

DS04 - From Artifact to Metadata: A Divergent Approach to Automation through Standardization

- **Nicolas Rouillé**, Cytel Inc., Geneva, Switzerland
- **Camie Britton**, Cytel Inc., Waltham, USA
- **Angelo Tinazzi**, Cytel Inc., Geneva, Switzerland

PHUSE EU Connect 2022

13th-16th November 2022

Paper DS04 – Data Standards (DS)



Presentation Discussion

- **Overview: Leveraging Metadata**
 - From a Sponsor perspective
 - From a CRO perspective
 - Perspective-specific challenges
- **Addressing CRO MDR usage challenges**
 - Cytel's approach
 - Case study: production of summary tables, listings and figures
- **Metadata Automation Outcomes**
 - Impact on automation
 - Timelines
 - Resources
 - Automation project oversight
- **Additional Considerations**
- **Conclusion & Summary**



Overview: Leveraging Metadata



Leveraging Metadata: Sponsor Perspectives

Metadata Repository (MDR) Requirements

- ✓ **Data Standards Governance & Quality Assurance**
 - Standardize and reuse content across a single Sponsor's technology stack
 - Compliance between a Sponsor's standards and CDISC requirements
 - Version control
- ✓ **Rapid study start-up / EDC build and timely delivery of submission-ready outputs**
 - Parallel protocol development and EDC development using metadata
 - Automating aCRF & SDTM generation → standardized notation and mapping
 - Elimination of double programming
- ✓ **Flexible and efficient MDR solutions**
 - Off-the-shelf software with Sponsor-specific customization capabilities with rapid deployment timeframes and lean maintenance requirements
 - Ability to deploy across wide-range of roles (data management, biostatistics, statistical programming) and franchises within the Sponsor organization

Challenges

- Data standards defined, maintained and used across many stakeholders
 - Internal: various Therapeutic Areas (TA) with disease / indication-specific endpoints
 - External: multiple CROs and outsourcing modalities (FSP, project-based service providers, etc.)
- Quantification and assurance of Sponsor standards alignment with CDASH / CDISC requirements
- Increased use of non-EDC data sources (biomarkers, PROs, EMR/claims, wearables, etc.) within a Sponsor's clinical trial portfolio

Leveraging Metadata: CRO Perspectives

Metadata Usage Requirements

✓ Data Standards Governance & Quality Assurance

- Standardize and content reuse across multiple EDC platforms, driven by Sponsor EDC system selection for studies awarded to CROs
- CROs must accommodate each Sponsor's standards, while aligning with CDASH & CDISC requirements
- Version control and audit trail capabilities must support each Sponsor's delivery and audit expectations

✓ Need for mature, scalable and cost-effective MDR solutions

- MDR solutions still maturing with respect to use of metadata to standardize and automate beyond protocol development, eCRF generation and basic SDTM mapping activities
- MDR use within CROs must accommodate and partition the varying standards of all Sponsors within a CRO's customer base
- MDR software must (quickly) deploy across wide-range of Sponsor projects and biometrics functional staff who often reside in different companies
- MDR user interface must be simple with modest training demands
- Metadata solutions must deliver value and drive efficiencies within CRO margin requirements

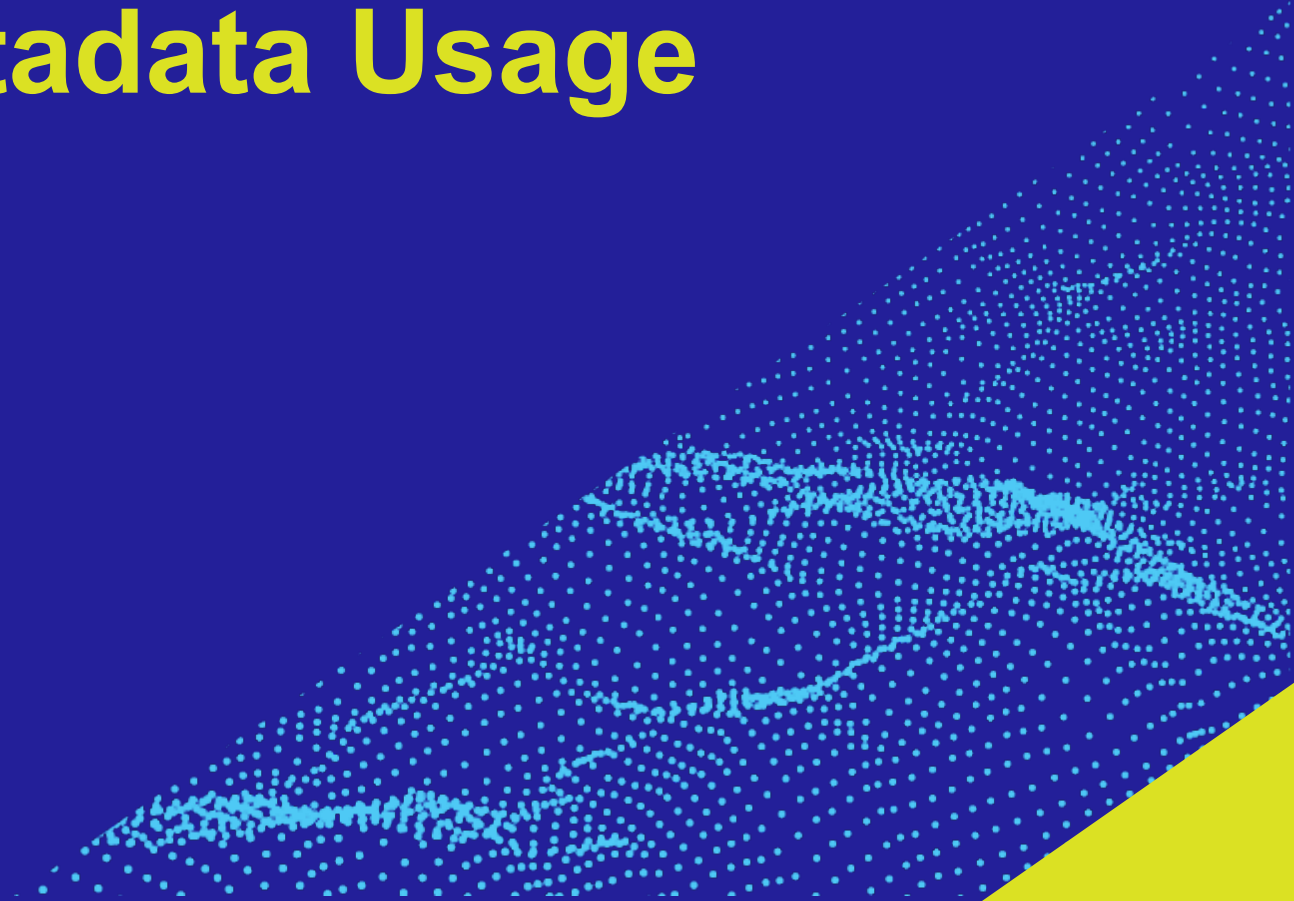
Challenges

Biometrics services are routinely outsourced across multiple organizations

- Protocol and eCRF development often sit in different organizations, with Data Management and biostatistics/stats programming services routinely outsourced to different CROs
 - Limits the flow down of metadata between protocol development, eCRF creation, and SDTM generation activities
 - CROs must deploy numerous standards libraries for eCRFs / edit checks across multiple EDC platforms to service a wide range of Sponsor outsourcing demands
- Sponsor project teams often deviate from CDASH / CDISC standards during eCRF review & approval
 - Hampers automation of SDTM mapping + ADaM generation via MDR solutions
- Elimination of double programming must be agreed with Sponsor
 - Some reluctance to dispense with what is seen as an industry gold standard for quality assurance when outsourcing



Addressing CRO Metadata Usage Challenges



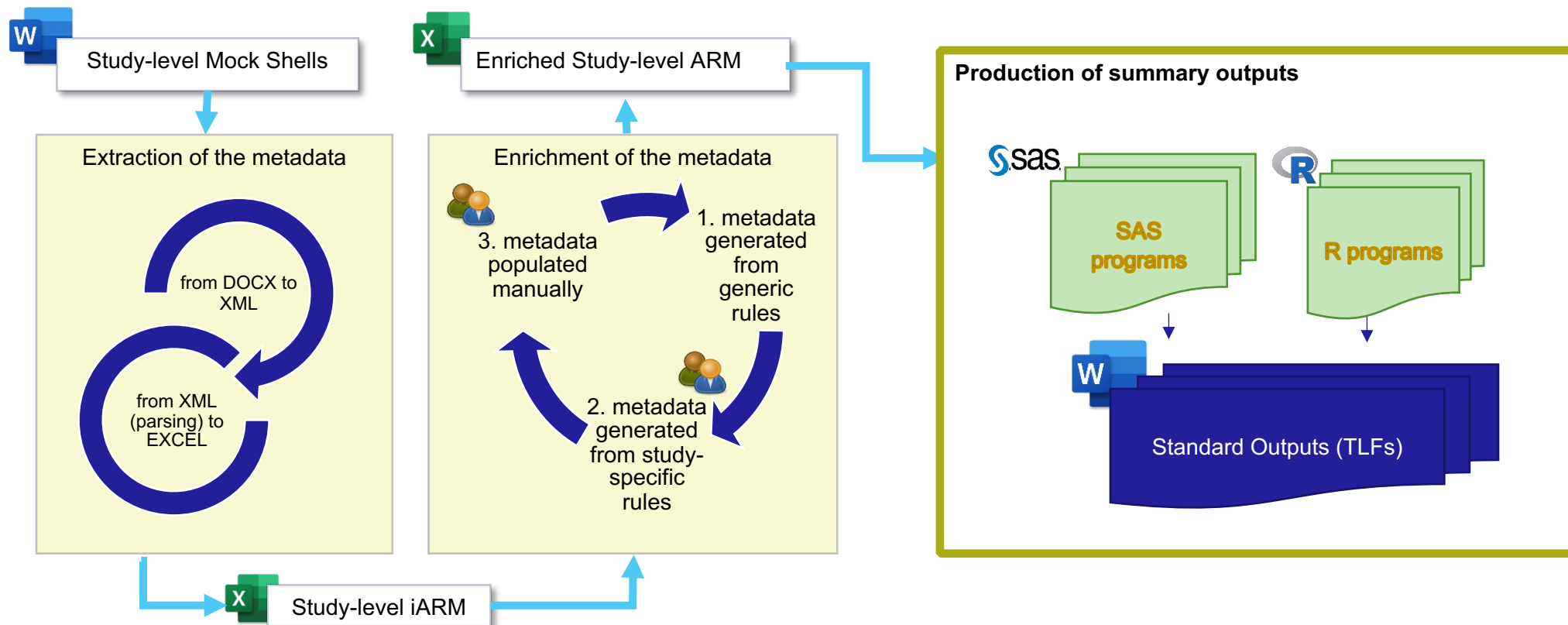
Addressing CRO-specific Metadata Use Challenges

Cytel's Approach

- Our automation team began by **partitioning the data workflow, based on the distinct types of deliverables** we produce for our clients (CDASH, SDTM, SAP, study mock shells, ADaM, TLFs). We then identified what metadata could be captured from each of these artifacts, **pinpointing the metadata which should be enriched in order to fully automate the production of a given deliverable**.
- This strategy allows the implementation of a **user-centered approach** by distilling the overall challenge into smaller pieces. Each artifact is attributed to a specific function (e.g., Data Management, Biostatistics, Statistical Programming), which provides subject-matter-expertise and directly supports automation tool conceptualization, development and testing.
- This functional segmentation **accelerates the development of new tools**, since multiple tools can be developed in parallel and released **asynchronously**, depending on the progress of their development. This helps to significantly **speed the release cadence of tools**, thanks to quick validation of the use cases and alignment of interdependencies.
- It also **accommodates tool development failure** in the event that a tool does not meet artifact production requirements, without preventing the use of other tools developed for other upstream or downstream steps of the workflow.

Case Study: TLF Generation driven by Analysis Result Metadata (ARM)

‘PRISM’ : Process Overview for automated production of TLFs





Metadata Automation Outcomes



Metadata Standardization & Automation Outcomes

Cytemation

Launched a multi-year strategic initiative (“Cytemation”) spanning 2021 through 2024

- ✓ Dedicated project management oversight and coordination
- ✓ Assigned functional subject-matter experts who also lead and support tool programming / development / testing
- ✓ Support provided by IT where needed for tool integration within Cytel’s statistical computing environment
- ✓ Quality Assurance engagement to support tool validation and testing in accordance with Cytel software deployment SOPs
- ✓ Efficiency (hours saved) targets and associated delivery metrics set for automation in each fiscal year with C-suite reporting and visibility up to Cytel’s Board of Directors
- ✓ Automation efforts linked to operational delivery (SOPs) + quality metrics
- ✓ Productivity gains shared with Cytel customers in terms of lower unit costs for affected deliverables

2021

- SAP Generator – Analysis plan automation
- eCRF / edit check standards & libraries
- CDASH compliance measurement to enable SDTM mapping automation
- Lighthouse – Mock shell automation
- TabMacro – Stats programming macro library
- PRISM – TLF automation
- MDR technology assessments by vendor

2022

- DMC Charter automation
- ADaM automation
- Next releases of Lighthouse & PRISM
- Company wide Innovation Drive focused on automation opportunity identification

2023 (Road Map)

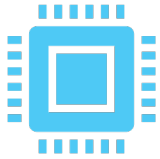
- Version Control system deployed within SCE
- Randomization project automation
- Searchable metadata/artifacts database
- SAP Catalogue (augment SAP Generator)
- SAP Generator next release
- Risk-based TLF QC



Additional Considerations



Additional considerations



Opportunities created by innovative technologies

For instance, R-markdown

- R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into pdfs, Word files and more.
- Cf <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

R Markdown

from  Studio



Integration within sponsors eco-systems



Cytel is providing access to automation tool instances for certain types of deliverables on behalf of our sponsors

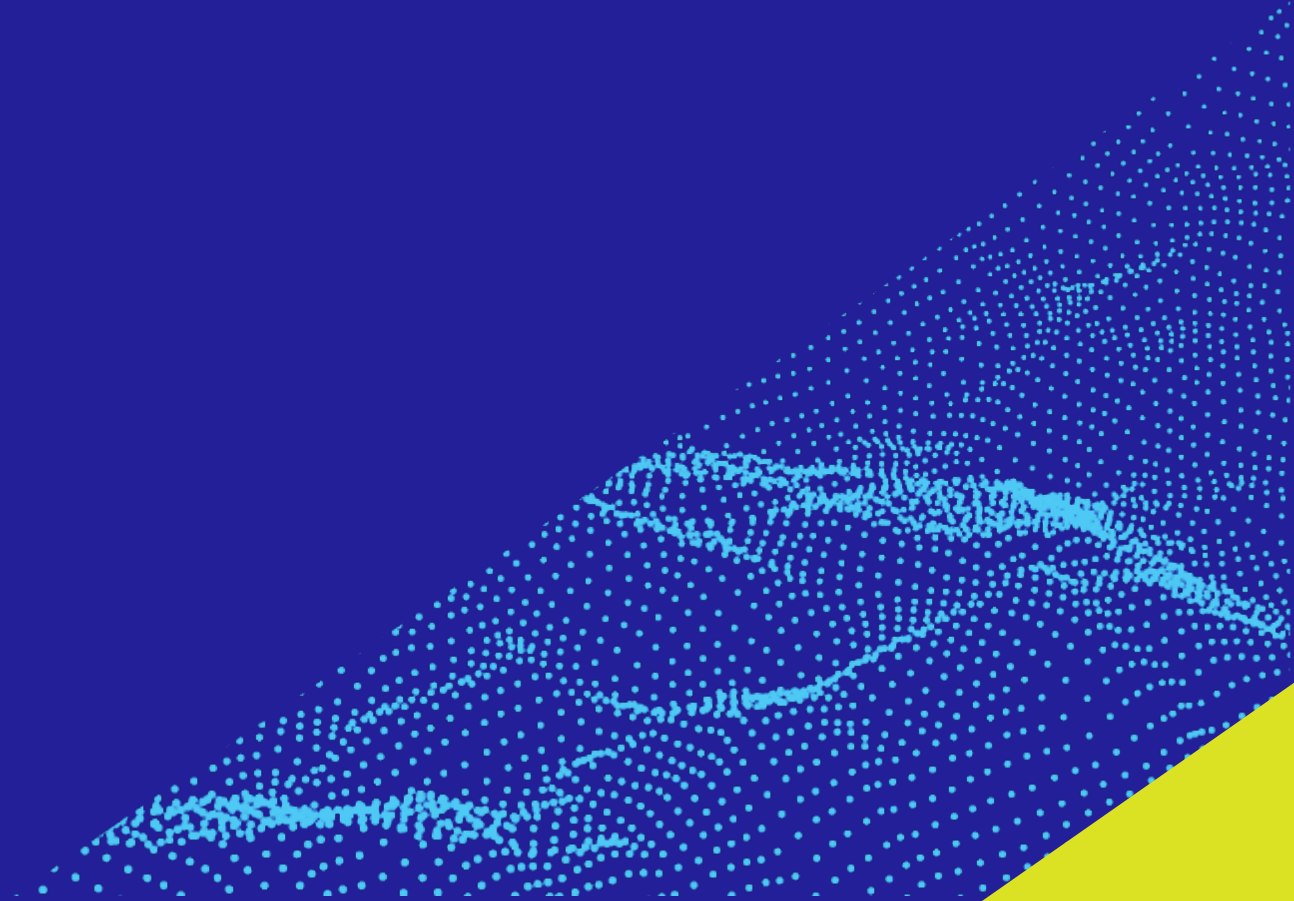
This can be extended to enable sponsor access to these applications for their use across multiple outsourcing configurations (e.g., Project-based services, FSP, etc.).



Scale and adapt to internal organizational changes and growth curves



Summary: Q&A





Thank you!

Contact information:

Nicolas Rouillé

e-mail: nicolas.rouille@cytel.com

Camie Britton

e-mail: camie.britton@cytel.com

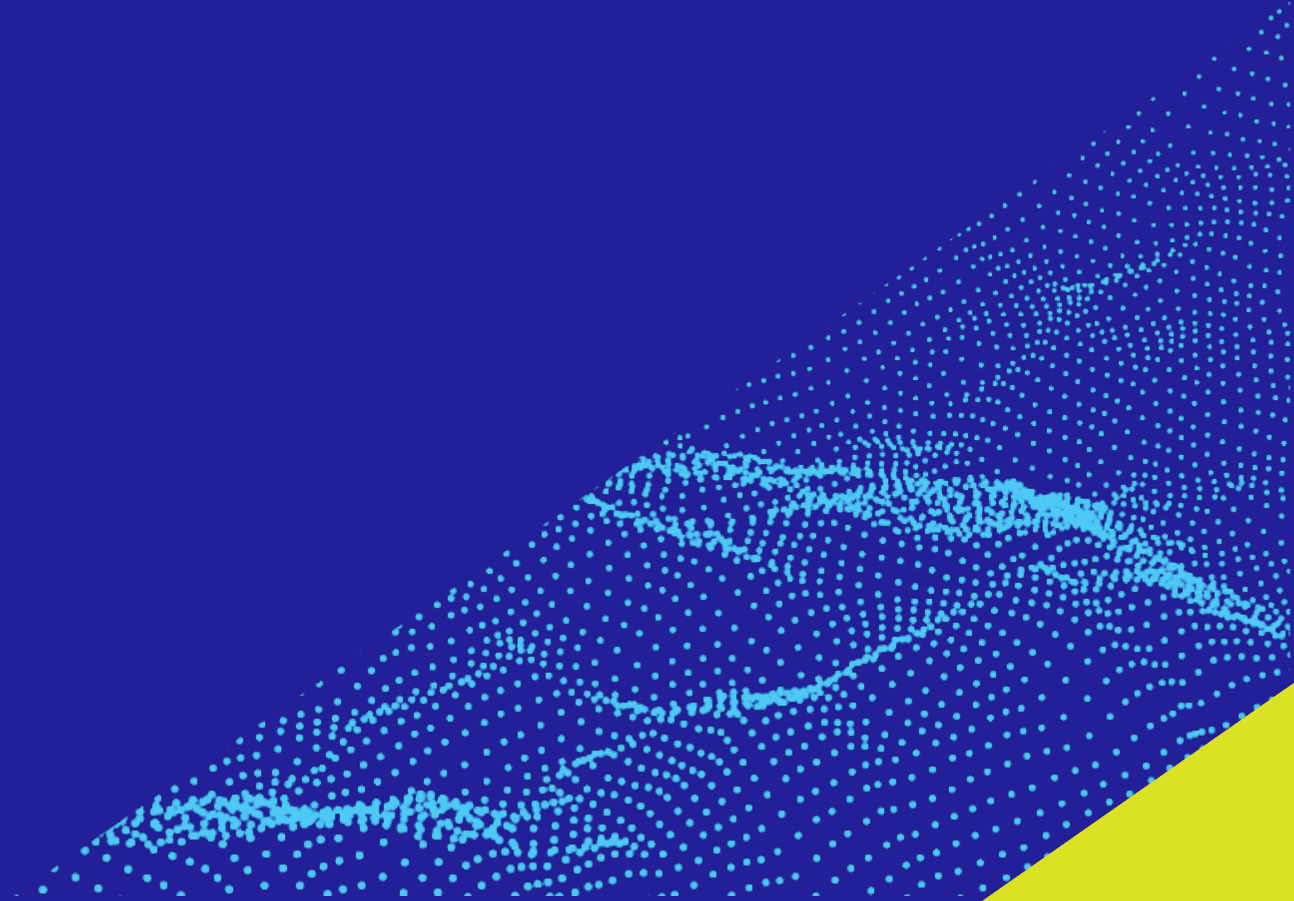
Angelo Tinazzi

e-mail: angelo.tinazzi@cytel.com

Web: www.cytel.com



Appendix



From Artifact to Metadata: A Divergent Approach to Automation through Standardization

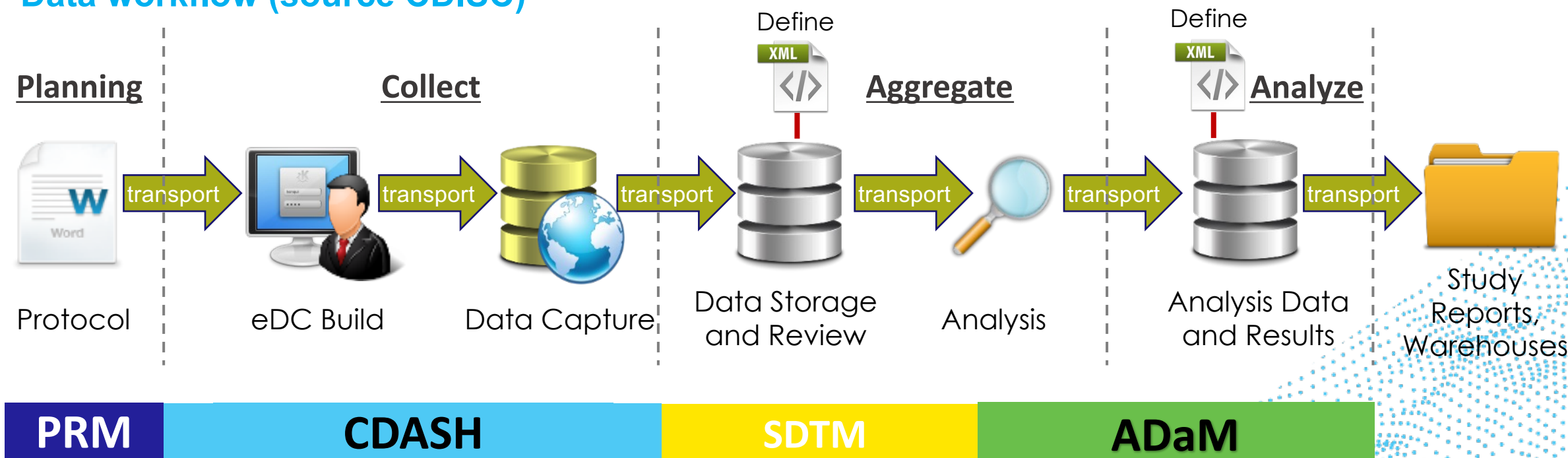
Preamble

The data /metadata engineering pipeline is a hot topic for any company in our industry; usage of metadata is key, with its promise to speed time-to-market and improve adherence to the increasing requirements of our regulated industry.

Deployment of metadata repositories (MDR) for standards management is currently the ‘go-to’ solution to facilitate collaboration along the data workflow between cross-functional stakeholders to increase adherence to industry and sponsor standards while accelerating study timelines.

The challenges associated with leveraging metadata to enable standardization and automation differ from sponsor and CRO perspectives. We explore the **differences and similarities faced by pharma, biotech and services in this regard through examination of the approach Cytel used in the development of its automation tools** and the current and future opportunities the approach enables for sponsors and CROs alike.

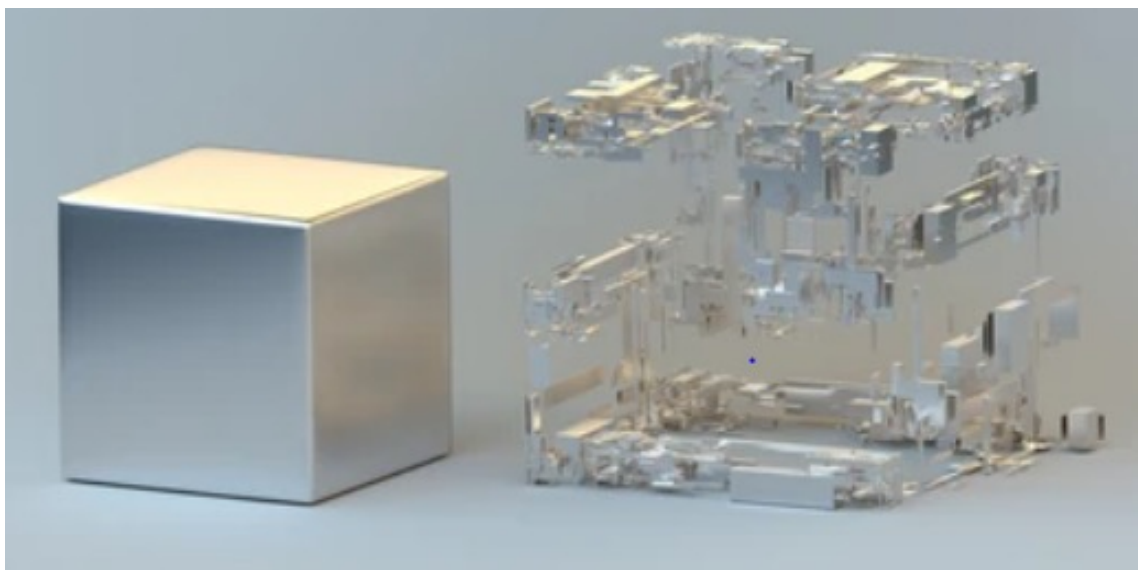
Data workflow (source CDISC)



Controlled Terminology



MDR Model = super-cube



Expectation

Reality

**MDR Model is having multiple dimensions;
roughly, dimensions we might list:**

- Study level (study A, study B, etc)
- Artifact level (Protocol Representation Model (PRM), CDASH, SDTM, ADaM, etc.)
- Standards level
 - International standard (e.g., ISO)
 - CDISC (incl. Multiple IG versions)
 - Company Standards (what release?)
 - Study-specific conventions? Exceptions?
- ... and much more! Some examples:
 - Compound level (NB one compound could have multiple TAs level),
 - TA level (NB one TA could have multiple compounds level),
 - reporting task levels (DMC, IA, FA)

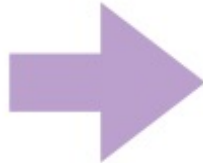


R Markdown : a format for writing reproducible, dynamic reports with R

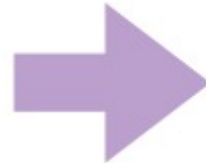
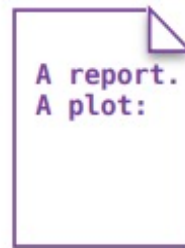
Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more.

Source: <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

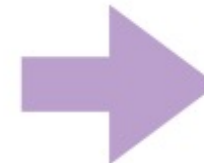
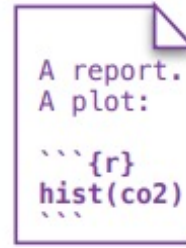
i. Open - Open a file that uses the .Rmd extension.



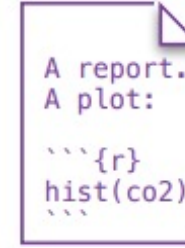
ii. Write - Write content with the easy to use R Markdown syntax



iii. Embed - Embed R code that creates output to include in the report



iv. Render - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.



=



=

