Paper TT-06

From SDTM to a Graph-Based, Multi-Purpose Interface for Data

Eva Kelty, Capish, Malmö, Sweden

Catharina Dahlbo, Capish, Malmö, Sweden

ABSTRACT

As a clinical trial is completed, data is commonly organized and stored according to the SDTM standard, designed to standardize the organization and formatting of data to streamline collection, management, analysis, and reporting. However, because the data remains stored in a traditional table/relational database structure, it can be challenging to gain a comprehensive view of data.

By moving data into a graph-based structure, new possibilities emerge for enhancing cross-functional collaboration. Unlike traditional table-based structures, where relationships between data points are manually defined, a graph-based interface connects related information, making complex relationships easier to visualize and explore through dynamic querying. By improving the accuracy and efficiency of data exploration, these tools support better decision-making and strengthen patient safety.

This session will explore a method for transitioning from tables to a graph-based data structure and provide insight into how this approach can be applied in different contexts to better understand available data.

INTRODUCTION

Clinical research generates an immense volume of data, spanning demographics, adverse events, treatments, laboratory results, and more. Industry standards such as SDTM provide a well-defined framework for organizing this information to ensure regulatory compliance and facilitate downstream analyses. Over time, SDTM has brought structure, consistency, and control to clinical data management. Yet as the industry has mastered structure, a new challenge has emerged - connection.

The data captured so precisely across SDTM domains often remains fragmented, with each domain (AE, LB, EX, DM) representing only part of a patient's story. Understanding the full clinical context frequently requires manual linking between datasets, slowing insight generation and obscuring relationships. As studies become more complex and the demand for real-time analysis increases, this traditional, table-based design can limit exploration.

To support faster, more intuitive understanding, teams across clinical development and safety now need ways to navigate data as they think - through relationships rather than static listings. When connections between events, timing, and outcomes are difficult to see, opportunities for earlier insight and proactive decision-making can be missed.

REFRAMING SDTM DATA AS A GRAPH

Moving SDTM data into a graph structure calls for careful planning to ensure data content is maintained while revealing the natural relationships within the data.

MAKING RELATIONSHIPS EXPLICIT

In SDTM datasets, relationships are typically implied through keys like USUBJID, VISITNUM, and timing variables. A graph makes these relationships explicit. It turns hidden connections into visible pathways, allowing reviewers to see how patients, events, and outcomes relate at a glance.

- **Nodes** represent meaningful entities such as patients, visits, lab tests, and adverse events.
- **Edges** capture the relationships: a patient "experienced" an event, a lab test "occurred at" a visit, or a treatment "was administered at" a certain time.

This explicit mapping mirrors how scientists and reviewers ask questions, following a path of relationships rather than manually joining tables.

DEFINING A FLEXIBLE GRAPH SCHEMA

Unlike rigid relational schemas, a graph schema can evolve as the study progresses or as new types of data are integrated (e.g., biomarker results or digital health data).

- Additional nodes and relationships can be added without re-engineering the whole structure.
- Metadata (e.g., severity grades or causality assessments) can be attached directly to relationships, preserving clinical meaning.

For example, an edge between a patient and an adverse event can store both the severity and the timing of the event.

BUILDING FOR EXPLORATION

The transformation into a graph is designed for exploration rather than predefined outputs. Full lineage to the original SDTM datasets is maintained for traceability. The goal is to empower interactive exploration, where reviewers can dynamically follow relationships, pivot between connected concepts, and ask new questions without waiting for static listings or custom programming.

PRACTICLE CHALLENGES

Transitioning SDTM data into a graph structure presents practical challenges. Generic table designs must be mapped carefully to preserve meaning, and consistency across visits, particularly when planned and actual visits differ - needs attention. Finally, deciding whether to anchor by visit or by date affects how patterns are interpreted.

BENEFITS OF GRAPH-BASED INTERFACE

A connected, graph-based interface transforms how teams interact with clinical data.

CROSS-DOMAIN CONTEXT

Complex clinical narratives involve multiple data sources - demographics, dosing, labs, AEs, concomitant meds. In a graph:

- These data points are interconnected rather than siloed.
- Reviewers can visually follow a patient journey, understand what happened and when, and identify patterns across domains.

For instance, a reviewer can trace an SAE back to the corresponding dosing visit and view concurrent lab values and concomitant medications – all without navigating multiple listings or tables.

FASTER AND MORE INCLUSIVE REVIEWS

Because the graph interface is visual and interactive, it reduces reliance on programming teams for ad hoc outputs.

- Clinicians, safety teams, and operations staff can explore directly.
- Queries like *"Show all subjects with liver enzyme elevations within 14 days of dosing"* become point-and-click queries rather than new SAS programs.

This democratizes access to insights and helps stakeholders collaborate in real time.

SUPPORTING PATIENT SAFETY PROACTIVELY

Early detection of safety signals requires seeing relationships. A connected view highlights outliers and correlations that might otherwise go unnoticed until later stages.

- A pattern of related AEs across multiple patients can be visualized and flagged earlier.
- Integrated context (timing, treatment, and lab values) helps determine causality faster.

By visualizing connections, reviewers can move from static reviews to proactive monitoring - seeing emerging patterns earlier.

MULTI-PURPOSE REUSE

A key advantage is reuse beyond single studies. Once structured as a graph:

- **Cross-trial analysis:** Compare safety outcomes across protocols without re-engineering datasets.
- **Knowledge building:** Build institutional memory of relationships between treatments and outcomes.
- **Development support:** Inform study design, patient inclusion criteria, and risk mitigation.

This approach does not replace SDTM. Instead, it builds on it - preserving compliance while enabling real-time, connected exploration.

USE CASES BEYOND A SINGLE TRIAL

While the initial focus is trial-level SDTM data, the same model extends naturally:

- **Cross-Trial Safety Reviews:** Aggregate AEs across multiple studies in a portfolio. Identify patterns such as recurring signals with similar compounds.
- **Pharmacovigilance Integration:** Link post-marketing safety reports to trial data and literature for continuous monitoring.
- **Protocol Optimization:** Use connected insights to refine inclusion/exclusion criteria and avoid known safety pitfalls in future studies.

Each of these use cases depends on the same principle: the graph provides a holistic, connected view that table listings cannot. When data is connected, every study contributes to the next one.

CONCLUSION

Clinical trial data is rich but often underutilized because of structural silos. SDTM fulfills essential regulatory and standardization needs, yet its tabular design can limit how easily relationships are seen and explored. When SDTM's structured foundation is complemented by a connected graph, entirely new ways of interpreting and understanding data emerge.

A graph-based perspective enables stakeholders to explore dynamically, uncover insights earlier, and make decisions that could enhance patient safety and trial outcomes. It provides a framework not only for faster exploration but also

for institutional learning, creating a connected knowledge base that supports reuse across studies, programs, and therapeutic areas.

This evolution represents more than a technical advancement; it is a cultural shift toward curiosity-driven exploration and collaborative problem-solving. SDTM provides the structure, and graphs bring the connection. Together, they transform static data into living insights, enabling deeper understanding, stronger collaboration, and more informed decision-making.

REFERENCES

1. CDISC Study Data Tabulation Model (SDTM) Implementation Guide.
2. PHUSE test data, CDISCPILOT https://github.com/phuse-org/TestDataFactory/tree/main/Updated
3. New Approach to Graph Databases, A. Berg, H. Drews and C.Dahlbo, PP05 PHUSE EU Connect 2018
4. Graph vs Table – A Complex Question, C.Dahlbo, E.Kelty, PP05 PHUSE EU Connect 2023
5. Unique Technology for the Future, C.Dahlbo, A.Workneh, H.Drews, PP04 PHUSE EU Connect 2018

ACKNOWLEDGMENTS

Thanks to the data managers, clinicians, and safety reviewers whose commitment to improving data accessibility and insight generation inspired the development of this approach.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Author Name: Eva Kelty

Company: Capish

Address: Lokgatan 8, Malmö, Sweden

Work Phone: +46 40 10 88 80

Email: eva.kelty@capish.com

Website: www.capish.com

Author Name: Catharina Dahlbo

Company: Capish

Address: Lokgatan 8, Malmö, Sweden

Work Phone: +46 40 10 88 80

Email: catharina.dahlbo@capish.com

Website: www.capish.com