



Automation for ADaMs: Building a Solid Metadata Foundation for Future AI Integration

Manish Deole

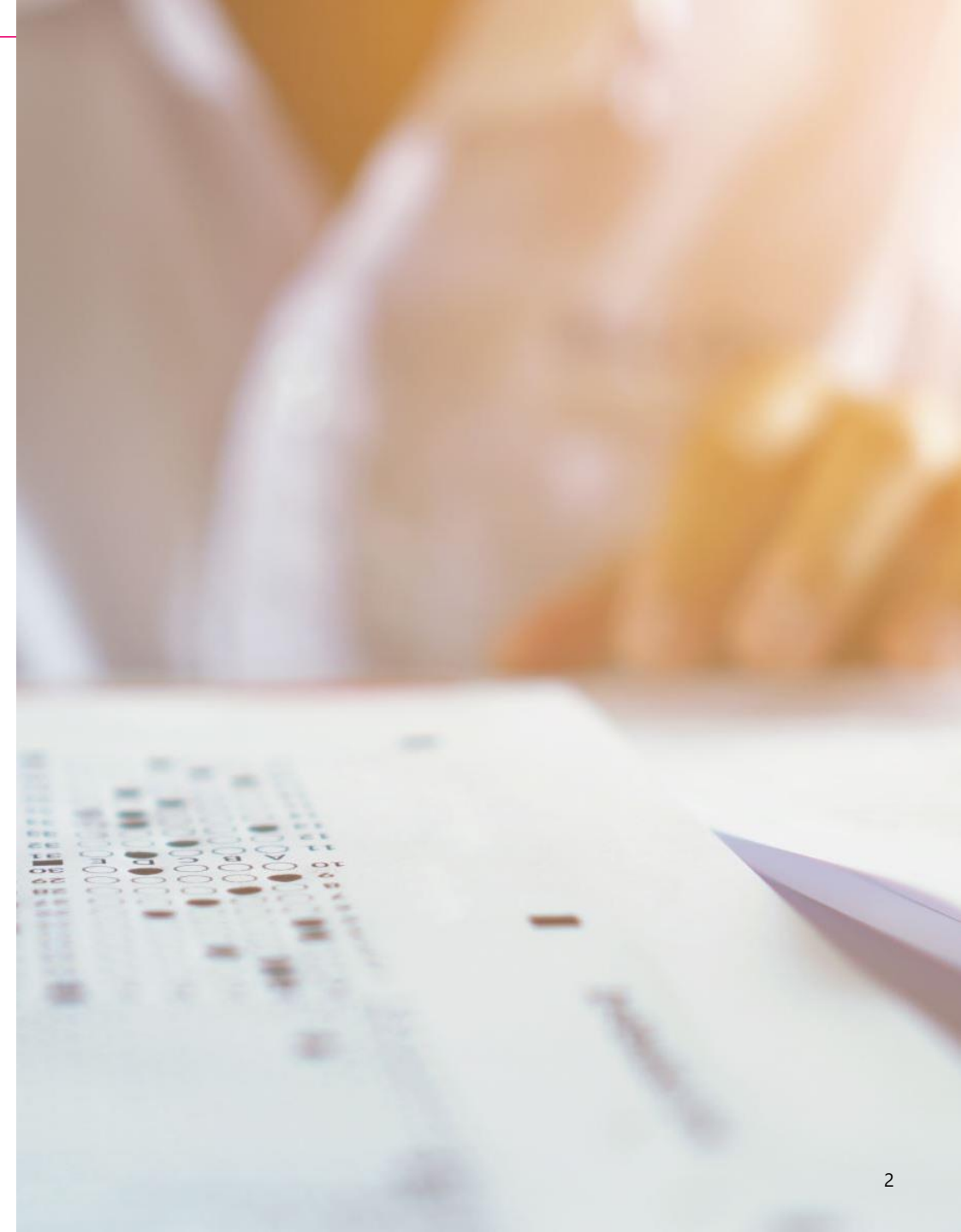
11-Jun-2025

PHUSE SDE (Boston)

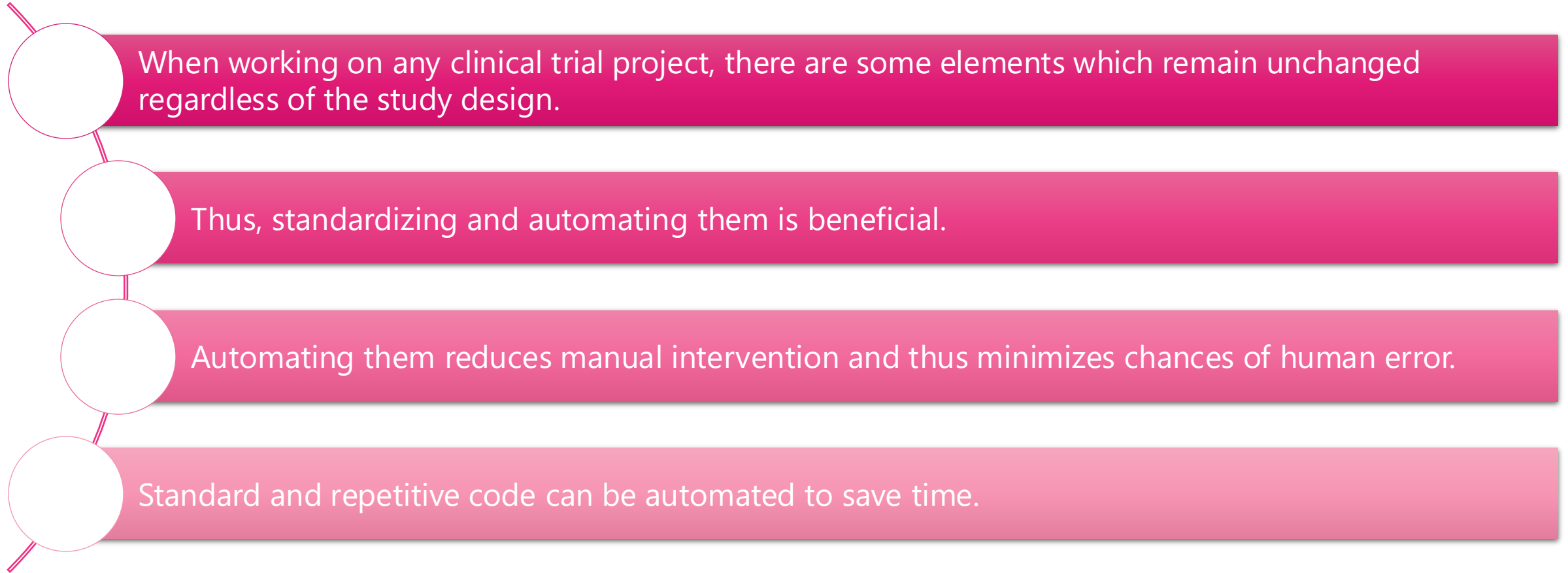


Introduction

- ADaM (Analysis Data Model) datasets are standardized, analysis-ready datasets specifically designed for statistical analysis, making it easier for analysts to perform calculations and generate tables, listings, and figures (TLFs).
- ADaM provides a standardized approach to dataset creation, ensuring uniformity across different studies and facilitating interoperability between software and analysts.
- By providing analysis-ready datasets, ADaM streamlines the statistical programming process, saving time and resources.



The need for Automation



Automation of ADaM Dataset Generation



ADaM automation allows for the automatic generation of analysis datasets from SDTM data, reducing manual effort and time.



Streamlined workflows eliminate the need for manual data entry, freeing up staff to focus on more complex tasks.



By automating data derivation and metadata population, ADaM automation minimizes the risk of human error.



Automated ADaM generation allows for faster preparation of analysis datasets for regulatory submissions.

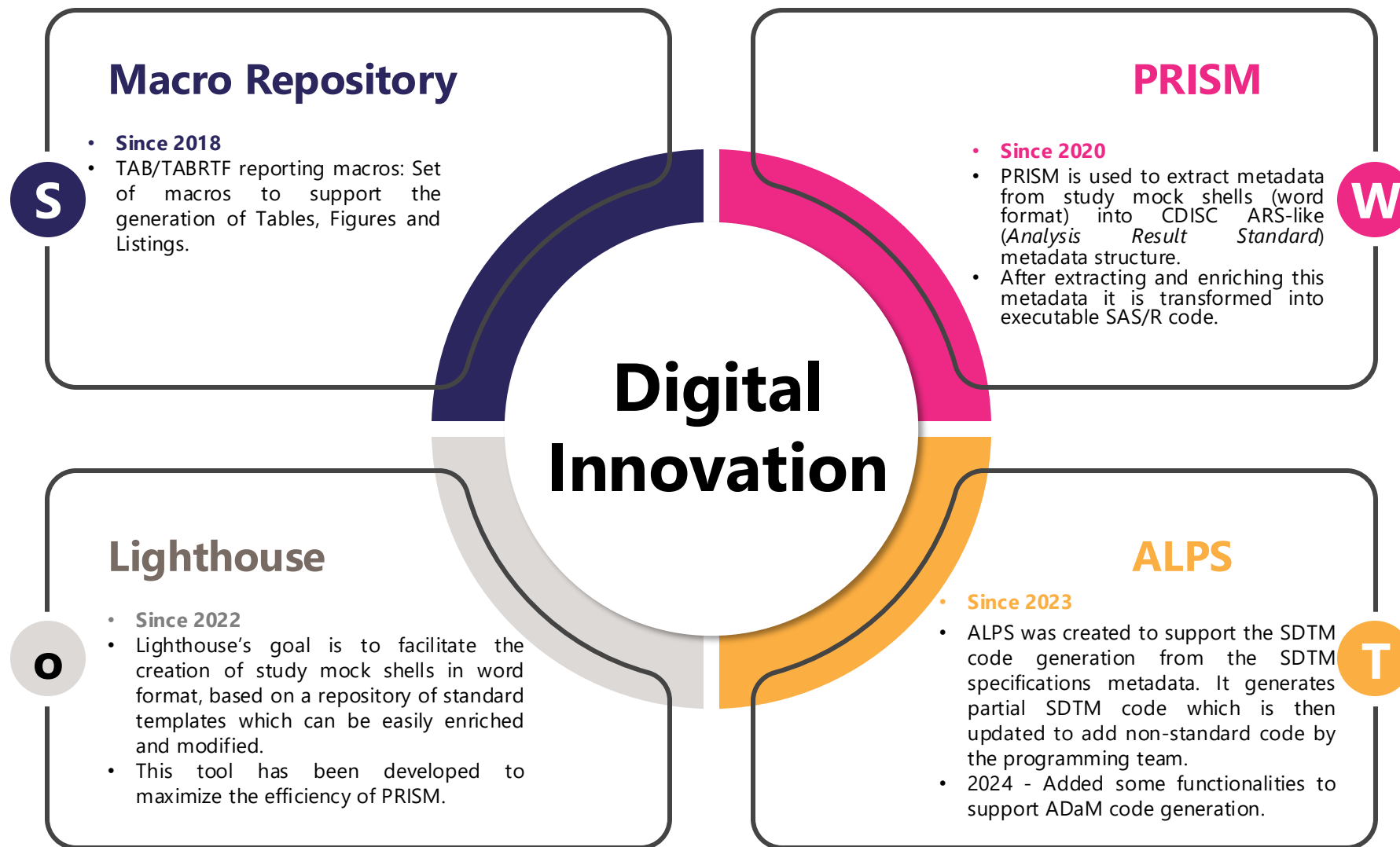


Clear lineage and consistent metadata make it easier for reviewers to assess and approve submissions.

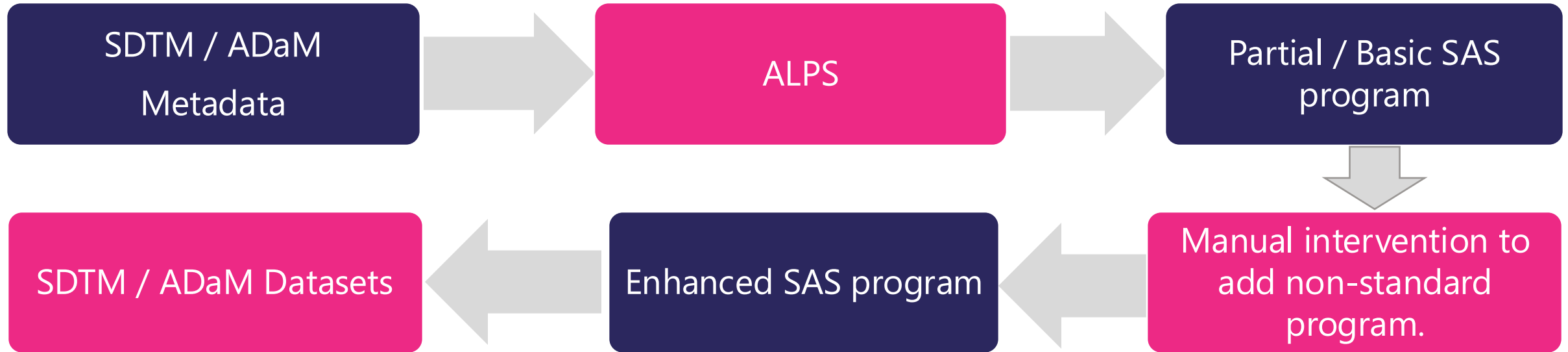


Well-organized and traceable ADaM datasets can lead to faster review times and quicker regulatory approvals.

Our Journey towards standardization at Cytel



ALPS* (SDTM & ADaM)



- Proper metadata preparation enhances both program development and specification processes.
- Metadata serves as a foundation for efficiently preparing define.xml.
- Metadata-driven automation requires fewer resources and accelerates the SDTM and ADaM mapping process.

*ALPS: **A**ccelerated **L**egacy to CDISC SDTM/ADaM **P**rocessing **S**ystem

Metadata-Driven Approach

A metadata-driven approach to ADaM dataset automation streamlines clinical data transformation by leveraging predefined metadata standards. This method enhances efficiency, consistency, and compliance with CDISC requirements.

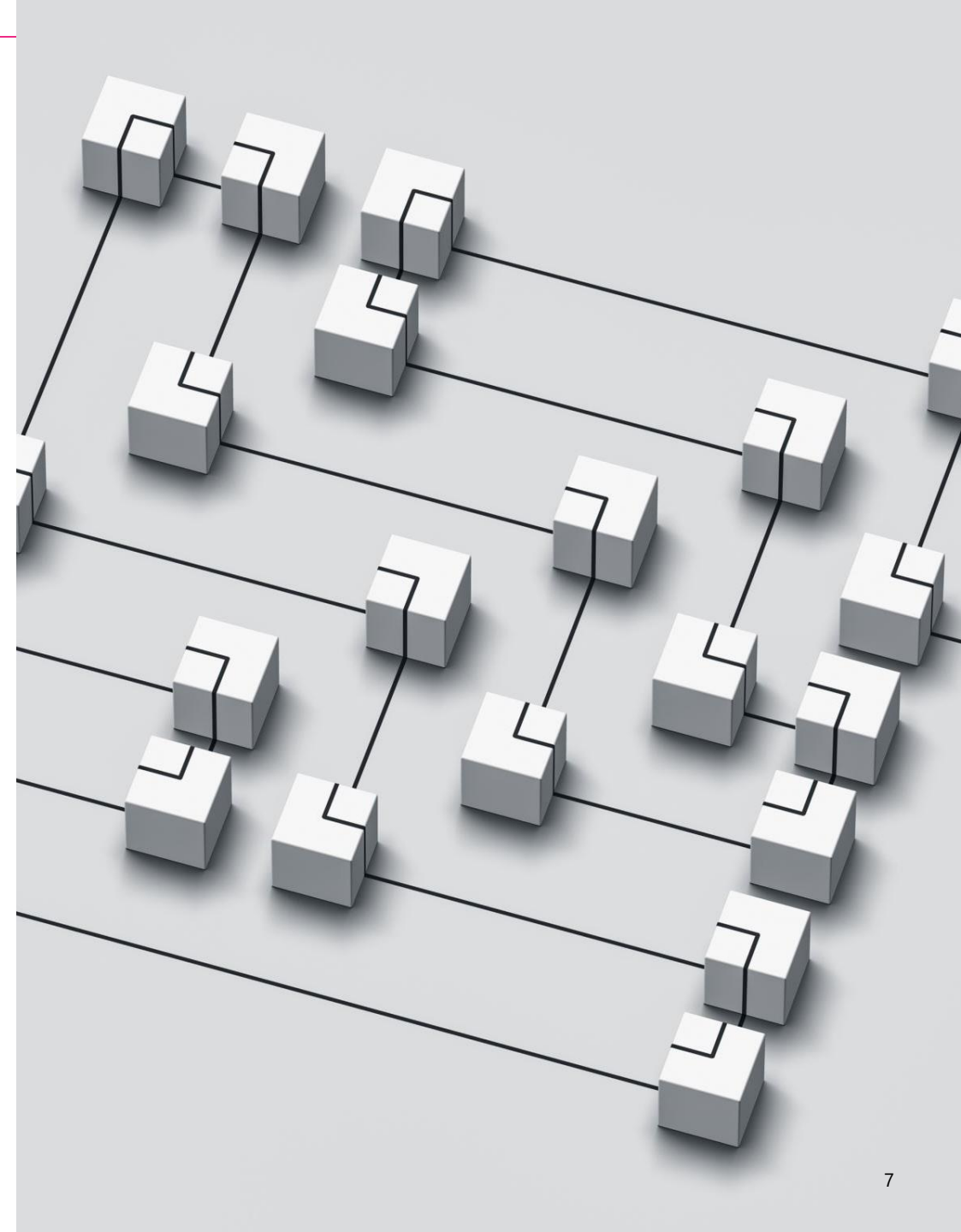
Cytel's proprietary standard macros can be integrated by the tool to facilitate consistency and efficiency.

To minimize errors, we ensure that the input datasets are correctly referenced, required variables are available and the variable names and types match the actuals in the datasets.

Lead programmer must ensure correct data selection conditions are specified in metadata.

ADaM automation is in nascent stage and we are enhancing it further.

Limitations: The current version of the tool supports simple transformations and derivations. More complex derivations like LOCF, Multiple Imputation, etc. are manually added to the partially generated code.



SDTM implementation

Advantages:

- 40% of the current SDTM projects are mapped efficiently and smoothly with the ALPS automation.
- Specifications are compliant towards generating define.xml.
- Developer code is generated from the specifications via standard macros.
- Quick production of first draft of developer code.
- High level of details of programming in specifications from the beginning of the mapping process.
- Easy maintenance of traceability between specifications and mapping.

Limitations:

- High level of study specific data collection.
- Complex mapping can't be managed directly from specifications and need post-processing programming steps.
- External vendor data reconciliation can't be managed directly from the specifications and need post-processing programming steps.
- Standard specifications / pre-filled derivation rules need standard CRF pages.



ADaM implementation



ADaM automation is in pilot phase.



So far, we have tested it in 4 projects.



Initial focus is on standard safety datasets.



Learnings from SDTM automation are helping build a solid foundation for ADaM automation.

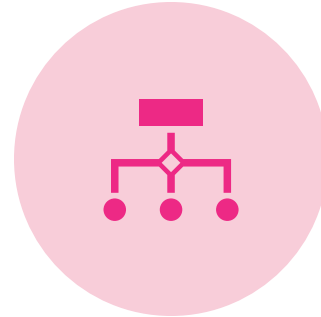
Benefits of Automation



Efficiency: Since the team can focus more on the non-standard parts of the outputs, the overall efficiency of the team is increased.



Quality: Since the tools have been rigorously tested, they produce extremely high-quality and reliable outputs.



Consistency: Since automated tools generate standard code based on a set of rules, the resulting code remains highly consistent across various projects. This makes it easier to understand and debug (in case of any updates).

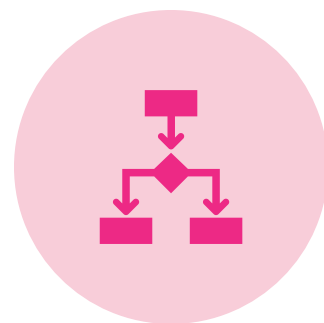


Reduced manual intervention: Since manual intervention is limited, the possibility of human error is minimized. As long as the specifications are correctly drafted, the output generated by the standard code should be error-free.

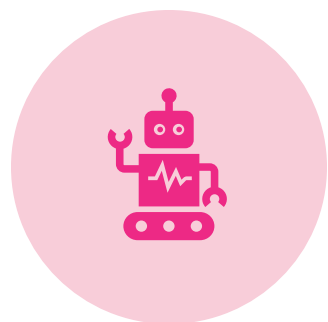
What next?



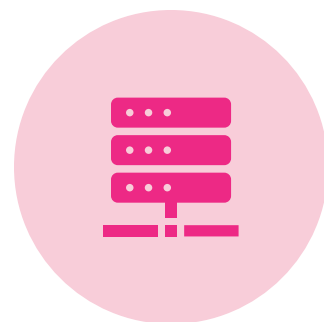
Expand ALPS to incorporate AI-based derivation suggestions.



Handle complex derivations such as LOCF, multiple imputations, adding new records (e.g., new parameters), etc.



Train AI models on completed SDTM/ADaM mappings to generate reusable derivation libraries.



Easy integration of external data.

Questions?

