Paper ET18

Rethinking Data Quality - From Fragmented Checks to Connected Insights

Eva Kelty, Capish, Malmö, Sweden
Matthew Kramer, Capish, Boston, USA

ABSTRACT

Data quality has always been central to clinical research, but today it carries new urgency. With decentralized trials, real-world inputs, and multiple sources feeding complex systems, ensuring accuracy is no longer a back-end task, it shapes every decision that follows. When inconsistencies go unnoticed or context is missing, the impact ripples across timelines, compliance, and ultimately, patient outcomes.

Meeting these challenges calls for non-traditional approaches. Instead of relying on fragmented checks and static reports, teams benefit from connected data where relationships are visible and issues surface through exploration rather than escalation. Leveraging innovative technologies such as graph databases and interactive data visualization makes it possible to uncover context, patterns, and dependencies often hidden in traditional structures.

This session will highlight emerging practices that enhance operational efficiency and support compliance, moving data quality beyond cleanup - toward an insight-driven process enabling confident decisions and continuous improvement in the life sciences sector.

INTRODUCTION

Data quality is entering a new era, one defined by opportunity rather than limitation. As clinical research continues to evolve through decentralized models, digital technologies, and broader data partnerships, organizations now have access to richer and more timely information than ever before. This expanding data landscape presents an important shift: data quality is no longer just about preventing errors; it is about enabling insight, strengthening decision-making, and improving confidence in study outcomes.

In practice, however, many teams continue to experience recurring challenges. The same data points may be queried repeatedly across review cycles, not because the data is incorrect, but because its context is unclear. Values appear out of sequence, relationships between events are difficult to interpret, or timing is abstracted in ways that obscure how data elements relate. These situations increase review effort and slow interpretation, even when data meets traditional quality criteria.

Historically, data quality has focused on foundational principles such as accuracy, completeness, and consistency. These remain essential. At the same time, modern trials highlight an additional, increasingly important dimension: how data is aligned, connected, and interpreted over time. The way timing is represented influences how relationships are understood, how patterns are recognized, and how confidently conclusions are drawn.

Clinical data is commonly reviewed using visit-based structures. These abstractions provide a standardized framework that supports aggregation, comparison, and reporting across participants. While invaluable, visit-based summaries can also simplify timing in ways that reduce visibility into true sequence - particularly when data is collected asynchronously, continuously, or within broad visit windows. As a result, data that is valid and complete may still require repeated clarification to support confident interpretation.

This paper builds on connected approaches to data quality by demonstrating how preserving temporal context, alongside structural consistency, enhances understanding. By enabling movement between visit-based and day-based perspectives, connected insight allows teams to maintain familiar workflows while gaining clearer visibility into timing, relationships, and study conduct, reducing repetitive review and supporting more informed decisions.

THE LIMITATIONS OF FRAGMENTED DATA QUALITY

Traditional data quality processes often rely on isolated checks performed across separate datasets, tools, or static summaries. While these approaches have long supported operational consistency and regulatory expectations, their fragmented nature can limit broader understanding as trial complexity increases. Many traditional tools are designed around predefined review structures and fixed summaries, frequently organized by domain or visit. Although this structure supports standardization and reporting, it can limit visibility into how data points relate across domains and over time. When review begins with static summaries rather than underlying relationships, additional context often must be reconstructed through repeated queries or manual investigation.

WHEN CONTEXT IS MISSING

Data points that are missing, delayed, or flagged repeatedly may appear as isolated discrepancies when reviewed independently. Without broader operational or temporal context, underlying contributors, such as visit scheduling challenges or workflow delays, can remain obscured, leading to repeated reconciliation rather than resolution.

WHEN PATTERNS ARE HARD TO SEE

Protocol deviations or recurring inconsistencies may be reviewed individually, even when they reflect systematic patterns related to study design or site processes. Without connected insight, recognizing these trends can require multiple review cycles and extended investigation.

WHEN TIMING IS ABSTRACTED

Visit-based summaries group data under protocol-defined milestones, offering efficiency and consistency. However, measurements associated with the same visit may have been collected days or weeks apart, particularly when visit windows are broad or data is captured asynchronously. While this abstraction supports reporting, it can obscure important differences in timing and sequence that influence interpretation.

Across these scenarios, the challenge is consistent: data may be accurate and complete, yet difficult to interpret when relationships and timing are not easily visible.

CONNECTED DATA QUALITY: A MODERN APPROACH

Connected data quality emphasizes relationships across data elements, supports exploratory review, and preserves visibility into timing and sequence. Rather than treating data points as independent entries to be validated in isolation, this approach considers how information connects across subjects, domains, workflows, and study events to form a more complete picture of study conduct.

Modern data architectures support this approach by explicitly representing these relationships. Graph-based models, for example, are well suited to capturing how subjects, visits, assessments, dosing, operational processes, and events are interconnected, providing a flexible structure for preserving context as data evolves. When these connections are maintained, reviewers can more easily recognize how issues cluster, how patterns develop, and how dependencies across domains influence data behavior over time.

Exploratory review is central to this model. Reviewers can move fluidly between summarized information and underlying detail, examining related records, timelines, and sequence as questions arise – allowing context to surface naturally.

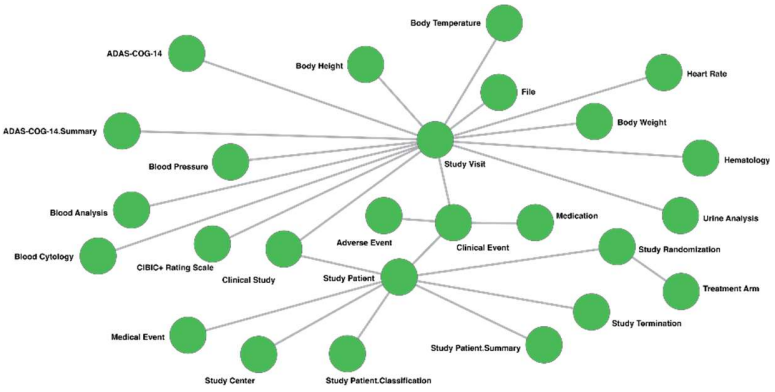


Figure 1. In a graph-based model, the domain data are interconnected rather than siloed. Normally hidden dependencies become visible paths, enabling reviewers to immediately understand how subjects, events, and outcomes are connected.

THE ROLE OF TIME IN DATA INTERPRETATION

As trials increasingly rely on decentralized and continuous data collection, understanding how events relate to one another becomes an important factor in interpretation across safety, efficacy, and operational domains. When sequence and proximity are unclear, even accurate and complete data can require additional effort to interpret.

In practice, questions that span domains often hinge on when events occurred relative to one another. For example, determining whether a laboratory change preceded or followed a dose adjustment, or how closely an adverse event aligned with a related assessment, may require manual reconstruction when dates are abstracted or summarized.

Clinical data is commonly reviewed using visit-based frameworks, which provide structure and consistency for aggregation and reporting. However, data points associated with the same visit may reflect different points in the study timeline, particularly when visit windows are broad or data is captured asynchronously. This abstraction can make it more difficult to assess how related events align without additional context.

DAYS VS. VISITS: COMPLEMENTARY VIEWS OF THE SAME DATA

Visit-Based Perspective

Visit-based views organize data around protocol-defined milestones, supporting aggregation, comparison, and reporting across participants. This structure enables teams to answer questions such as:

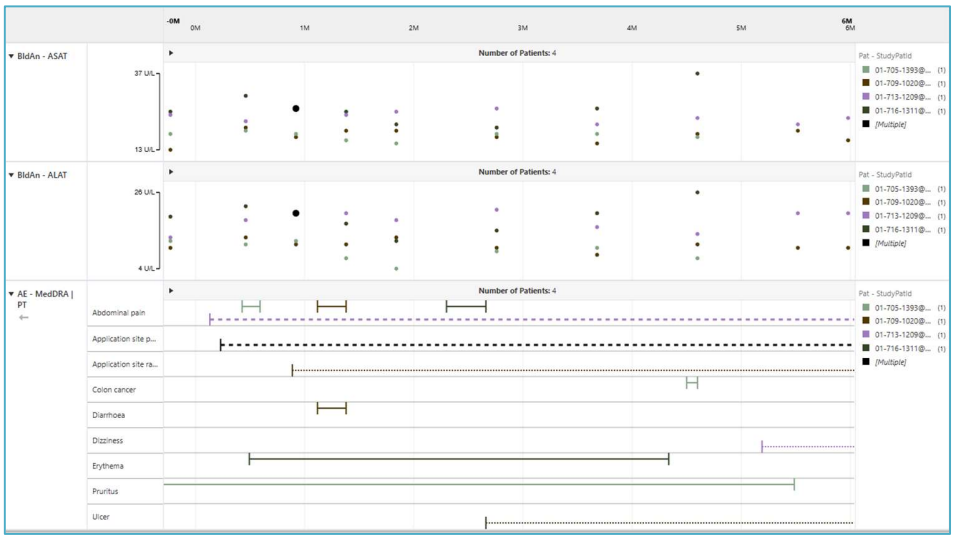
- What do measurements look like at Visit 8?
- How do participants compare at predefined study points?

Chronological (Day-Based) Perspective

A day-based view places each data point on its actual collection date, revealing:

- Natural variability in data collection timing
- Proximity between related events
- True chronological sequence across domains

This perspective supports deeper questions, such as how closely a lab value aligns with a dosing change or whether an event occurred before or after a safety signal.



Visits

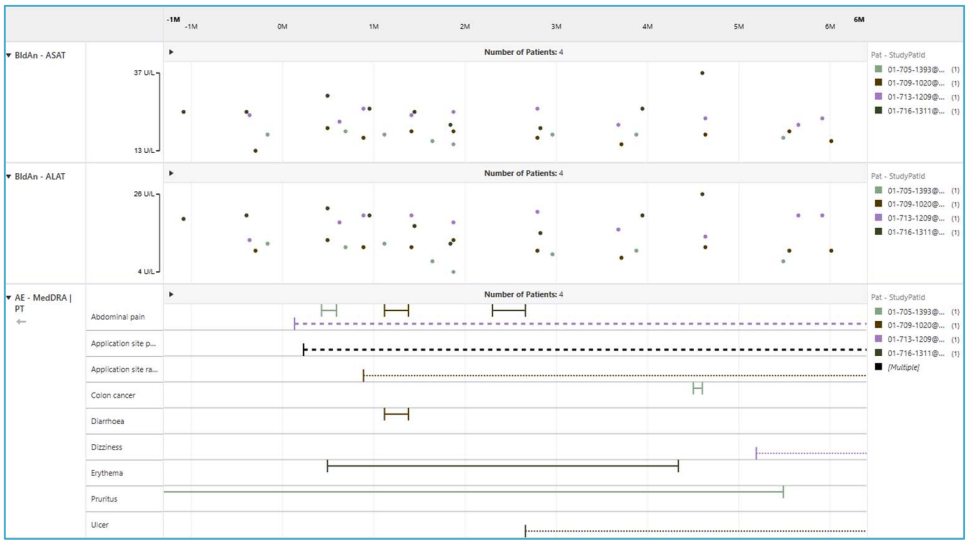


Figure 2. Visit-based perspective versus chronological (day-based) perspective.

Strength Through Combination

Data quality is strongest when teams can move easily between these perspectives. Visit-based structure provides consistency, while chronological context enhances interpretation and confidence.

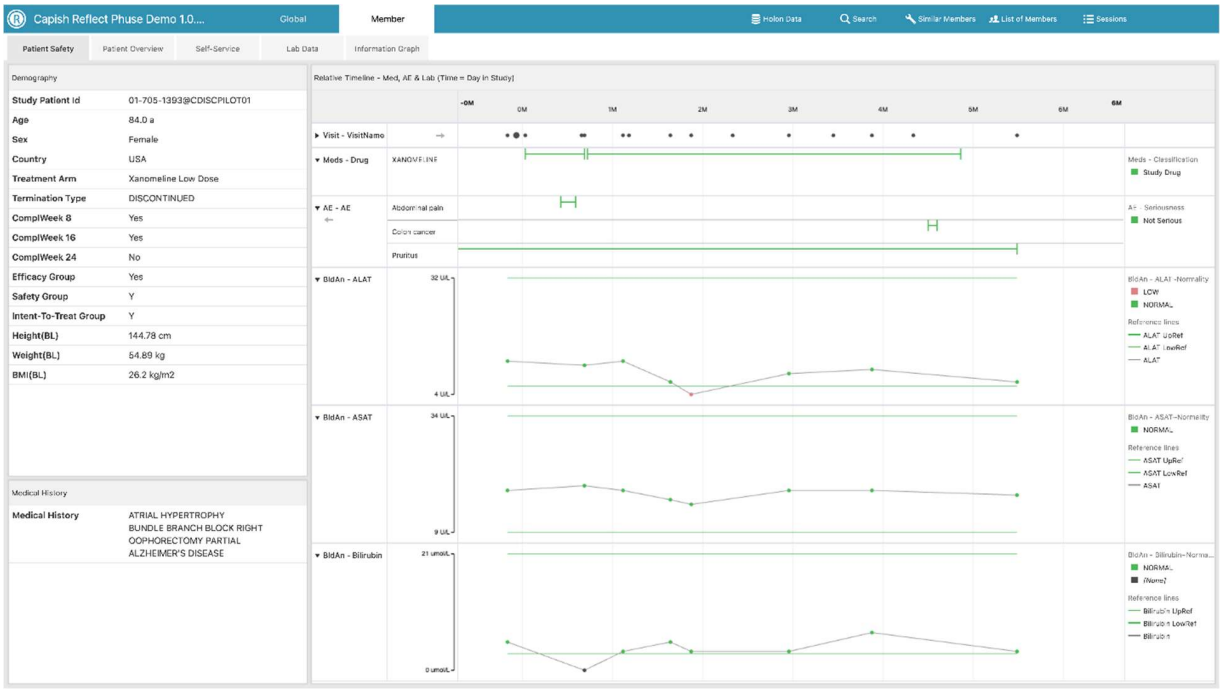


Figure 3. An example of a timeline for one patient where lab values are evaluated together with adverse events and medications.

WHY TEMPORAL FIDELITY STRENGTHENS DATA QUALITY

Making timing and sequence visible enhances multiple dimensions of data quality:

- More meaningful comparisons that acknowledge actual collection timing

- Greater confidence in summarized outputs by understanding what they represent
- Clearer safety interpretation where sequence and proximity matter
- Stronger downstream decisions supported by accurate context

When data relationships and timing are transparent, teams can interpret findings with greater clarity, even when the data itself is technically valid and complete.

BRIDGING VISITS AND DAYS THROUGH CONNECTED INSIGHT

Connected data quality does not position visit-based and day-based views as alternatives. Instead, it brings them together. High-quality review environments support:

- Visit-based summaries for structured review and reporting
- Immediate access to actual dates when deeper context is needed
- Visibility into timing differences that may influence interpretation

This dual perspective reinforces data quality as a forward-looking, insight-driven discipline, supporting both operational efficiency and scientific rigor.

A SHIFT TOWARD CONTINUOUS IMPROVEMENT

With connected insight and modern review tools, data quality becomes a continuous learning process rather than a series of isolated fixes. Teams can evaluate trends over time, identify workflow improvements, and apply lessons learned across studies.

Machine-assisted review and visual exploration further guide attention toward areas that merit deeper examination, allowing teams to focus on insight rather than manual reconciliation. This emphasis on ongoing improvement supports smoother trial execution, stronger compliance, and more confident decision-making.

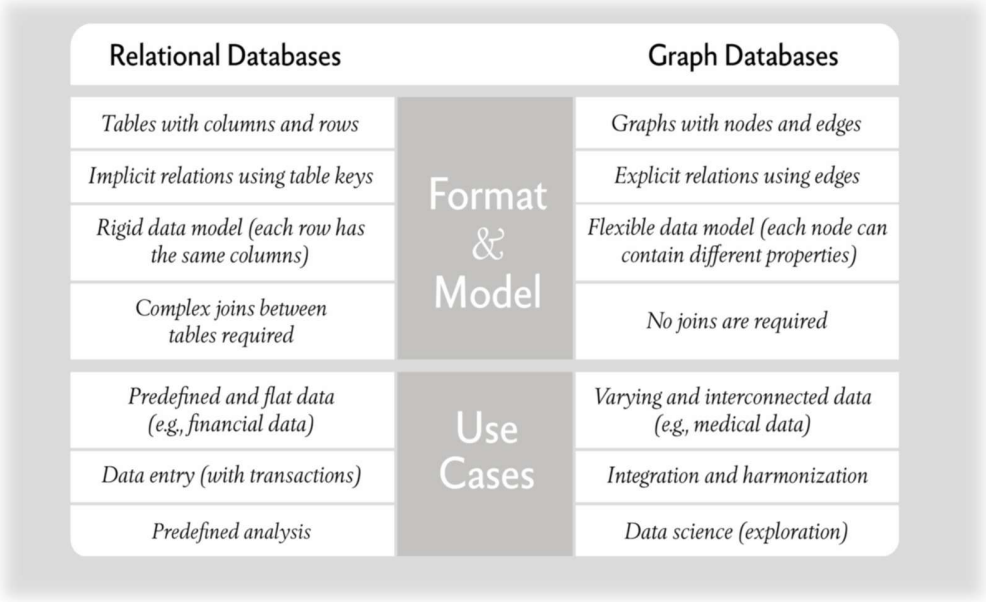


Figure 4. A comparison of a traditional relational database structure and a graph database structure.

CONCLUSION

As clinical trials become increasingly complex and data-rich, organizations have an opportunity to rethink data quality in ways that enhance clarity, efficiency, and confidence. By moving beyond fragmented checks and embracing connected insight, including temporal fidelity, teams can better understand not only what issues occur, but how they relate and why they persist.

Visit-based abstractions remain essential, but when complemented by true chronological context, they become even more powerful. Connected data quality transforms validation into insight, supporting clearer interpretation, stronger safety assessment, and continuous improvement throughout the clinical research lifecycle.

REFERENCES

1. From SDTM to a Graph-Based, Multi-Purpose Interface for Data, E. Kelty, C. Dahlbo, TT06 PHUSE EU Connect, 2025, 2025 PAP_TT06

2. The Table Has Turned, C. Dahlbo, A. Berg, PP18 PHUSE EU Connect, 2022, POS_PP18.pdf

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:	Matthew Kramer
Company:	Capish
Address:	100 Montgomery Street, San Francisco, CA, USA
Work Phone:	+6179972550
Email:	matthew.kramer@capish.com
Website:	www.capish.com