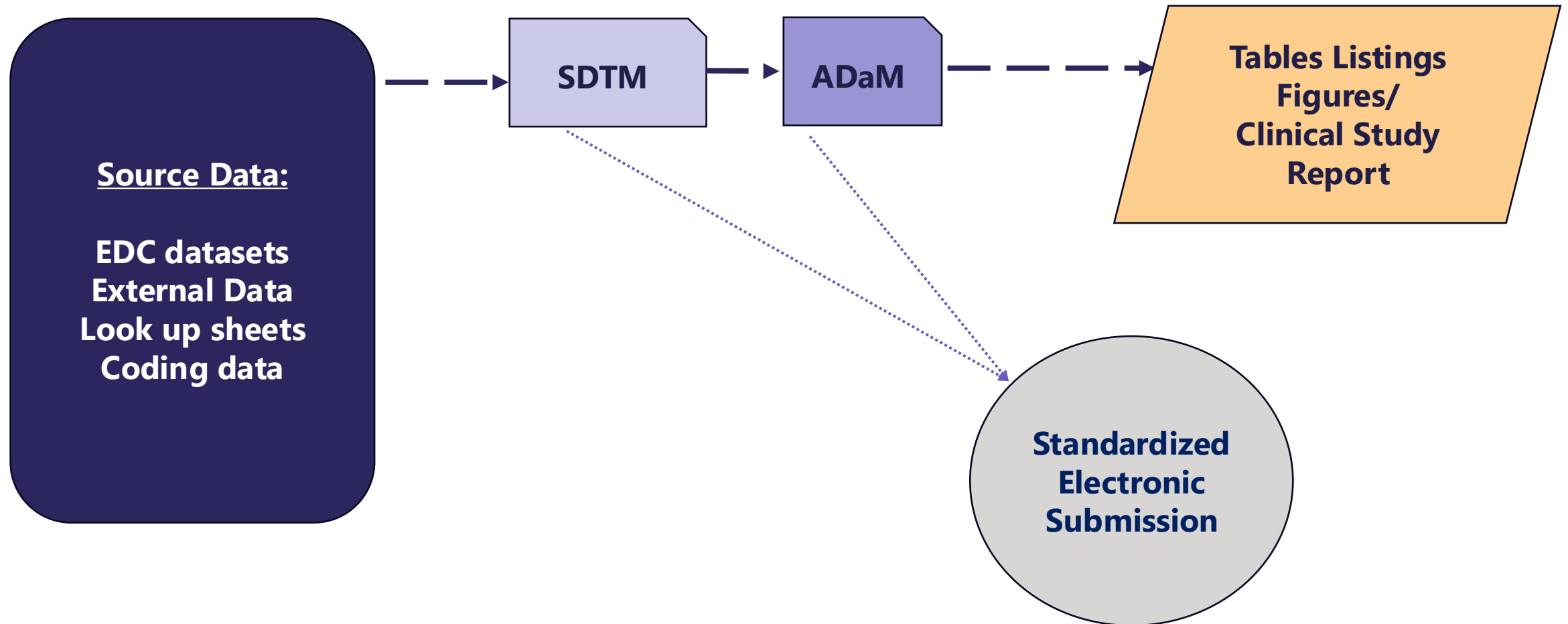
Producing TLFs and CSR prior to CDISC Standardization (Legacy)



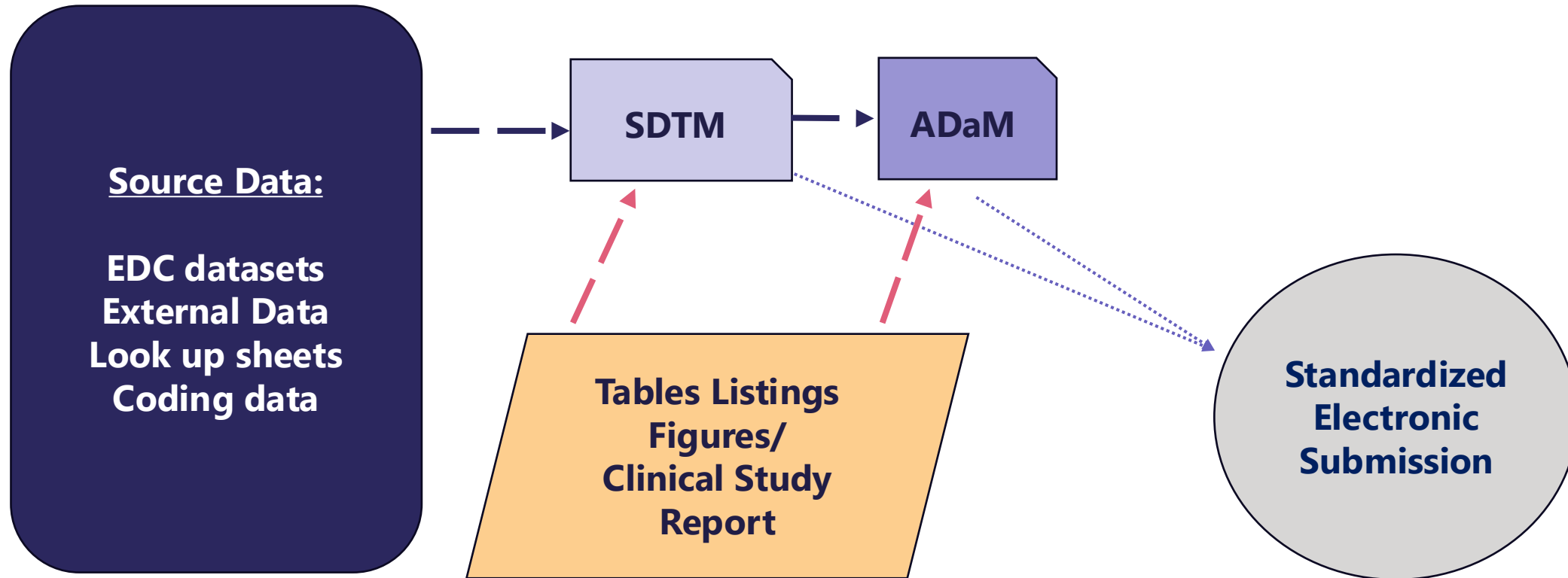
Another Variation of Legacy Data



Producing TLFs and CSR using CDISC Standards



Conversion of Legacy Data



02

**Why does this
happen and
what do
regulators say?**



Situations

- The study was performed and reported before CDISC was mandated
- Study results and data were not originally intended for submission (e.g., investigator-initiated studies)
- Study was terminated by the Sponsor before the study team had developed the SDTM and ADaM datasets
- Originally submission accepted via FDA waiver of CDISC requirement
- Financial or practical reasons led to decision to report without CDISC compliant data.

Regulatory Perspectives

FDA accepts post-CSR created datasets for submission and understand conformity may not be perfect.

- Datasets must be technically CDISC-compliant
- Fully traceable to study results, reproducibility
- Consistency between CDISC datasets and CSR outputs
- Transparent documentation of derivations
- Clear justification for deviations from CDISC guidance

**See Section 8.1.4 – FDA Study Data
Technical Conformance Guide, Technical
Specifications Document, December 2025.**

Updated Guide

New guideline calls for the following

- Create and include Legacy Data Conversion Plan
- Create and include Legacy Data Conversion Report
- Include CRF annotated with legacy data

Include the first two in the SDRG

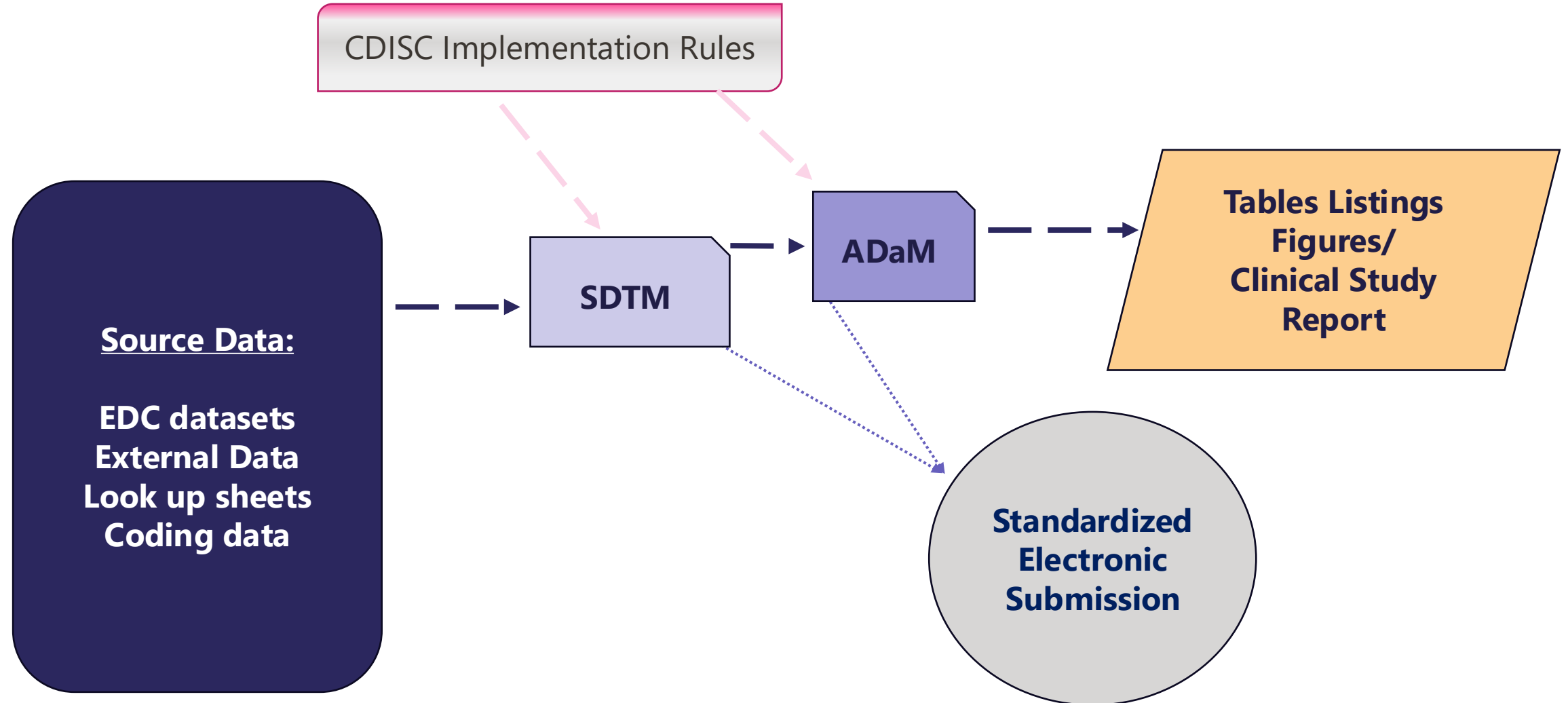
**See Section 8.1.4 – FDA Study Data
Technical Conformance Guide, Technical
Specifications Document, December 2025.**

03

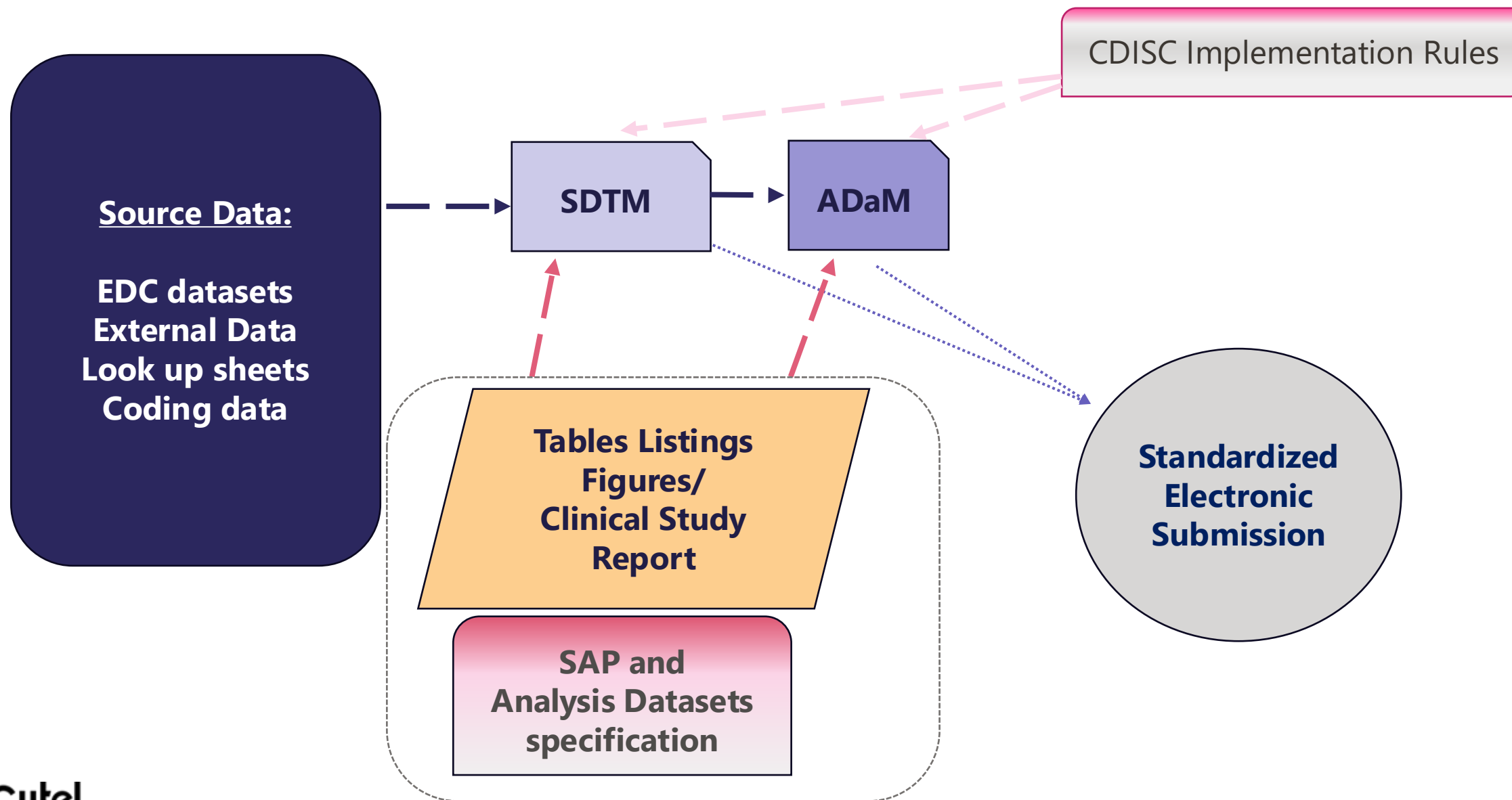
Distinct Challenges



Producing TLFs and CSR using CDISC Standards



Conversion of Legacy Data to CDISC



04

Approaches



Gather Material

Must Haves

- 1) CSR/Study Report
- 2) TLFs
- 3) Source Data
- 4) Case Report Form

Nice to Have

- 1) Statistical Analysis Plan
- 2) Manuals (or DTAs) for non-EDC data
- 3) TLF Mock Shells
- 4) Programming instructions

Christmas Comes Early

- 1) Legacy Analysis Datasets (ADs)
- 2) Specifications for ADs
- 3) SAS programs for ADs
- 4) SAS programs for TLFs

Get Familiar with the CSR (study results and write-up)

Endpoints

Imputations

Analysis Population Criteria

**Derivations & Other
Important data handling
rules**

**Censoring Rules (if
applicable)**

**Statistical methods used to
generate the results**

Create, Compare, and Document

Deliverables

- SDTM datasets, define.xml,
 - Annotated CRF (for SDTM)
 - Annotated CRF (for Legacy)
 - Reviewer's Guide
 - Legacy Data Conversion Plan
 - Legacy Data Conversion Report
- ADaM datasets, define, and Reviewer's Guides

Comparison to CSR

Confirm that the selected results from the CSR can be reproduced from the CDISC data.

Recommended:

Verify endpoints supporting the study objectives and select safety summaries

Document

Provide the reviewers clear documents of significant data issues, clarifications, explanations of traceability, and adjudications.

05

Challenges & Recommendations





“

“Plans are useless, but planning is indispensable.”

Dwight D. Eisenhower

“To improve is to change; to be perfect is to have changed often”
Winston Churchill

”

“The best-laid plans of mice and men often go awry”

Robert Burns, *To a Mouse* 1785

Challenges and Recommendations

Challenges

Legacy analysis data involved hard coding

Recommendations

- **Leverage SAS programs**
- **Leverage (available) specs**
- **Compare to CSR listings**

Chronological Data Entry Errors

- **Ensure the data match the CSR**
- **Highlight issue in the Reviewer's Guides and Legacy Data Conversion Report**

Challenges and Recommendations

Challenges

Values do not conform to the non-extensible CDISC code list (CT)

Recommendations

- **Include both original and CT values**
- **Highlight issue in the Reviewer's Guides and Legacy Data Conversion Report**

Look-up lists used to classify are not available

- **Look through all available sources for insight**
- **Reverse-engineer list from CSR**
- **Include in define xml and RGs**

06

Pro Tips

“All surprises will come from within the data”

Angelo Tinazzi



Some Considerations

Standardize the SDTMs as much as possible to other submission-compliant datasets in the compound.

Use the submission documentation to clearly document significant data issues and how they were handled or represented in the data

Ensure the Legacy Plan and Report are included

Ensure the ADaM datasets are “one proc away” from the results for re-productibility

Plan your comparison plan to TLFs ahead of time; revisit as the process is ongoing

Give yourselves enough time; be prepared to dig into the source data and documentation

Cytel

Thank You



Contact

- Veena.Palle@cytel.com
- Khalif.Halani@cytel.com