



2026 Europe Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



MILAN, ITALY | MAIN CONFERENCE: 20-21 MAY | TRAININGS & WORKSHOPS: 18, 19, & 22 MAY

Automating Define-XML Generation in the CDISC 360i Program

Pierre Dostie, Tools & Standards Specialist, UCB

Speakers



Pierre Dostie
Tools & Standards
Specialist
UCB

A senior statistical programming specialist with nearly 30 years of experience, including over 20 years in the pharmaceutical industry. Currently focused on guiding teams in ADaM specifications, define.xml, and ADRG, while contributing to CDISC initiatives such as the CDISC 360i project.



Agenda

1. Scope of the 360i Data Definition Engine Project
2. Inefficiencies in Today's Process
3. What Are We Trying to Accomplish and Why it Matters
4. The Project: The Data Definition Engine (DDE)
5. Data Definition Specification (DDS)
6. Future Work
7. Key Takeaways
8. Questions?



Scope of the 360i Data Definition Engine Project

Made of multiple feature teams:

- **Define-XML generation**
- ODM-based CRF generation
- Dataset Shell generation
- Trial Design datasets generation
- Generating SDTM datasets from Lab DTAs
- Data Definition Specification (DDS) model

This presentation focuses on SDTM Define-XML generation



Inefficiencies in Today's Process

Manual and Inefficient Workflow

- Currently, most Define-XML workflows rely heavily on spreadsheets, local conventions, and manual editing, causing inefficiencies.

Error-Prone Processes

- Manual copy-paste and ad hoc edits lead to mismatches and inconsistencies in metadata and datasets, increasing QC churn.

Limited Maintainability and Reuse

- Spreadsheet templates are study-specific and not machine-interpretable, limiting standardization and systematic artifact generation.

Automation Opportunity

- Generating Define-XML from a consistent metadata backbone reduces errors, streamlines updates, and scales across projects.

What Are We Trying to Accomplish and Why This Matters

What Are We Trying to Accomplish?

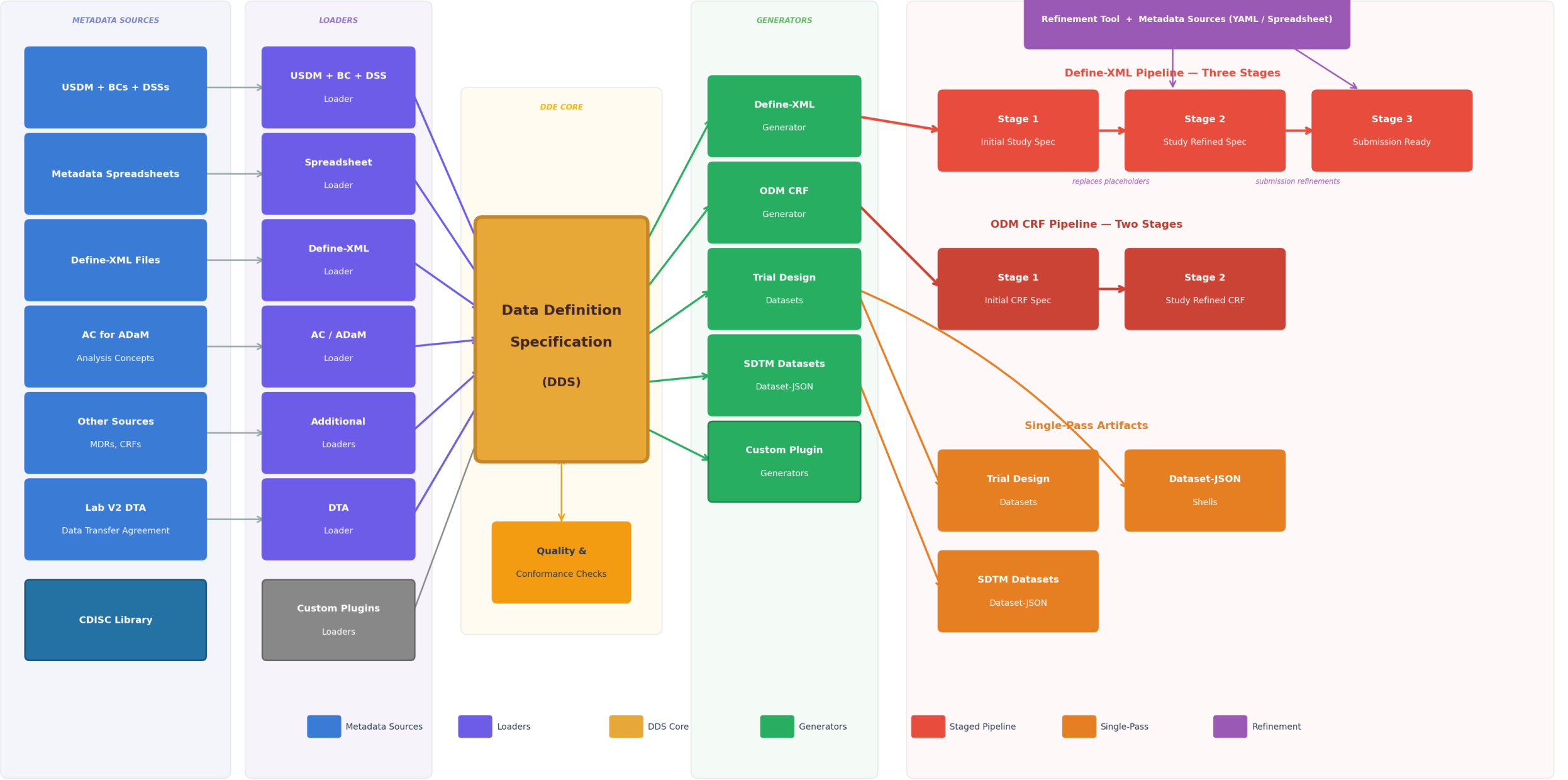
- Maximize automation / minimize manual work
- Generating Define-XML: USDM → Biomedical Concepts → SDTM Dataset Specializations
- Support using multiple sources of metadata (i.e. metadata spreadsheets or MDRs)
- Create a new Data Definition Specification model to support metadata-driven automation
- Identify and resolve metadata/standards gaps

Why This Matters

- Bridging future methods of generating Define-XML with current ways of working
- Open-source where a team of experts performs ongoing development and maintenance
- Higher quality Define-XMLs generated more efficiently
- Realizing the benefits of your investment in standards via metadata-driven automation

Data Definition Engine (DDE) Architecture

CDISC 360i — Define-XML / CRF Generation Sub-team



Data Definition Engine (DDE) Architecture

Generating Define-XML from the Study Design



Plug-In Architecture

Extensible by design — add new capabilities without modifying the core

Loaders

Plug in new data sources to feed the engine

- MDR APIs
- Metadata Spreadsheets
- Proprietary Sources

DDE

Core Engine

Generators

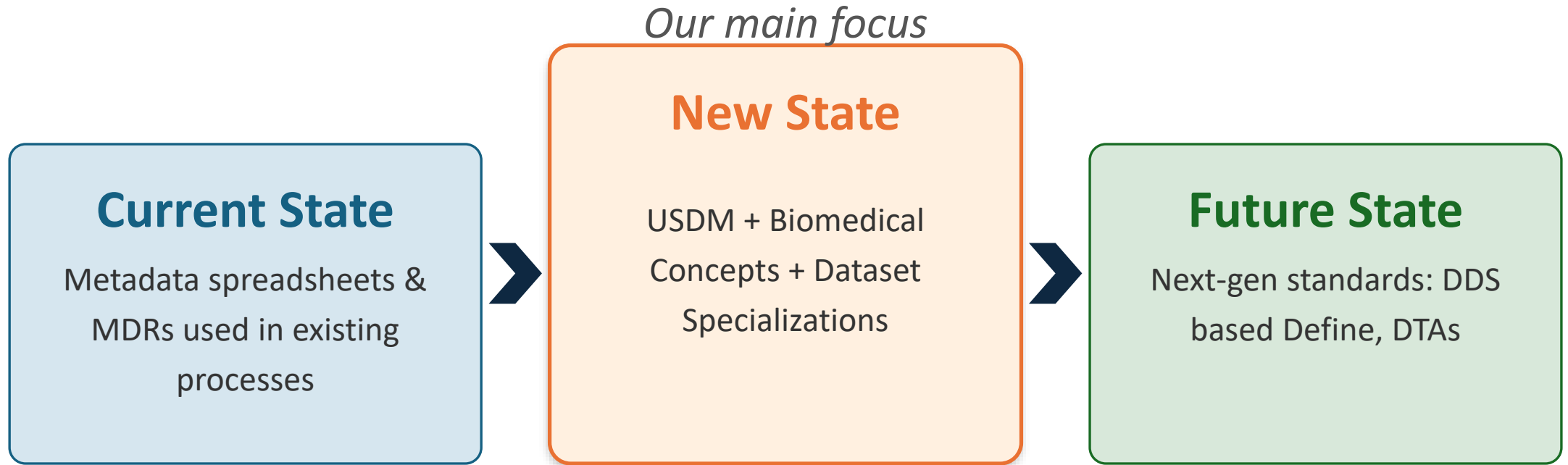
Create new study artifacts or extend existing ones

- Define-XML
- ODM-based CRFs
- Dataset Shells
- Trial Designs Datasets
- Lab DTAs to SDTM

Refinement Pipelines

Custom workflows for study-level metadata refinement and submission preparation

The Solution Targets Multiple Time Frames



Supporting current workflows while bridging to more modern ways of working

Metadata Gaps Identified

KEY FINDINGS

01 Gaps Identified

Numerous gaps in metadata needed to automate Define-XML were identified across multiple areas

02 Placeholder Approach

Placeholders used to address missing metadata in the first Define-XML generated

03 Standards Impact

Gaps may drive updates to standards, improving metadata coverage for future automation

EXAMPLES OF GAPS

Variable Attributes

- KeySequence
- Length
- Code list DataType
- Repeating

Versions of Standards Used

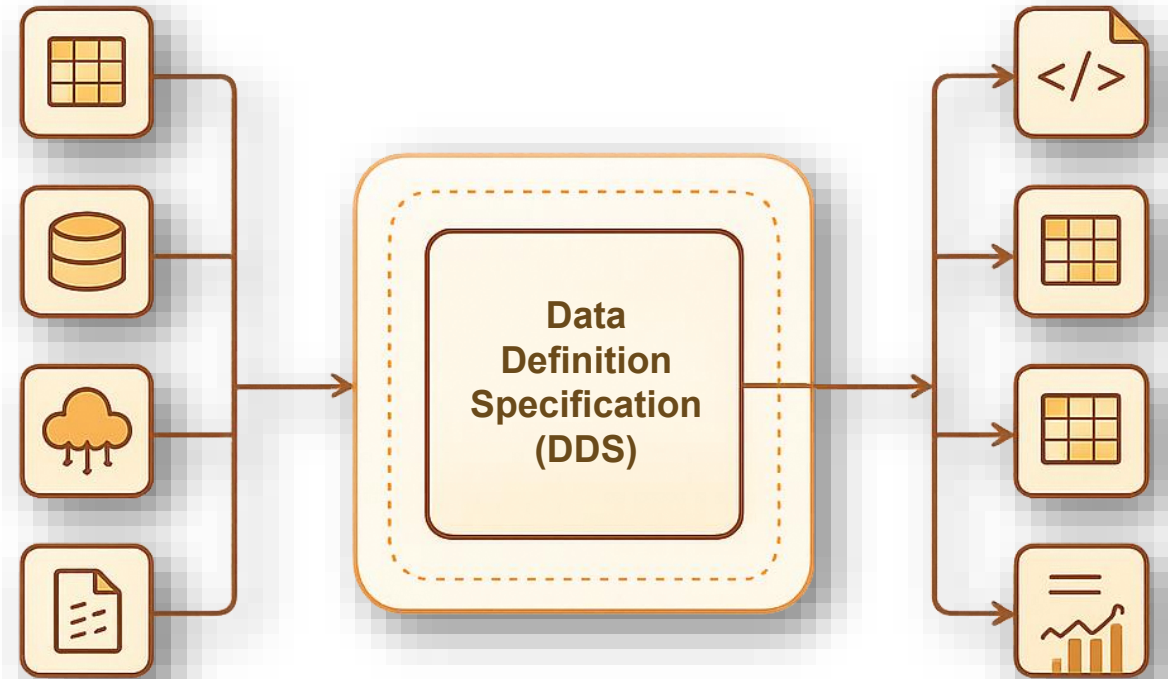
- Dictionaries and their versions
- IG version used
- CT versions used

Study Context

- What VLM is relevant to the study
- Some BCs have multiple DSSs
- Relevant code list subsets

Data Definition Specification

- Links upstream conceptual models to downstream artifacts, replacing unreliable spreadsheet intermediaries.
- Defines consistent metadata structures for data demand and supply.
- Supports standards metadata-driven automation.



Future Work



01

Refine & Enhance

Strengthen the current work-in-progress into a polished, production-ready solution

02

ADaM Define-XML

Extend automation support to ADaM Define-XML generation

03

DTA Transformations

Enable DTA to SDTM dataset transformations for end-to-end processing

04

Plug-In Architecture

Support custom loaders and generators through an extensible plug-in system

05

EOY Release Target

Deliver an initial open-source release by end of year

Key Takeaways

Proof of Automation

Define-XML can be generated consistently from structured, standards-based metadata in a machine-consumable model.

Phased Progress Achieved

Phase 1 delivered automated SDTM Define-XML generation using new metadata specification models and biomedical concepts.

Forward Extension Plans

Phase 2 will automate ADaM Define-XML, incorporating analysis concepts as structured metadata.

Process Improvement

Metadata-driven automation replaces spreadsheets, enhancing reuse, maintainability, and quality while reducing errors.

Standards Feedback Loop

Automation efforts reveal gaps in CDISC standards, guiding standards evolution and better implementation.

Open-Source Adoption

Open-source solutions promote interoperability, accelerate learning, and reduce duplicated efforts across organizations.



Thank You!



Questions?

