Paper ET12

The Knowledge Graph - The Power of an Information Network driving Biometric workflows

Nekzad Shroff, Saama Inc, Campbell, CA, USA
Jaya Inampudi, Saama Inc, Campbell, CA, USA

ABSTRACT

Clinical trial programming requires meticulous management of code/specification lineage, dependencies, and traceability between components like SDTM, ADaM, TLFs, logs, and outputs.

This paper explores Knowledge Graph databases as a transformative solution for this use case. Blending empirical research and a case study with Saama's Biometrics research and analysis information network, we demonstrate Knowledge Graphs' unparalleled ability to map relationships, lineage, and dependencies among SDTM, ADaM, TLFs, outputs, logs, and outline files as they get copied/mutated during their lifecycle in the system. This visual graph-based granularity provides comprehensive, intuitive navigation, reusability, and metadata management previously unattainable. Knowledge Graphs not only enhance traceability and integrity of code/specification lineage but significantly improve compliance and operational efficiency. In addition, the availability of this digitized relationship metadata in the Knowledge Graph is an important step for future automation and streamlining of manual steps in the submission lifecycle.

INTRODUCTION

In the realm of clinical trial programming, the management of code and specification lineage, dependencies, and traceability between various components such as Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), Tables, Listings, and Figures (TLFs), logs, and outputs is a complex and meticulous task. Traditional methods often involve the use of spreadsheets, flat text files, and XML files, which, while functional, lack the ability to effectively manage traceability and dependencies. This not only poses challenges in terms of operational efficiency but also impacts compliance with regulatory standards.

The advent of Knowledge Graphs presents a transformative solution to these challenges. As a form of semantic graph database, Knowledge Graphs are designed to store information in a way that is easily understandable by both humans and machines. They offer an unparalleled ability to map relationships, lineage, and dependencies among various components involved in clinical trial programming.

This paper aims to explore the potential of Knowledge Graphs as a tool for enhancing the efficiency and integrity of clinical trial programming. Through a blend of empirical research and a case study with our Analysis and Reporting environment, we will demonstrate how Knowledge Graphs can provide a level of visual graph-based granularity that offers comprehensive, intuitive navigation, reusability, and metadata management that was previously unattainable.

The following sections will delve deeper into the power of Knowledge Graphs, presenting a detailed case study, discussing future implications, and finally concluding with the transformative potential of this technology in the field of clinical trial programming.

THE POWER OF KNOWLEDGE GRAPHS

Knowledge graphs are a type of data storage and representation concept which stores information in "nodes" as well as "relationships" between nodes. They are especially good at representing datasets with multi-dimensional relationships, where new relationships can be added without needing to be predefined. This type of structure is highly efficient at representing entities such as programs, datasets and specifications which have multiple interconnected metadata elements (Analyses, domains, results datasets, endpoints etc.) and is also flexible in its representation unlike the rigidity of predefined rows and column-based representations.

Knowledge Graphs represent a paradigm shift in the way we manage and interpret data in clinical trial programming. At their core, Knowledge Graphs are a form of semantic graph database designed to store information in a manner that is easily understandable by both humans and machines. Thus, they offer an unparalleled ability to map and visualize relationships, lineage, and dependencies among various components involved in clinical trial programming.

One of the key strengths of Knowledge Graphs lies in their uniquely intuitive structure of nodes and relationships mentioned above. This enables us humans to easily interpret and understand the content through a visual,

graph-based representation of data and relationships. This comprehensive and intuitive navigation can be used throughout the various components of clinical trial programming, such as SDTM, ADaM, TLFs, outputs, and logs. As these components get copied or mutated during their lifecycle in the system, Knowledge Graphs can effectively track these changes at a granular level which was previously unattainable.

In addition to enhancing navigation, this graph-based approach also significantly improves reusability and metadata management. By clearly mapping the relationships and dependencies between different components, Knowledge Graphs allow for more efficient reuse of code and specifications. This not only enhances operational efficiency but also ensures the integrity of code and specification lineage.

Furthermore, Knowledge Graphs play a crucial role in improving compliance. In the highly regulated field of clinical trial programming, maintaining traceability is of utmost importance. Knowledge Graphs enhance traceability by providing a clear and accurate record of code and specification lineage, thereby ensuring compliance with regulatory standards.

In conclusion, the power of Knowledge Graphs lies in their ability to transform the way we manage and interpret data in clinical trial programming. By providing a visual, graph-based representation of data, they enhance navigation, improve reusability and metadata management, and ensure compliance, thereby significantly improving operational efficiency. The following sections will delve deeper into the application of Knowledge Graphs through a detailed case study.

CASE STUDY: A KNOWLEDGE GRAPH BASED ANALYSIS AND REPORTING ENVIRONMENT

At Saama, we have used knowledge graphs as the foundational metadata store for our Analysis and Reporting Environment. The objective of this case study is to illustrate how this design enhances traceability and integrity of code/specification lineage and improves operational efficiency in a real-world setting.

Saama's Biometrics research and analysis information network is a complex ecosystem involving various components such as Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), Tables, Listings, and Figures (TLFs), logs, and outputs. These components undergo numerous transformations and mutations during their lifecycle in the system. Tracking these changes and maintaining the integrity of code and specification lineage is a challenging task.

The introduction of Knowledge Graphs into our solution has revolutionized the way we manage and interpret data. By mapping relationships, lineage, and dependencies among various components, Knowledge Graphs provide a visual, graph-based representation of data that allows for comprehensive and intuitive navigation.

For instance, consider the process of transforming raw data into SDTM, then to ADaM, and finally into TLFs for regulatory submission. Each transformation involves copying and mutating data, creating new versions of files, and generating logs and outputs. With Knowledge Graphs, we can visually track these transformations, providing a clear and accurate record of code and specification lineage.

This enhanced traceability not only ensures compliance with regulatory standards but also significantly improves operational efficiency. By clearly mapping the relationships and dependencies between different components, Knowledge Graphs allow for more efficient reuse of code and specifications. This reduces redundancy, minimizes errors, and saves valuable time and resources.

Furthermore, the digitized relationship metadata available in the Knowledge Graph is an important step towards future automation and streamlining of manual steps in the submission lifecycle. By automating routine tasks and streamlining workflows, Knowledge Graphs have the potential to further enhance operational efficiency and productivity.

In conclusion, this case study demonstrates the transformative potential of Knowledge Graphs in our Analysis and Reporting environment. By enhancing traceability and integrity of code/specification lineage and improving operational efficiency, Knowledge Graphs represent a significant advancement in the field of clinical trial programming.

BUSINESS VALUE OF THE KNOWLEDGE GRAPH

The digitized relationship metadata available in the Knowledge Graph represents a significant step towards the future of clinical trial programming. This metadata, which provides a detailed record of relationships, lineage, and dependencies among various components, has the potential to automate and streamline manual steps in the submission lifecycle.

For example, think about the process of converting raw data into SDTM, then into ADaM, and finally into TLFs for regulatory submission. Each conversion requires copying and changing data, creating new versions of files, and producing logs and outputs. The Knowledge Graph can automate these conversions with its digitized relationship metadata, reducing the need for human involvement. This improves operational efficiency and lowers the chance of errors.

Additionally, the availability of this digitized relationship metadata opens up new possibilities for the development of advanced tools and applications. For example, machine learning algorithms could leverage this metadata to predict the impact of changes in one component on other related components. This could significantly improve the efficiency of the development and testing process.

Also, Knowledge Graphs could also facilitate the development of more sophisticated data visualization tools. By providing a visual, graph-based representation of data, these tools could offer more intuitive and interactive ways to navigate and interpret data.

In conclusion, the digitized relationship metadata available in the Knowledge Graph has the potential to revolutionize clinical trial programming. By automating and streamlining manual steps in the submission lifecycle, and facilitating the development of advanced tools and applications, Knowledge Graphs represent a promising direction for the future of clinical trial programming.

CONCLUSION

In conclusion, Knowledge Graphs represent a significant advancement in clinical trial programming. By providing a visual, graph-based representation of data, they not only enhance operational efficiency but also ensure the integrity of code and specification lineage. The availability of digitized relationship metadata in the Knowledge Graph is a promising step towards future automation and streamlining of manual steps in the submission lifecycle. As such, Knowledge Graphs hold immense transformative potential for the field of clinical trial programming, promising a future of increased efficiency, compliance, and innovation.

CONTACT INFORMATION

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

Nekzad Shroff
Saama Inc
900 E Hamilton Ave #430
Campbell, CA 95008, USA
Work Phone: +1 888.205.3500
Email: nekzad.shroff@saama.com
Web: www.saama.com

Jaya Inampudi
Saama Inc
900 E Hamilton Ave #430
Campbell, CA 95008, USA
Work Phone: +1 888.205.3500
Email: jaya.inampudi@saama.com
Web: www.saama.com