Paper PP09

The Utilization of an Interface Based on a Graph Database

Catharina Dahlbo, Capish, Malmö, Sweden
Eva Kelty, Capish, Malmö, Sweden

ABSTRACT

A graph database-powered interface provides dynamic, connected views of clinical trial data, tailored to specific roles and trial stages. It gives users direct access to a rich data environment, removing the need for constant report or graph requests. By enabling independent data exploration, it empowers each role throughout the trial lifecycle. During the planning, design teams explore historical data to identify patterns and optimize study design. During monitoring, site coordinators and data managers track enrolment, deviations, and adverse events. Investigators and safety managers gain contextual insights through interactive timelines and safety signal queries. In the analysis stage, medical writers ensure accurate reporting, streamlined preparation for regulatory submission and final study filing. This platform transforms complex data into actionable insights, revealing hidden relationships and supporting informed, confident decisions.

The poster demonstrates how a graph-enabled interface can serve as a responsive partner across the entire trial, fulfilling different purposes and requirements.

INTRODUCTION

The life sciences and healthcare sectors continue to face significant challenges in managing the complexity of their data-intensive processes. Ensuring that data is easily accessible, comprehensible, and valuable for a wide range of end-users remains a constant difficulty. The data involved in these fields often comes from diverse sources, includes numerous variables, and requires integration to provide a broad and meaningful overview for stakeholders.

Data is no longer seen as something stored away and only used by specific departments. Increasingly, it's being recognized as a valuable tool that helps people make better decisions, generates new ideas, and leads to better medical treatments. This change means that data needs to be stored in ways that make it easy to access, understand, and evaluate, so it stays useful and relevant for current research and development.

Clinical trials, which form the backbone of new drug development, exemplify these challenges. They involve multiple stages, from planning and execution to reporting and archiving, and engage a variety of stakeholders including researchers, sponsors, data managers, statisticians, medical writers and so on. Each group approaches the data from different perspectives, asking diverse questions and seeking distinct insights.

It requires a scalable, interoperable system that allows seamless querying and delivers fully annotated and context-rich data. Interoperability, achieved through standardization and the adoption of common terminologies, is critical for harmonizing and integrating heterogeneous data sources. Moreover, preserving the original context in which data was collected ensures that its meaning and purpose remain intact, enabling users to explore, analyze, and draw informed conclusions.

The FAIR (Findable, Accessible, Interoperable, Reusable) principles provide a strong foundation for building such data repositories. In this context, graph databases offer unique advantages for modeling, storing, and navigating the intricate and often unstructured data characteristics of clinical trials. Their inherent flexibility supports data exploration throughout the entire clinical trial lifecycle, from planning to archiving.

The poster and the paper explore the utilization of an interface powered by a graph database to enhance data accessibility and usability in clinical trials. By integrating such an interface, users are empowered to interact with complex datasets intuitively, unlocking the full potential of clinical trial data and promoting data reusability across all stages of the research process.

DEMONSTRATING THE CONCEPT WITH CURATED PHUSE DATA

To illustrate how a graph database with an intuitive interface can support the user we have used publicly available PHUSE test data [1]. The data consists of CDISC-compliant datasets from which we have used the SDTM dataset to work with, CDISCPILOT. All major domains are represented, and the data has been carefully curated to improve usability and exploration.

Through this curation, relationships between domains have been standardized, variable names harmonized, and inconsistencies reduced. This ensures that the users can focus on analyzing connections and identifying insights rather than spending time cleaning or preparing data. The curated structure makes it easier to navigate across datasets, visualize links between entities, and quickly understand the context of each data point, ultimately supporting more efficient and insightful analysis.

REVEALING HIDDEN LINKS IN CLINICAL TRIAL DATA

Graph databases offer a fundamentally different way to structure and explore clinical trial data by modeling information as interconnected entities [2]. Instead of organizing data into flat tables, they represent subjects, visits, interventions, adverse events, and other trial components as nodes, while the relationships between them, such as temporal links, causality, or hierarchical dependencies, are represented as edges.

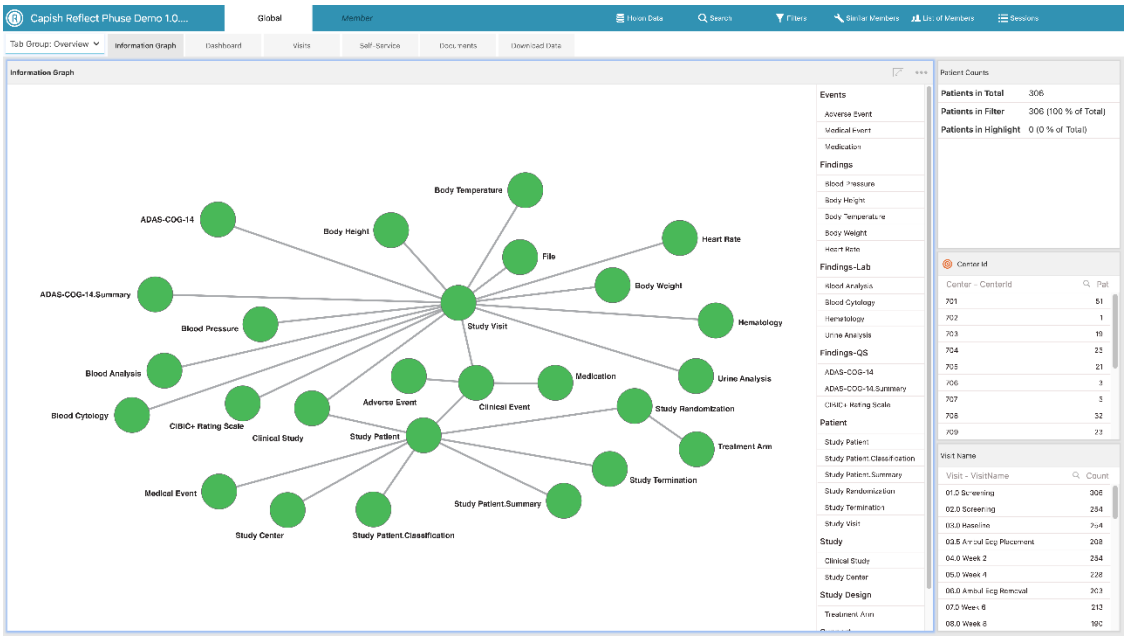
A key distinction is that graph databases store relationships as first-class objects. This means that the connections between entities, such as a subject having a visit or a treatment leading to an adverse event, are stored in the database rather than being inferred through joins or key matching. These relationships can also carry their own attributes (for example start and end dates, or source), making them analyzable. As a result, the database can query these connections directly and efficiently, preserving both structure and context.

This structure naturally reflects the inherent complexity of clinical trials. For example, subjects participate in studies involving multiple visits where assessments are performed, interventions may lead to adverse events linked to both treatments and visits, and longitudinal data such as medications records create additional layers of temporal relationships critical for safety and outcome evaluation [3].

Unlike traditional relational databases that depend on rigid schemas and often complex joins to reconstruct relationships, graph databases enable real-time traversal of data, supporting dynamic and flexible exploration. For clinical trials, this model preserves the original context of the data and reduces the risk of meaning being lost during transformation. Users can intuitively navigate data across domains, exploring links between subjects, treatments, outcomes, and timelines without needing custom reports for every new question. Different roles, such as data managers, clinicians, or medical writers can be supported through customized queries and visualizations tailored to their needs, while new data sources can be integrated without altering existing schemas.

Most importantly, this relationship-driven approach aligns with how clinical experts naturally conceptualize trial data, not as isolated points, but as connected events evolving over time. Depending on the customers' needs, the data could be modeled differently.

The figure below shows via an information graph how the study is modeled.



THE VALUE OF AN INTERFACE BUILT FOR GRAPH DATA

A graph database alone is not sufficient, without an intuitive interface, the data remains inaccessible to many users. The real value emerges when the interface is designed to harness the strengths of the graph model, making relationships and connections between data points clear and actionable.

By extending a property graph database with an information-based ontology and multiple indices [4], users can easily define complex cohorts, a process that would otherwise be time-consuming and technically demanding. This combination transforms how users interact with their data, enabling powerful, flexible, and user-friendly exploration of complex relationships.

ADDRESSING INCREASING COMPLEXITY AND DIVERSE USER NEEDS

Clinical trial data is no longer confined to static, single-source datasets. Modern studies increasingly integrate information from a growing array of sources such as traditional CRF data captured via EDC systems, real-world evidence drawn from electronic health records and registries, biomarker and genomic profiles, imaging data, wearable device data, and decentralized trial platforms.
While this diversity adds significant value to clinical research and enables a more holistic understanding of patient outcomes, it also introduces new challenges in data integration, harmonization, and contextual interpretation.

Compounding this challenge is the fact that different stakeholders across the trial lifecycle approach data with distinct needs and perspectives.

- **Study designers** seek historical insights to refine protocols and optimize trial design.
- **Site monitors and data managers** require real-time visibility into data quality and protocol adherence.
- **Investigators and safety teams** focus on rapid detection and contextualization of emerging safety signals.
- **Statisticians and medical writers** depend on transparent data paths to ensure accurate reporting and submission readiness.
- **Regulatory reviewers** expect clear documentation of data lineage and traceability.

Each role generates specific analytical questions often involving complex filtering, aggregation, and longitudinal analysis. Traditional static data tables, with their rigid structure, struggle to deliver these tailored insights efficiently. A graph-based, curated, and interface-supported approach provides a scalable and flexible solution, empowering all user types to explore data intuitively, comprehensively, and in real time.

USE CASES ACROSS THE TRIAL LIFECYCLE

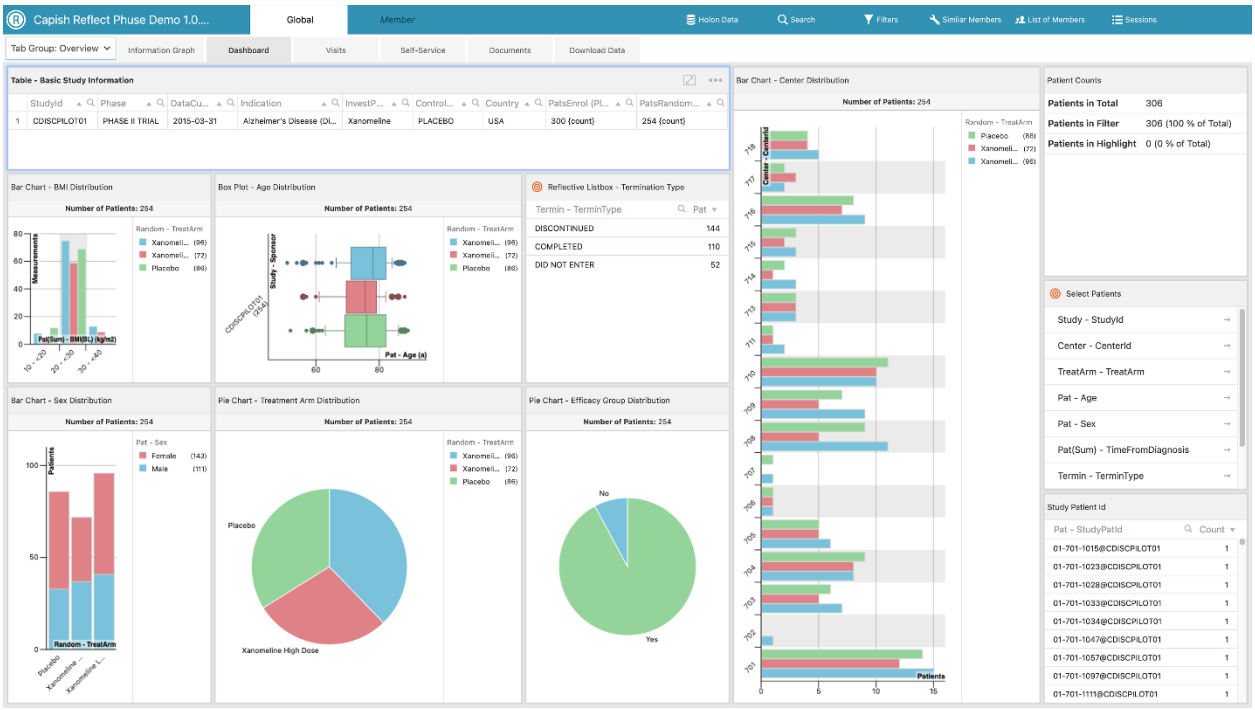
Study Planning and Design

During the study planning phase, teams need to draw on historical data from prior studies or external sources to make informed design decisions. These include defining inclusion and exclusion criteria, determining endpoint definitions, and estimating sample size.

Using a graph-enabled interface, teams can explore historical data interactively to uncover meaningful patterns very quickly, such as:

- The timing and frequency of adverse events across population
- Common patterns of protocol deviations.
- Correlations between baseline characteristics, treatments, and outcomes.

By presenting these insights visually and interactively, the interface helps teams design studies that are both scientifically rigorous and operationally feasible.



In the figure above we can see the PHUSE dataset, curated and standardized to ensure consistency across domains. The curation simplifies exploration, ensuring that users can focus on insight generation rather than data cleaning. The dashboard is only one of many interactive tabs where data immediately can be filtered, highlighted, and grouped for further exploration.

The interface can be tailored to specific study needs and extended with self-service features, enabling users to define custom queries, visualize data relationships, and download for decision reports.

The dashboard in the example provides an **interactive overview of patient-level clinical trial data**, allowing users to explore and compare key study variables across multiple dimensions.

It displays:

- **Demographics and baseline characteristics,** such as age, sex, and treatment group distributions.
- **Adverse event summaries** by category, severity, and relationship to treatment.
- **Efficacy and safety outcomes** represented in bar, box**,** and pie charts for easy comparison.
- **Dynamic filtering options** across multiple domains (e.g., treatment arm, visit, country, adverse event type), enabling users to refine patient cohorts interactively.
- **Real-time cohort statistics** (total, filtered, and highlighted patients) and contextual data summaries in the side panels.

Overall, the dashboard offers a comprehensive and customizable visualization environment for analyzing study populations, patterns, and outcomes in an integrated, user-friendly way.

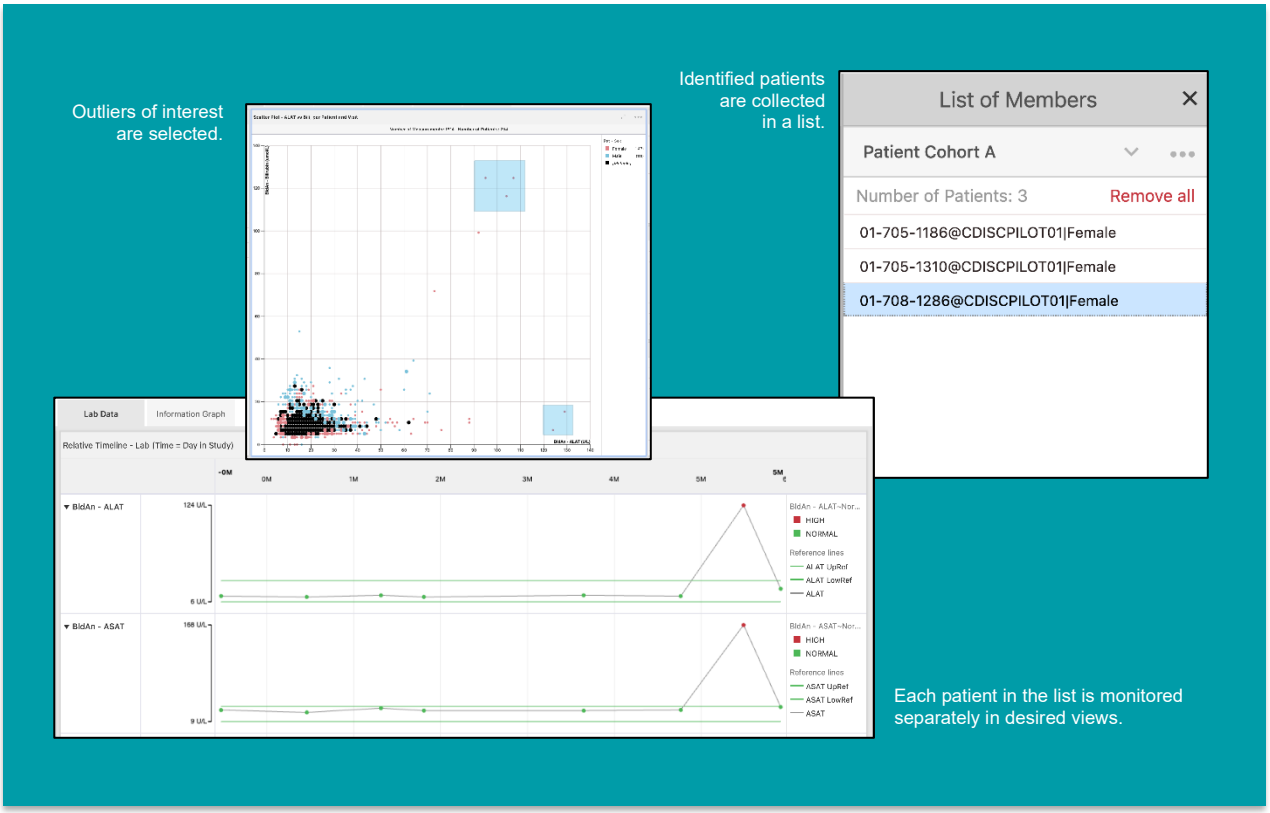
Data Monitoring

As trials progress into the conduct and monitoring phase, real-time visibility becomes critical. Site monitors, data managers, and study leaders need to track

- Enrollment rates and subject demographics
- Data completeness and quality

- Protocol deviations or emerging trends
- Clusters of adverse events by site, country, or population

A graph-based interface supports this by allowing users to interactively filter and visualize relationships between variables.



The figure above illustrates how users can identify and monitor specific patients of interest within the interface.

- **Outliers** in the data visualization are selected directly from the scatterplot.
- **The identified patients** can then be collected into a list or cohort for focused analysis.
- Each patient in the list can be **monitored individually** across different visualizations, such as lab timelines (e.g., ALAT and ASAT levels), to understand their data patterns and clinical progression.

This demonstrates how the interface enables targeted exploration, allowing users to isolate unusual cases and follow them across multiple data views for deeper insight. This enables faster issue detection and context-rich understanding of data anomalies without complex programming or static reports.

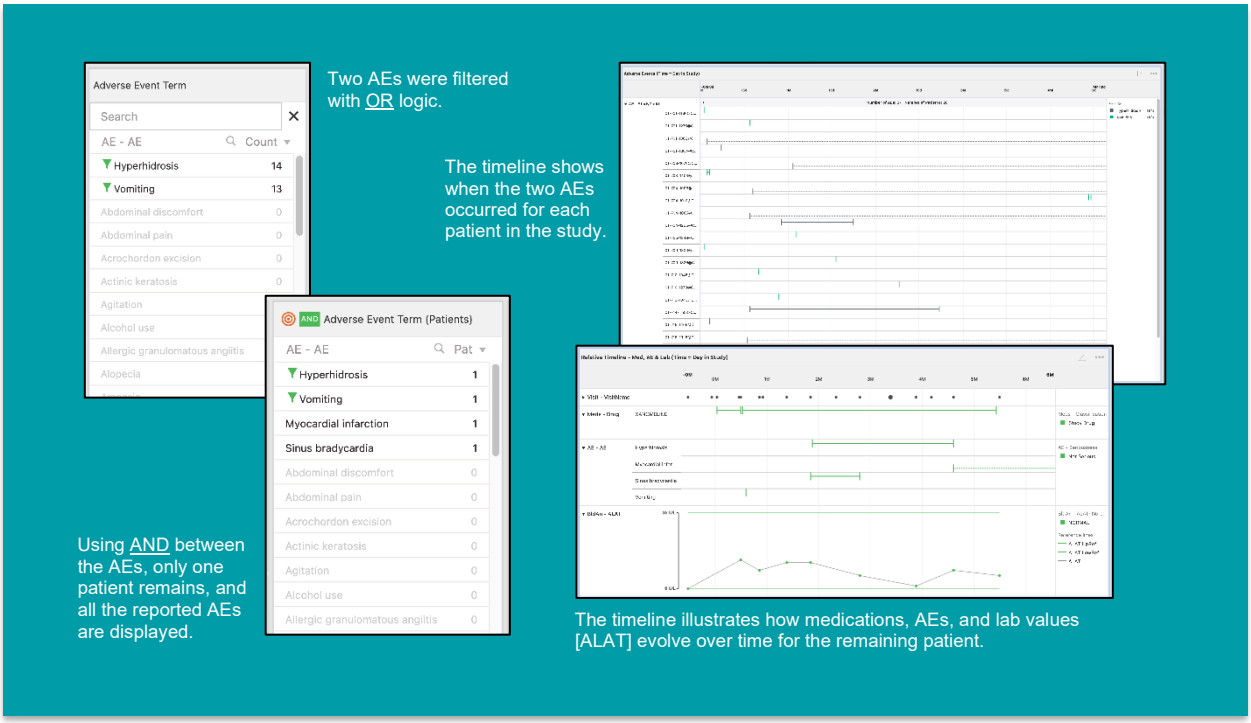
Safety Assessment

In safety control, investigators and pharmacovigilance teams need to identify and assess potential safety signals quickly.

A graph structure enables temporal and relational analysis, linking adverse events to treatments, visit timelines, and patient characteristics. This allows users to:

- Visualize event sequences and causal links
- Compare safety patterns across populations or sites

- Detect merging trends



The example in the figure above demonstrates how users can interactively filter and explore adverse events (AEs) within the interface.

- When two AEs (e.g., *Hyperhidrosis* and *Vomiting*) are selected with **OR** logic, the system displays all patients who experienced either of the two events.
- When the **AND** condition is applied, only the patient who experienced both AEs remains visible, and all other reported AEs for that patients are displayed as well.
- The **timeline visualizations** shows when these AEs occurred for each patient, and for the remaining patient, it also displays **medications, adverse events, and lab values (ALAT)** in relation to one another over time.

Interactive timelines and event network visualizations help translate complex data relationships into intuitive actionable insights, supporting faster and more accurate safety assessment. This approach helps you better to know your data and be ready for the regulatory authorities.

Study Reporting and Regulatory Submission

In the reporting and submission phase, traceability and transparency are essential. Graph-based systems inherently preserve data lineage, allowing every derived value or analytical result to be traced back to its original source, along with its associated metadata, processing steps, and dependencies.

This end-to-end transparency:

- Builds confidence in reported results
- Simplifies responses to regulatory queries
- Minimizes inconsistencies during submission preparation

By maintaining context and provenance within a single connected data structure, graph databases help ensure both regulatory compliance and scientific reproducibility.

The figure shows how the interface can be used to identify and analyze patients with serious adverse events (SAEs).

- The data is **filtered for patients with SAEs**, and the three affected patients are saved as a subgroup. Two of them also experienced non-serious AEs.
- The system then displays **all adverse events** for these selected patients, with an option to further filter for events occurring on the same date.
- The **highlighted subgroup** is visually marked in yellow throughout the application, allowing users to easily track and compare the subgroup across different views.

Together, these capabilities streamline the final stages of the trial, ensuring that results are transparent, traceable, and ready for efficient regulatory review, while preserving the full scientific context for future data reuse.

CONCLUSION

Integrating a graph database with an intuitive, interface represents a major step forward in how clinical trial data is accessed and understood. By combining the flexibility of a graph model with the reliability of standardized CDISC structures such as SDTM and ADaM, sponsors and CROs can gain real-time, contextual insights throughout the trial lifecycle.

The ability to customize views for different roles, such as clinical operations, safety, or biostatistics, enhances collaboration and ensures that each stakeholder can interact with the data in a way that supports their specific objectives. This approach allows users to explore data directly, without relying on static reports or complex

programming. It accelerates insight generation, ensures transparency into data lineage, and strengthens both regulatory confidence and cross-functional collaboration.

Building on existing standards rather than replacing them, the graph-based framework enhances flexibility, usability, and compliance. It transforms data from a passive record into an active resource, driving continuous learning, data reuse, and more agile, informed decision-making across studies.

REFERENCES

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

ACKNOWLEDGMENTS

We would like to thank Patric Moreau for his dedication to the poster design.

CONTACT INFORMATION

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

Author Name: Catharina Dahlbo
Company: Capish
Address: Lokgatan 8, 211 20 Malmö, Sweden
Work Phone: +46 40 10 88 80
Email: catharina.dahlbo@capish.com

Author Name: Eva Kelty
Company: Capish
Address: Lokgatan 8, 211 20 Malmö, Sweden
Work Phone: +46 40 10 88 80
Email: eva.kelty@capish.com

Website: capish.com