Paper PP13

Harness the Data's Potential in a Clinical Trial

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

ABSTRACT

Clinical trials and medical data often have many dimensions, as they seek to represent and explain the complexities of biological processes. Using a graph database for handling complex and unstructured data, like clinical trials, can get flexibility across the entire lifecycle from planning to archiving.
A repository containing many trials, built, and modelled for searching in data, can be used in different phases of a clinical trial such as planning, execution, reporting and archiving. Accessing data independent of the trial's outcome could be of interest for many purposes like planning new studies or preparing for discussions with authorities, where comprehensive understanding and explanation of results are vital.

The poster will explore the significance of rapid and streamlined data access at various stages of the clinical trial process. Additionally, it will emphasize the importance of integrating metadata within the same system, ensuring a holistic and efficient approach to clinical trial management.

INTRODUCTION

The life sciences and healthcare sectors face persistent challenges in simplifying complex and time-consuming processes to ensure that data is accessible, comprehensible, and valuable for end-users. Frequently, the data involves a multitude of variables from diverse sources, making it difficult for end-users to fully understand and obtain a comprehensive overview of all available information [1].

Data serves as a valuable resource, empowering informed decisions, accelerating innovation, and facilitating the development of improved drugs for the market. Recognizing this, there's a shift in how we treat data. Previously, it was stored solely for archival purposes within specific departments, leading to inaccessible data silos. Today, data should be stored in a manner that allows easy access for those who need it, and it should be assessable to determine its relevance and usefulness for the current objectives [2].

Clinical trials are essential for new drug development, involving distinct stages from planning to archiving. In order to support ongoing medical progress, ensuring post-report data availability is crucial. Clinical trials are collaborative efforts involving a lot of project members e.g., researchers, sponsors, data managers, data statisticians, medical writers etc. Everyone involved has different perspectives on the data and wants to ask different questions. What should a repository look like to be helpful in all stages of a clinical trial? And what help can the users get from an available data repository? The answer to these questions will be discussed in this paper.

To fulfill that a repository can help project members to easily look at and compare data from different parts within a trial it requires a scalable system that facilitates easy querying and provides fully annotated data to end-users, ensuring comprehensibility at any time. Achieving this goal relies significantly on interoperability, which involves adopting a consistent approach to harmonize and integrate data. This is accomplished through standardization and the implementation of common terminologies [3]. Additionally, it is essential to model data objectively, preserving the context in which the data was collected. The context holds the meaning or purpose of the data, allowing users to effectively communicate, explore, analyze, draw conclusions, and make informed decisions.

A FAIR repository is a good starting point when trying to get a holistic view of clinical trials. Using a graph database for storing and exploring complex and unstructured data, like clinical trials, can get flexibility across the entire lifecycle from planning to archiving a trial.
By adding an application, powered by a graph database, to the top, the end-user provides the opportunity to interact with the database in the background.
All these requirements are desirable to be met for the best experience and efficiency when evaluating data in different stages of a clinical trial and make the data reusable.

REPOSITORY

FAIR

The ‘FAIR Guiding Principles for scientific data management and stewardship’ were published in Scientific Data in 2016. The intention was to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets [4].

A FAIR repository refers to a data repository that adheres to the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR). These principles aim to enhance the usability and impact of digital assets, particularly research data.

1. **Findability:** Data is easily discoverable by both humans and computers. By including metadata and a unique identifier efficient search and retrieval is facilitated.
2. **Accessibility:** Data and metadata are accessible to a range of users under clearly defined conditions.
3. **Interoperability:** Data and metadata are structured using standards, making it possible to be integrated with other datasets.
4. **Reusability:** Data is well-described with rich metadata, and provenance, enabling effective reuse by both humans and machines.

FAIR repositories play a crucial role in advancing data-driven research not only ensuring proper data storage but also maximizing usability and shareability. While these principles serve as a foundation, achieving data reusability requires a practical approach and solution.

COMPONENTS FOR AN EFFECTIVE REPOSITORY AND VISUALIZATION OF IT

Making an effective storage and visualization of a repository, includes a combination of different useful components:

- Information Model - How the data is modelled
- Ontology - How the data in the model is described
- Graph Database - How the data is stored
- Reflective Logic - How the graph database is queried and searched

These components in combination create a powerful and flexible solution [5].

AN INFORMATION MODEL THAT INCLUDES METADATA

The repository should be organized and modeled for effective data search, and it gains power when incorporating metadata with the data. Metadata tells us what we need to know about the data and to understand it. Using an information model that is built as a graph with focus on nodes, called Holons, instead of relationships creates other possibilities than a traditional relational database.

ONTOLOGY

An ontology is holding the rules for the information model and describes the structure of the content. It defines how the nodes of the graph relate to each other, which fields are included in each node, the datatypes of the fields, units to be used, etcetera. The applied oontology also provides a terminology controlling the naming of all the entities.

This model is based on a network of information units, so-called “Holons”, and their relations. Data that is tightly connected are stored together in the Holon. Examples of such data could be variables belonging to an individual event, process, or entity, e.g., a measurement result stored together with its unit and the date it was measured.

By utilizing the ontology, data will be easier to access, easier to explore, easier to share, and above all, easier to understand.

GRAPH DATABASE

Graph databases organize data in a network (like a mind-map), with each node linked to others via relationships. Unlike traditional relational databases, a property graph database internally structures components by adding properties, types, and labels to nodes and relations. Processing related data is more natural, eliminating the need for joins seen in relational databases. When combined with a conceptual ontology, it mirrors the human brain's organization of information. This framework allows seamless integration of metadata, making a graph database less sensitive to structural changes.

How the nodes and their relationships are modelled in a property graph depends on the purpose of the database. While nodes and relationships are equal partners in a graph database, models can choose to focus on either nodes or relationships. The model to choose depends on the questions you want to ask and whether the aim is to analyze the graph as a whole for analytical purposes or whether local transactions are the primary concern [1].

INTERFACE

Integrating a user-friendly interface, driven by a graph database, empowers end-users to interact seamlessly with the underlying database. A knowledge graph accessible through a web-based interface aligns well with the FAIR principles, ensuring Findability, Accessibility, Interoperability, and Reusability.

CLINICAL TRIAL

Clinical Trials are essential for new drug development and involve distinct stages from planning to archiving. To support ongoing medical progress, ensuring post-report data availability is crucial. Clinical trials are collaborative efforts involving a lot of project members e.g., project manager, sponsors, data managers, data statisticians, medical writers etc. Everyone involved has different backgrounds and interests. How could collaboration be stimulated and facilitated by a repository?

EFFECTS OF USING A FAIR REPOSITORY

Data plays a vital role from start to end of the R&D performance at pharmaceutical companies, not only for individual studies but also entire programs. Discovering and understanding all relevant data is therefore essential to successful R&D outcomes.

Utilizing a graph database for managing intricate and unstructured data, such as found in clinical trials, offers flexibility across the entire lifecycle, spanning from planning to archiving. This will prevent the need to restructure the whole data repository when new data is incorporated [6].

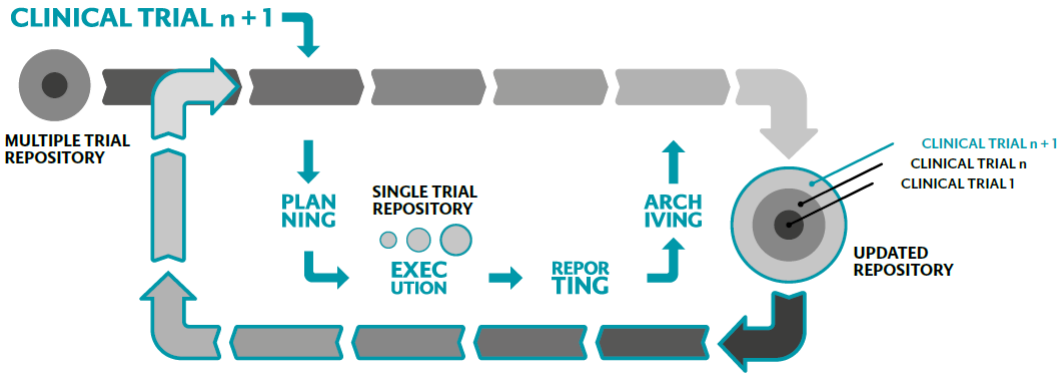

Figure 1: A multiple trial repository can be updated with another trial after it is completed. To benefit from aggregated data, it must be curated in the same way as the existing data.

PLANNING

In the planning phase, where the study is designed, study populations are chosen, patients are recruited etc., extracting valuable insights from existing data proves invaluable. The presence of a repository with diverse datasets becomes a key contributor to elevating the overall success of the trial.

EXECUTION

During the execution phase, trial performance involves the comprehensive handling of data, encompassing the collection, curation, standardization, harmonization, integration, and all aspects that contribute to making the data usable. When handling a new trial (clinical trial n+1 in figure 1), data could be added into a single trial specific repository while the trial is ongoing. The repository will grow over time. A single trial repository grants project members access and insights during the trial by integrating data from diverse systems, including biomarkers, lab results, and PROs.

This assures data quality, efficiency, transparency, and enables dynamic early decision-making. Being able to utilize an established multiple trials repository containing information from trials similar to the ongoing one, allows for an immediate initial assessment by comparing the outcomes directly.

REPORTING

In the reporting stage, project members can save time by having all data for the trial in a holistic and comprehensive view when they want to go beyond the data and explain and understand results.

ARCHIVING

Clinical trial data needs to be retained for an extended period, after the trial concludes. The completed trial (single trial) should therefore be added to the multiple trial repository for later reuse. This consolidation ensures that all pertinent studies are kept together, promoting easy accessibility and comprehensive management of relevant information.

CONCLUSION

The ability to extract insights from data empowers e.g., researchers, analysts, and decision-makers to make informed choices, advance scientific knowledge, and drive improvements in various domains. When data follows the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and is easily searchable, collaboration is seamlessly facilitated, adding significant value to clinical trials. The structured modeling, storage, and retrieval of data, tailored to its type, enable the discovery of answers to previously unforeseen questions.

Knowledge or information graphs prove to be an effective approach for working with life science data. They enable the integration of data and its metadata, offering flexibility by continuously expanding to accommodate new types of data. The combination of data, a property graph database, and an ontology establishes a robust framework for fostering innovation and collaboration in the field.

REFERENCES

1. Effective Data Modelling for Effective Data Visualization, Eva Kelty, Capish, Peter Tormay, DV07 PHUSE EU Connect 2018
2. Robust Data Stewardship with Capish Reflect, Peter Tormay, 2020
3. Data issues in the life sciences, A.E. Thessen, and D.J. Patterson, ZooKeys, 2011: p. 15-51.
4. The FAIR Guiding Principles for scientific data management and stewardship. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. Sci Data 3, 160018 (2016).
5. Unique Technology for the Future, C.Dahlbo, A.Workneh, H.Drews, PP04 PHUSE EU Connect 2018
6. New Approach to Graph Databases, A. Berg, H. Drews and C.Dahlbo, PP05 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	Eva Kelty
Company:	Capish	Capish
Address:	Lokgatan 8, 211 20 Malmö, Sweden	Lokgatan 8, 211 20 Malmö, Sweden
Work Phone:	+46 40 10 88 80	+46 40 10 88 80
Email:	catharina.dahlbo@capish.com	eva.kelty@capish.com

Website: capish.com

Brand and product names are trademarks of their respective companies.