Towards Automating Statistical Programming
and Data Transformation

Ganesh Gopal

Ephicacy Consulting Group

Agenda

- Components of data flow
- Current automation in the industry
- Challenges in automation
- Metadata and importance
- Towards a fully automated solution
- Questions

EPHICACY

Data Value Chain

Collect

Collect and store data as received from many sources and in several formats, e.g., annotated CRFs. EDC, Labs, Imaging IVRS, etc.

Process

Validate and accept data acquired from source from metadata driven and user defined rules e.g., DMW, JReview, JMP, RBM…

Transform

Use metadata driven and user defined transformation to standardize and integrate the data

- SDTM mapping

Analyze

Deliver the data as reports, visualizations, or analytics sources for querying, mining, and A.I. algorithms

- ADaM Derivation
- TFLs

Protocol → **eCRF** → **SDTM** → **ADaM** → **TLFs** → **CSR**

EPHICACY

Current automation status of various components and challenges…

Data Transformation & Programming Automation: *Lay of the Land*

Collect

- No collection standards
- No harmonization between CRFs and EDC set up

- Collection metadata in EDC system
- Processes to manage libraries
- In-house MDR

Process

- No intermediate data model between EDC and SDTM
- Minimum set-up for real time data review

- Having an efficient CDR
- Systems for query handling, visualization, RBM
- DMW, JReview, JMP, Spotfire, Cluepoints etc.

Transform

- Manual mapping for each study
- Specs are in spreadsheets
- Variability across studies

- In-house tools, e.g., SDTM mapping with P21 validation
- Off the shelf products
- Rely on CRO tools

Analyze

- Manual process for ADaM & TFL
- No standard ADaM definitions or TFL shells
- Submission docs at later stage

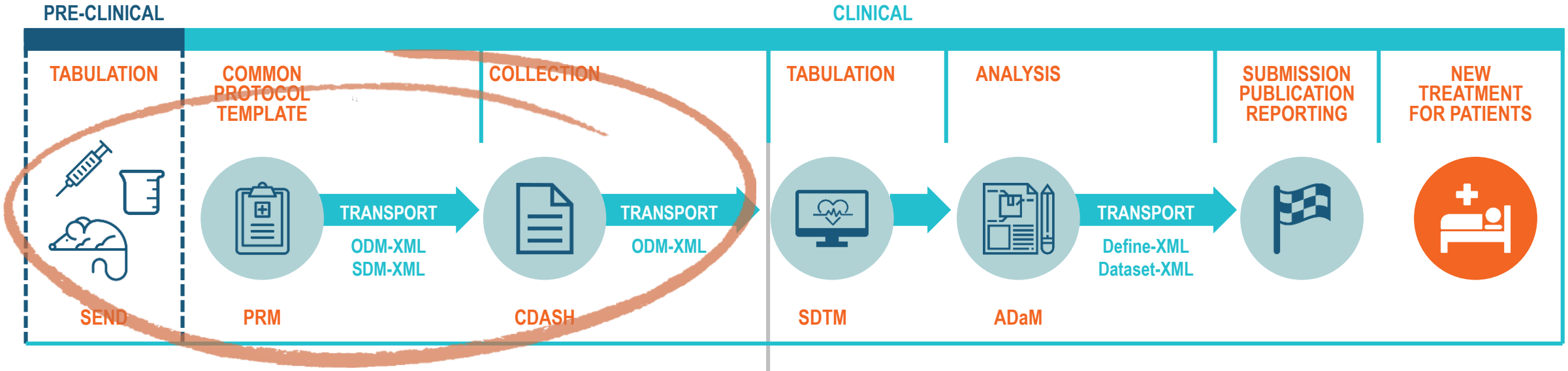
- In-house tools/ systems
- Standard macros
- Off the shelf products
- Rely on CRO tools

Manual

Some Automation

EPHICACY

Challenges in Automation



UPSTREAM

- No proper metadata governance
- Weak connection between collection & review model
- Lack of metrics around deviations / compliance
- Need single source of truth
- People, leadership & technology

DOWNSTREAM

- Need for an efficient CDR
- Standardization of ADaM & TFL standards
- Uniformity of standards between Sponsor & CRO

EPHICACY

Metadata, Metadata, Metadata!

- You are not the only one doing the *“Plumbing”*

Maintaining global and TA standards is an enormous task:

- Manual processes (spreadsheets) for content management, versioning and distribution of metadata
- Implemented in silos increasing variability
- Requires manpower to manage and govern the metadata
- Requires change management
- Periodic version releases in the standards, managing multiple versions

EPHICACY

Metadata

- continued…

Implementing standards is another challenge:

- Integration of multiple technology platforms
- Process change, training needs
- Diverse needs of study protocols
- Distributed teams (Different business functions, multiple vendors)
- Periodic version releases in the standards, managing multiple versions
- Leadership and management should champion the cause

EPHICACY

So…
We need to overcome the Metadata obstacle before we come up with an end-to-end solution for automation.

EPHICACY

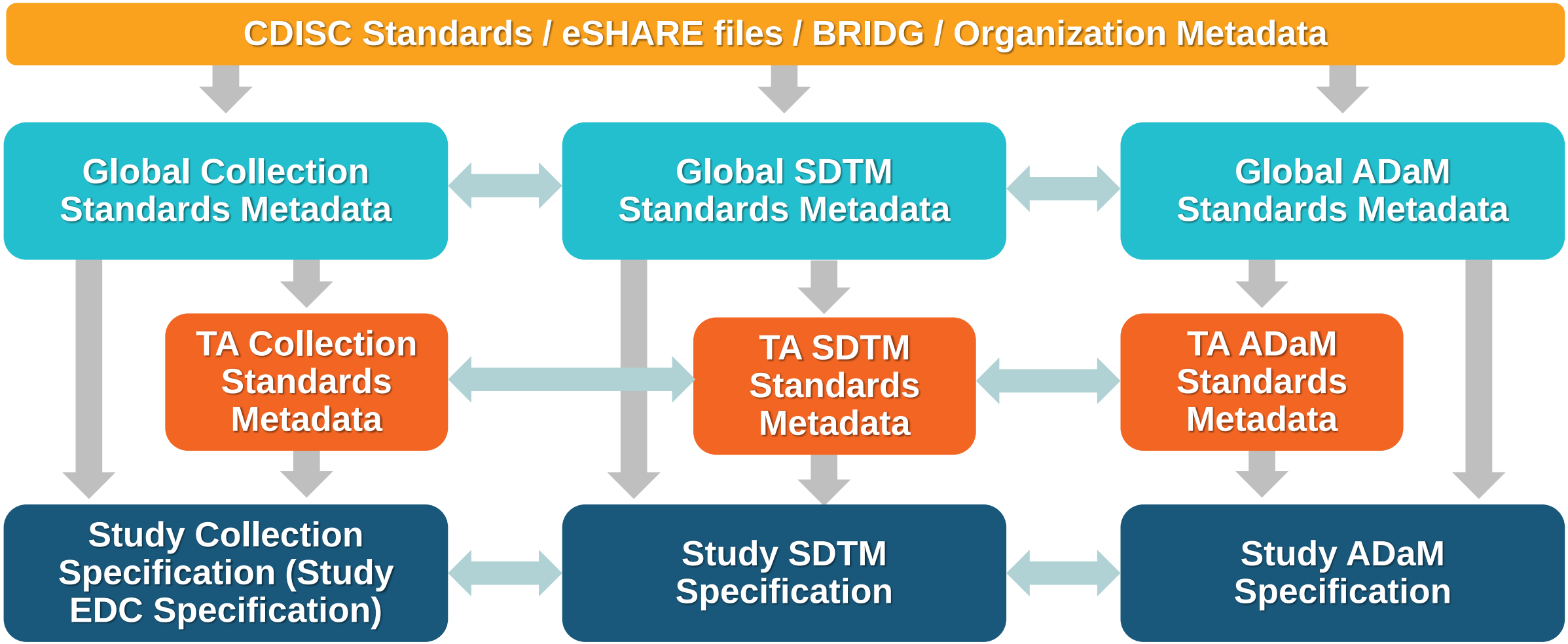
A Comprehensive Metadata Repository:

The Nerve Center of Automation

- Capture metadata standards including transformations / mappings
- Bulk import or cherry pick from standards into study metadata for consumption
- Improve operations by using advanced comparison and collaboration tools
- Import from key or legacy systems / Export for regulatory review and integration
- Capture traceability, versioning, and lifecycle
- Extend metadata, configure workflows, and integrate with Data Repository in a controlled environment

EPHICACY

The Ideal MDR Architecture



EPHICACY

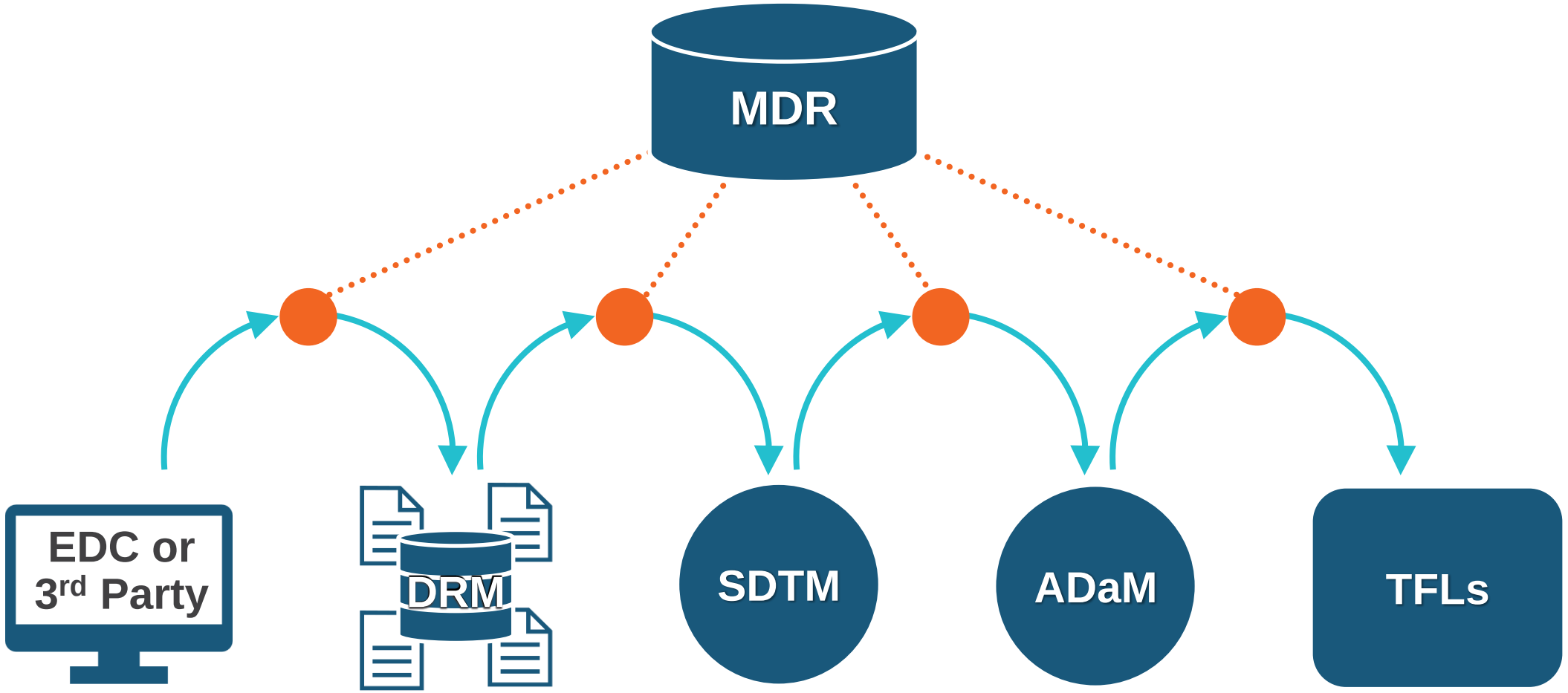
In addition…

Clinical Data Repository

- Standardize handling of trial data, program files, and data sets (collection to locking to archiving)
- Handle any clinical data model (relational data or files) uploaded cumulatively or incrementally
- Run programs and batch jobs with traceable dependencies and automated re-runs upon update
- A fully integrated data review model that connects well with the backend EDC

EPHICACY

A Metadata Driven End-to-End Value Chain



And …

Statistical Computing Environment

- Coordinate tasks, track work and enforce processes within the system
- Maintain traceability and lineage across programs, data and outputs with version control
- Easily define and manage role-based access, blinding, sequestering, and partner access
- Reliably audit access, changes and usage across all functions

EPHICACY

The Ecosystem

Governance Enrichment

Metadata Repository

- Collect metadata standards
- Store mappings
- Import / export of standards

Standardize Validate

Data Repository

- EDC to data review models
- Store program files & datasets
- Transform & stage

Establish Framework Process & Control

Computing Environment

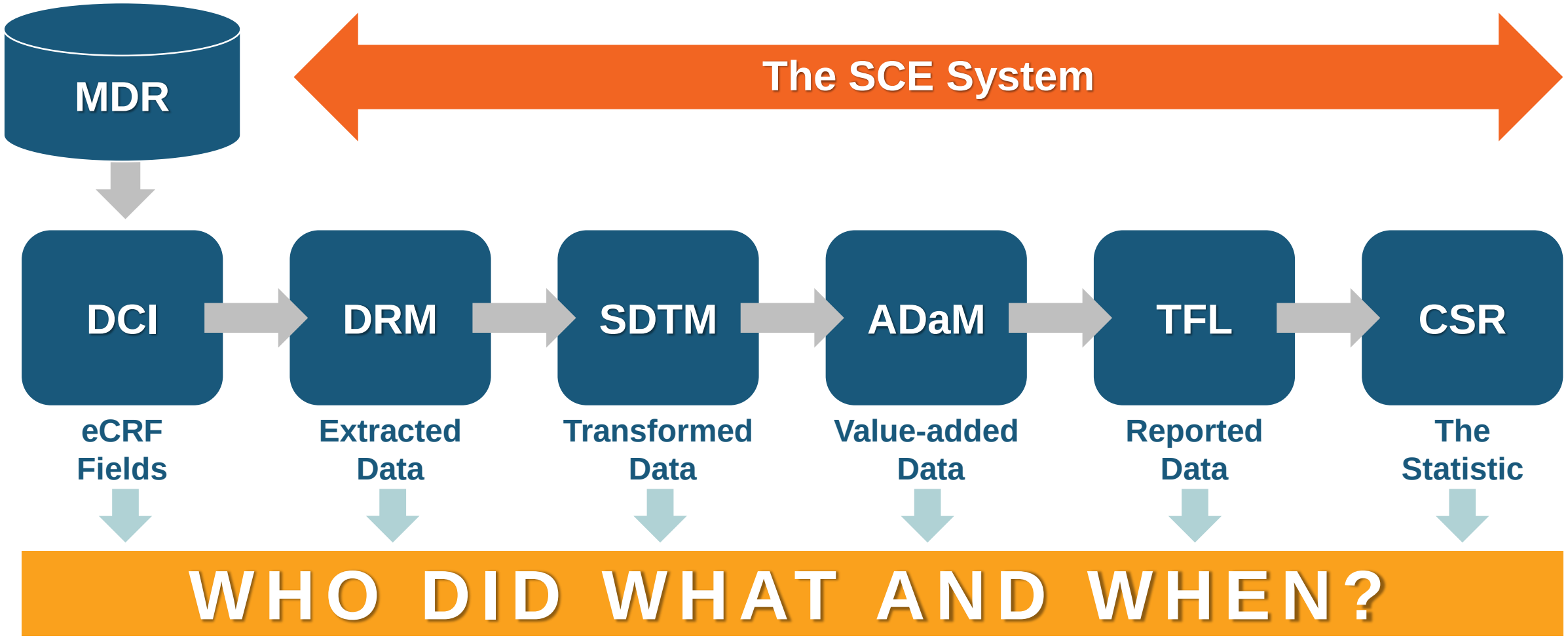
- Coordinate tasks
- Track dependencies
- Role based access

WORK FLOW
Assignments, Comments and Approvals

End to End Traceability

EPHICACY

From Data to a Statistic – The Story
Traceability and Dependency



Conclusion

- A fully automated data transformation and programming system is possible
- But, the upstream challenges need attention and effort
- Metadata governance, stewardship and enrichment is key
- A fully automated downstream system needs to be in partnership with a metadata driven upstream engine

EPHICACY

Questions?

EPHICACY