



# Metadata-driven Automation



# Objectives

- **Understand** how open-source and metadata can deliver automation to accelerates clinical data workflows
- **See** live demos across digital protocols, SDTM datasets, and analysis results (ARDs)
- **Recognize** reusable open-source components that enable consistency across studies
- **Appreciate** how metadata strengthens traceability and reproducibility end-to-end
- **Identify** where these approaches can deliver value for their own clinical study teams



# Agenda

|                                                                                   |                               |         |
|-----------------------------------------------------------------------------------|-------------------------------|---------|
|  | Welcome and intros            |         |
| <b>Use Case 1</b>                                                                 | Digital Protocol              | 25 mins |
| <b>Use Case 2</b>                                                                 | Dataset Generation (SDTM)     | 25 mins |
|  | Break                         | 10 min  |
| <b>Use Case 3</b>                                                                 | Metadata-Driven Results (ARS) | 25 mins |
| <b>Group Discussion</b>                                                           | Value Proposition             | 25 mins |
|  | Wrap up                       | 5 mins  |



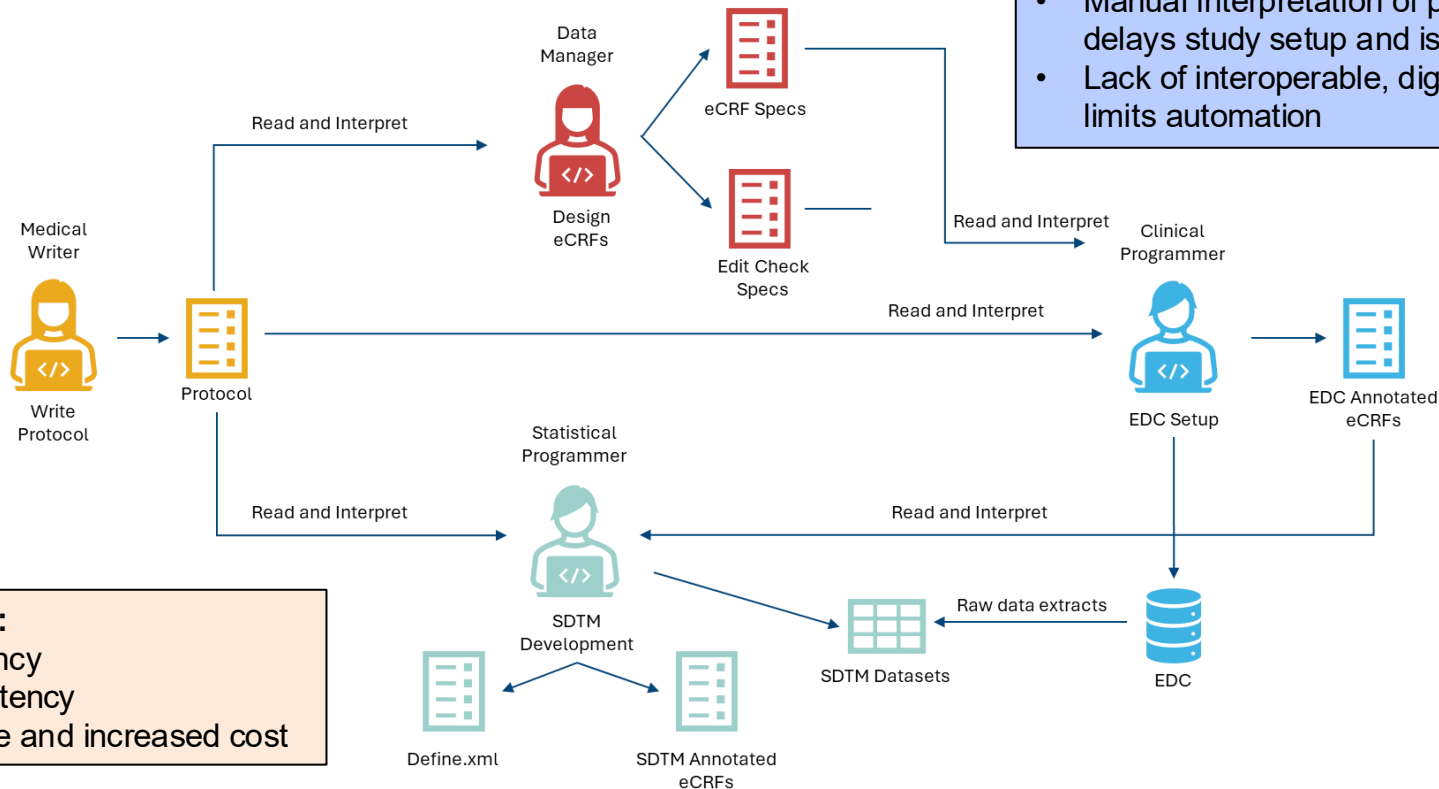
# Use Case 1 – Digital Protocol



# Current Challenges

## Today's Clinical Research Landscape:

- Study Design, Data Collection, Tabulations, and Analysis often operate in disconnected functional and data silos
- Manual interpretation of protocols delays study setup and is error prone
- Lack of interoperable, digital standards limits automation



## The Result:

- Inefficiency
- Inconsistency
- Lost time and increased cost



# CDISC 360i Vision

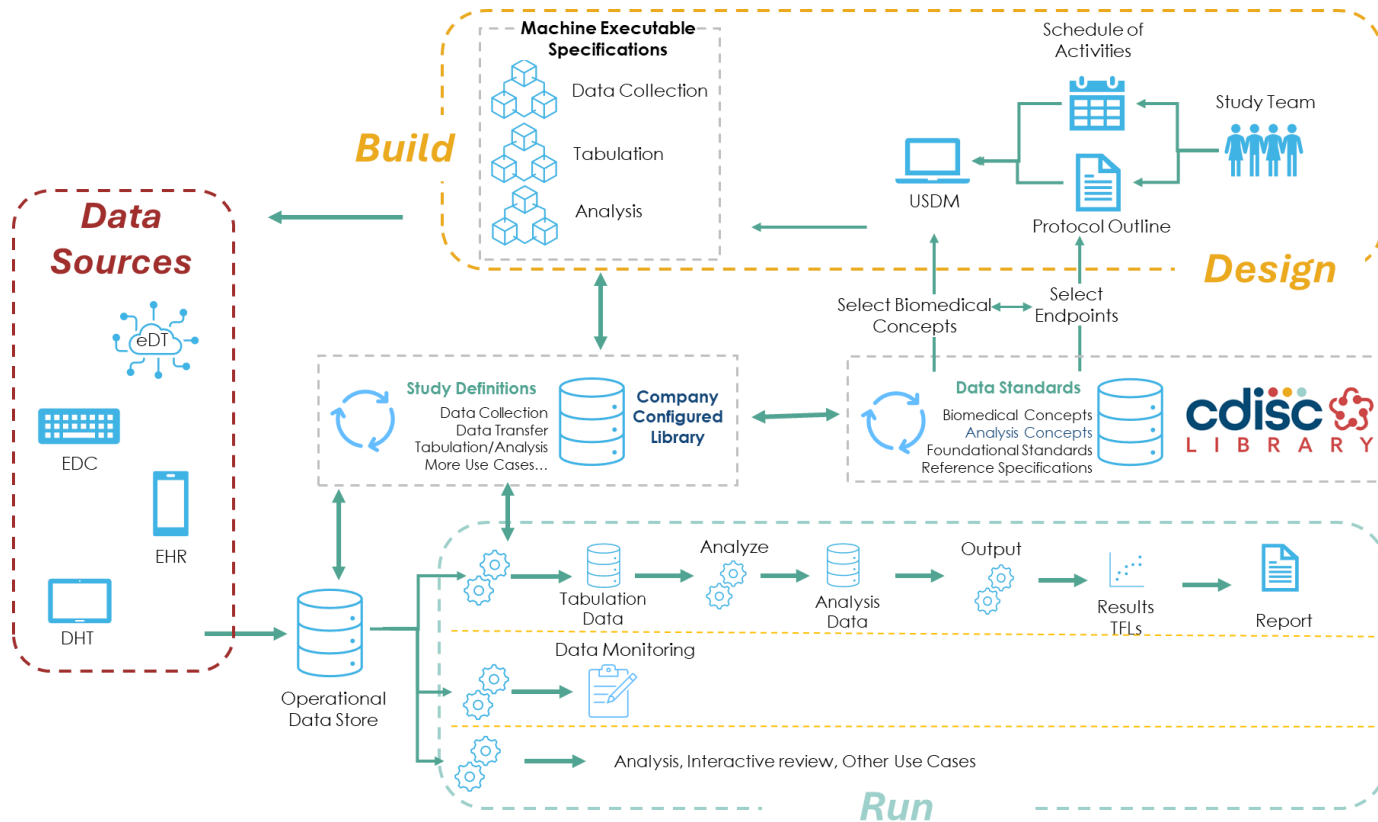
## **Creating a Digital Ecosystem:**

- Unified Study Definitions Model (USDM) as the foundation
- Digital Protocol feed downstream artifacts
- Linked Biomedical Concepts enrich meaning
- Machine-readable, interoperable data flows
- Open-source tools to support transparency and adoption

**Goal:** Fully digital, standards-driven clinical development lifecycle



# 360i Phase 1





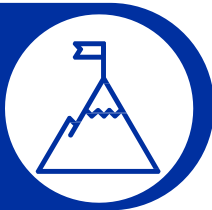
# Design

## Achievements



- Detailed user stories for Study Design and Study Amendments
- Digital Schedule of Activities linked to biomedical concepts

## Challenges



- Interpreting protocol information to identify what data is needed
- Fit for purpose Study Design software



## Insights

Linking concepts works, but **concept groupings** are needed for usability

- Collection method
- Data type
- Category
- Analysis

Open-Source USDM SoA **authoring tool** is needed



# 2025 AI Innovation Challenge

## Use Case 1: Key Results

- Most popular use case
- AI used to
  - Extract legacy protocols into USDM
  - Enable search, reuse, comparison
  - Support downstream automation

Submissions: 13

Winners:

- Faro Health
- Zifo RnD Solutions

### PROBLEM

Study design is manual, error prone, and time consuming and creates limited access to legacy study information

### SOLUTION

USDM-centric digital library of legacy study definitions

### CHALLENGE

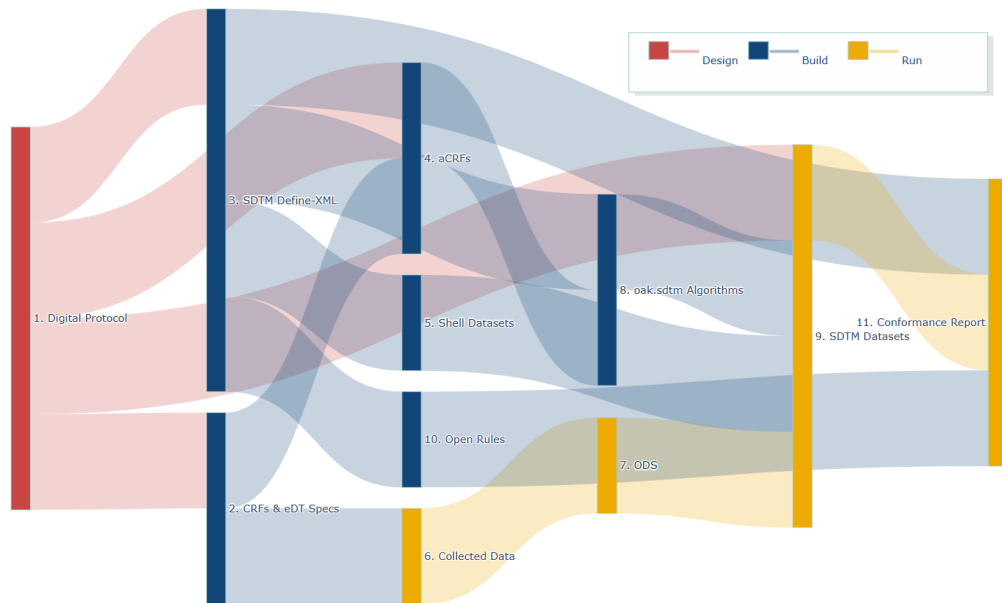
Demonstrate AI extraction of legacy study information in USDM and interface for searching and downloading content



# CDISC 360i Demos

cdisc CDISC 360i — Design / Build / Run

360i Flow from Digital Protocol



## Resources

[Open Google Colab Notebook](#)

## Deliverables (click to open demo)

1. Digital Protocol
2. CRFs & eDT Specs
3. SDTM Define-XML
4. aCRFs
5. Shell Datasets
6. Collected Data
7. ODS
8. oak.sdtm Algorithms
9. SDTM Datasets
10. Open Rules
11. Conformance Report

## Direct Video Links:

1. [Digital Protocol](#)
2. [CRFs & eDT Specs](#)
3. [SDTM Define-XML](#)
4. [aCRFs](#)
5. [Shell Datasets](#)
6. [Collected Data](#)
7. [ODS](#)
8. [oak.sdtm Algorithms](#)
9. [SDTM Datasets](#)
10. [Open Rules](#)
11. [Conformance Report](#)

[Download](#) diagram to access the notebook and videos



# GitHub Repositories Used in the Demo

| Purpose                       | Location                                                                                                                                                                                            |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Study Definition Workbench    | <a href="https://github.com/data4knowledge/study_definitions_workbench">https://github.com/data4knowledge/study_definitions_workbench</a>                                                           |
| Open Study Builder            | <a href="https://gitlab.com/Novo-Nordisk/nn-public/openstudybuilder/OpenStudyBuilder-Solution">https://gitlab.com/Novo-Nordisk/nn-public/openstudybuilder/OpenStudyBuilder-Solution</a>             |
| Study USDM documents          | <a href="https://github.com/cdisc-org/360i/tree/main/data/protocol/LZZT/usdm">https://github.com/cdisc-org/360i/tree/main/data/protocol/LZZT/usdm</a>                                               |
| USDM validation utility       | <a href="https://github.com/pendingintent/cdisc-json-validation">https://github.com/pendingintent/cdisc-json-validation</a>                                                                         |
| CDISC CORE Rules Engine       | <a href="https://github.com/cdisc-org/cdisc-rules-engine">https://github.com/cdisc-org/cdisc-rules-engine</a>                                                                                       |
| CRF creation                  | <a href="https://github.com/lexjansen/cdisc360i-pocs">https://github.com/lexjansen/cdisc360i-pocs</a>                                                                                               |
| Trial Design Dataset creation | <a href="https://github.com/pendingintent/cdisc-usdm-utils">https://github.com/pendingintent/cdisc-usdm-utils</a>                                                                                   |
| Define-XML template creation  | <a href="https://github.com/dostiep/360i">https://github.com/dostiep/360i</a>                                                                                                                       |
| Define-XML creation           | <a href="https://github.com/swhume/template2define">https://github.com/swhume/template2define</a>                                                                                                   |
| Raw subject data              | <a href="https://github.com/alidootson/UpdatedCDISCPilotData/tree/main/UpdatedCDISCPilotData/C-DASH">https://github.com/alidootson/UpdatedCDISCPilotData/tree/main/UpdatedCDISCPilotData/C-DASH</a> |



# Accessing the 360i Notebook

The notebook displayed in this demonstration can be found here:

<https://github.com/cdisc-org/cdisc-360i-notebooks>

- The repository is under continued development; follow repository for latest releases
- There are multiple notebooks
- Raise any issues encountered while using notebooks
- To contribute to the CDISC project:
  - Fork the repo and create a new branch
  - Once changes are made and tested, raise a Pull Request



# Using GitHub

**The GitHub repositories used in the demo are public and use open-source libraries**

The process for using these libraries is consistent:

- Fork the original public repository into personal GitHub repository
- Create new branch in personal GitHub repository
- Edit code to meet individual requirements, add features or fix issues
- Commit changes to personal branch
- If proposing change to parent (original) branch to share with wider community:
  - Raise a Pull Request with original repository owner
  - Original repository owner will review the Pull Request and work with requestor to evaluate and test Pull Request
  - Merge new code in Pull Request with original parent branch



# 360i Phase 2: Key Deliverables

---

**Digital Protocol & Impact Analysis Reporting:** Machine-readable protocol model with automated amendment and impact analysis

---

**Automated Define & Dataset Generation:** Generation of Define.xml and datasets integrated with study workflows

---

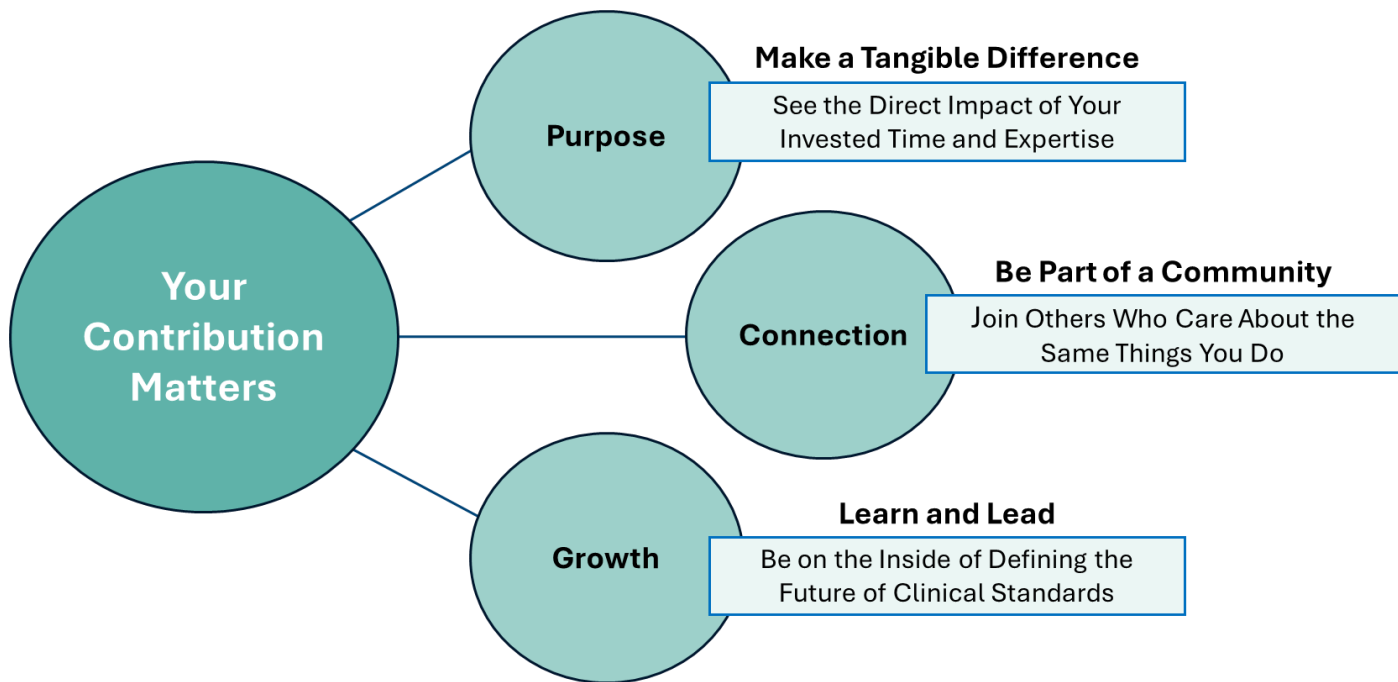
**Digital Data Transfer Agreement Model:** Technical DTA model and standardized template aligned with the digital protocol

---

**Standardized Analysis Concept Model:** Configurable, machine-readable analysis concepts driving automated analysis artifacts

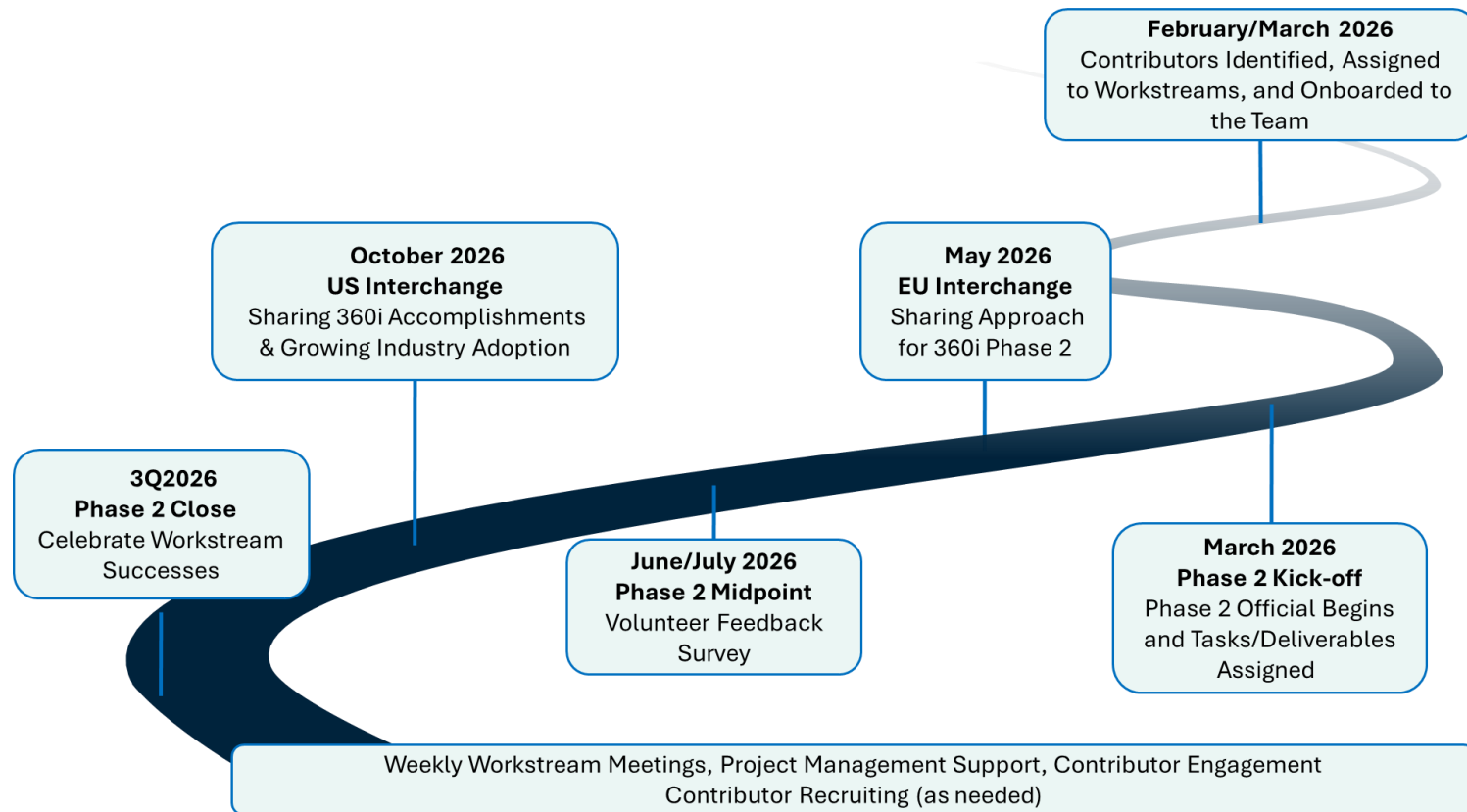


# Join us in this transformation



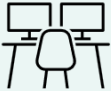


# 360i Phase 2 Journey





# Workstream Opportunities

|  | Study Design                                                                                                                                                                                   | eCRF and Define-XML Generation                                                                                                                                                 | Digital Data Transfer Agreement                                                                                                                                                | Analysis Concepts/ Derivation Concepts (AC/DC) Automation                                                   |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
|  | No additional needs currently                                                                                                                                                                  | <b>Recruiting:</b> <ul style="list-style-type: none"><li>• Subject Matter Experts/Opinion Leaders</li><li>• Content developers</li><li>• Reviewers</li><li>• Testers</li></ul> | <b>Recruiting:</b> <ul style="list-style-type: none"><li>• Subject Matter Experts/Opinion Leaders</li><li>• Content developers</li><li>• Reviewers</li><li>• Testers</li></ul> | <b>Recruiting:</b> <ul style="list-style-type: none"><li>• Subject Matter Experts/Opinion Leaders</li></ul> |
|  | <p>Effort is deliverable based.</p> <p>Impact comes from hands-on work that delivers results.</p> <p>Success requires a committed core group of contributors engaged ~ 20 hours per month.</p> |                                                                                                                                                                                |                                                                                                                                                                                |                                                                                                             |



# 360i Phase 2 Next Steps

**Complete** Volunteer Contributor Interest Form  
&  
**Share** the Form with Colleagues

**Link in Post-Webinar Email & on CDISC 360i webpage:** [https://www.cdisc.org/standards/cdisc-](https://www.cdisc.org/standards/cdisc-360i)



**CDISC 360i is transforming clinical research by digitalizing study design through analysis making metadata interoperable across the entire study lifecycle.**

With connected standards and end-to-end automation, 360i eliminates manual tasks, enables AI, ensures traceability and consistency, and empowers the industry to deliver results faster, with better quality, and at a lower cost accelerating the delivery of new therapies to patients.



**Volunteer**

**Join CDISC**





# Use Case 2 – SDTM

*Presented as a live demo*



# Use Case 3 – Analysis Results



# Our results often persist in presentation format

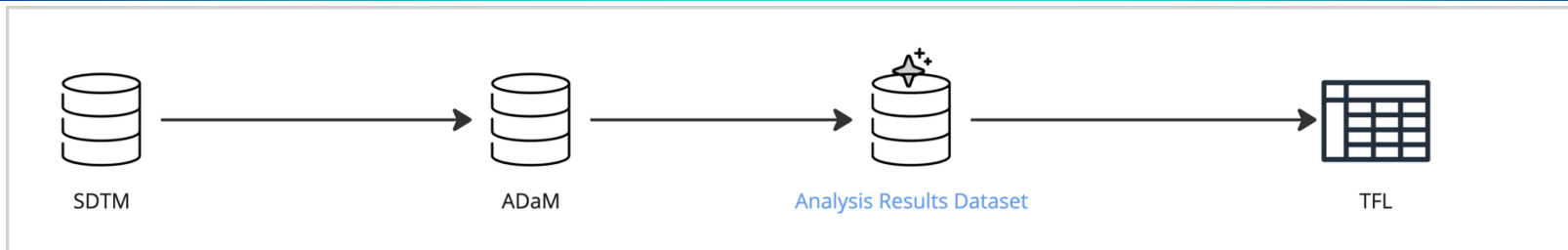


Image source: <https://www.cdisc.org/sites/default/files/pdf/CDISC-ARS-Webinar-June15-2023.pdf>

## Today: Results as TFLs

- Results are often stored in files that focus on report generation (rtfs, pdfs, etc.)
- Not readily machine readable
- The format follows the layout of the shell
- A change in rendering may require re-computation of results

# Store as results in datasets instead



## Analysis Results Datasets

- Separates results from presentation layer
- Enables traceability of results through metadata capture
- Enables reuse, comparison, and downstream applications (data viz, AI)

# Our journey to ARS at Novartis



**Metadata  
driven  
ARDs**

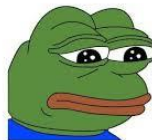


**ARS too  
complex**



**Build custom  
model**

REGRETS



THAT IS ALL I HAVE

**Too difficult**



**Back to the  
ARS**

# ARS Deepdive

## ARS Global

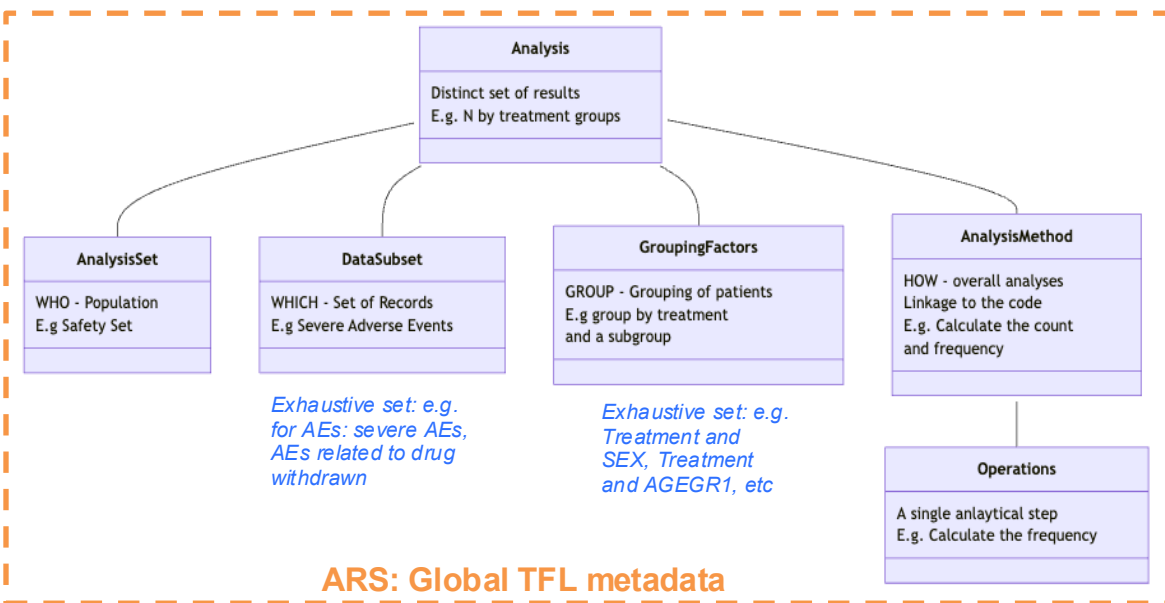
*TFL metadata*

## Operations

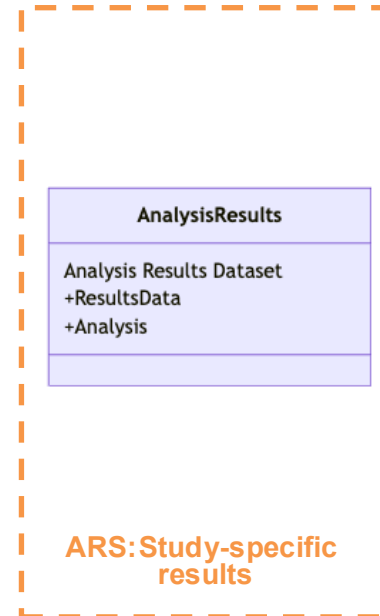
*Engine*

## Results

*ARDs*



**ARS: Global TFL metadata**





# Mosaic an open-source ecosystem

## Mosaic components

## Open-source building blocks

### Automation layer

**TFL  
Metadata**



**TFL  
Engine**



### User-facing layer

**User  
Interface**



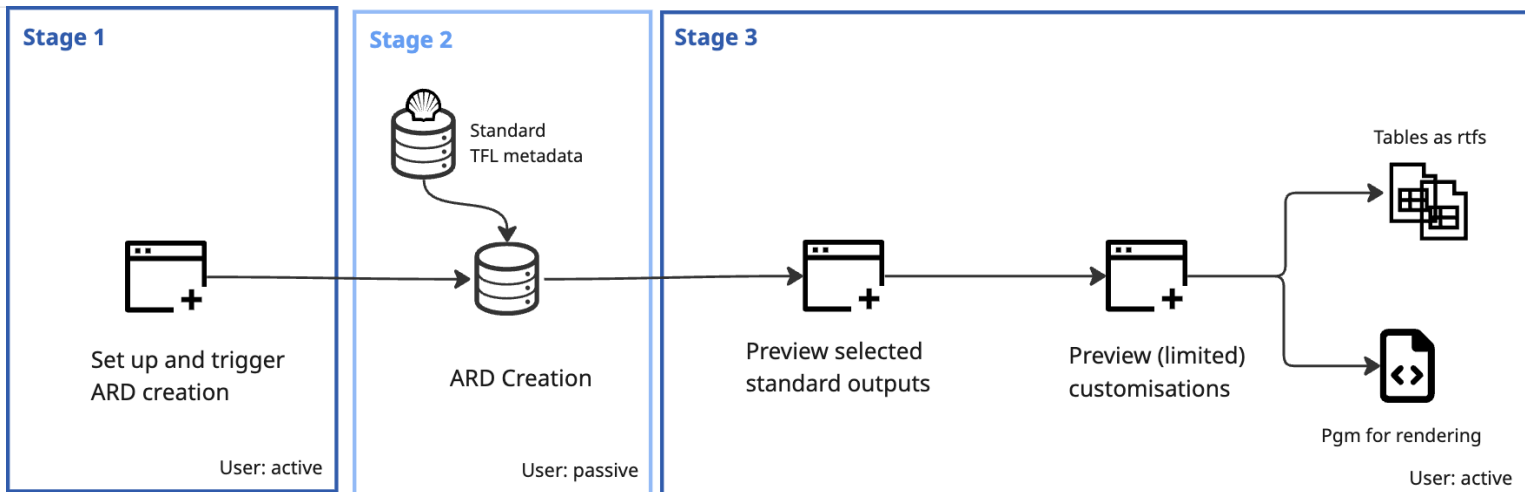


# Mosaic: user journey



## Mission

Mosaic: **automate the generation of standard TFLs** via Analysis Results Datasets while delivering excellent **user-experience**.



# Demo time!

CSTDCDSA / CDSA\_STUDY3 / csr\_5

2
 @DOYLEOR1

Home > Tables Dashboard > Table Configuration

< Table Configuration

Population Flag \*  
ADSL.SAFFL EQ Y

Filters \*  
ADAE.TRTEMFL EQ Y

Column Subgroup \*  
Grouping by `ADSL.SEX`

☒ Total Column

Sort by Treatment Column ⓘ

Clear All Regenerate Save

T\_AE\_002

HTML Preview
Annotated Shell

Table 14.3-1.1.1 Mosaic preview

Adverse Events by SOC and PT by gender

|                                            | Placebo        |              | Treatment A    |              | Total          |              |
|--------------------------------------------|----------------|--------------|----------------|--------------|----------------|--------------|
| Primary system organ class                 | Female (N=176) | Male (N=376) | Female (N=194) | Male (N=356) | Female (N=370) | Male (N=732) |
| Preferred term                             | n (%)          | n (%)        | n (%)          | n (%)        | n (%)          | n (%)        |
| Number of subjects with at least one event | 93 (52.8)      | 163 (43.4)   | 91 (46.9)      | 170 (47.8)   | 184 (49.7)     | 333 (45.5)   |
| Blood and lymphatic system disorders       | 1 (0.6)        | 1 (0.3)      | 3 (1.5)        | 3 (0.8)      | 4 (1.1)        | 4 (0.5)      |
| Leukocytosis                               | 0              | 0            | 0              | 2 (0.6)      | 0              | 2 (0.3)      |
| Splenomegaly                               | 0              | 1 (0.3)      | 0              | 0            | 0              | 1 (0.1)      |
| Lymphadenopathy                            | 0              | 0            | 1 (0.5)        | 1 (0.3)      | 1 (0.3)        | 1 (0.1)      |
| Splenic cyst                               | 1 (0.6)        | 0            | 0              | 0            | 1 (0.3)        | 0            |
| Iron deficiency anaemia                    | 0              | 0            | 1 (0.5)        | 0            | 1 (0.3)        | 0            |

# Breakout Group Discussions

## 1. Leaning validation

- How could metadata-driven approaches simplify validation without reducing confidence?
- What validation steps could be reduced or eliminated?
- Where is human oversight still needed in the process?

## 2. Efficiency vs readability

- As automation adds abstraction, how do we balance efficiency with transparency and inspectability?
- Should inspection focus on code, metadata, or outputs?
- Is readable code still the goal, or is “inspectable pipelines” the new standard?

## 3. Metadata automation and / or AI

- How do we blend AI within metadata-driven workflows?
- What are the core opportunities for AI e.g., extraction, mapping, suggestion,
- Are there areas where we want to prioritise a deterministic approach?

## 4. Uptake in regulatory submissions

- What level of validation, transparency, or traceability would regulators need to trust automated outputs?
- Where do current regulatory processes, guidance, or expectations lag behind what automation enables?
- Are the biggest barriers technical, organizational, or cultural (e.g. risk tolerance)?