

Flexible Solutions for Insight and Data Driven Innovation

Catharina Dahlbo, Capish Nordic AB, Malmö, Sweden

ABSTRACT

In order to analyze data from clinical trials, Clinical Data Scientists need insight into all sources of data associated with those clinical trials. Accessing all relevant and associated data, documents and images from a unified format, using fewer yet more flexible tools, can simplify what traditionally has been a complex and cumbersome process facing Clinical Data Scientists. Such tools need data integration as well as data visualization and exploration capabilities. Additionally, compatibility with existing standards, (e.g., CDISC, SDTM and ADaM) is important and often necessary. Powerful next generation data solutions will support Clinical Data Scientists in their efforts to evaluate clinical trial data more quickly, easily and completely. The ability to focus on gaining insight by reviewing the highest level down to the lowest level of data points, and understand those relationships provides the opportunity to significantly increase the level of understanding of data trends, patterns, correlations and hypotheses.

INTRODUCTION

Data is seen as raw data, in other words, an unrelated collection of values. By adding context, the collected data becomes information and starts to be meaningful. Information itself is useful only when the human brain can convert it into knowledge – a process filtered by experiences and the identification of relationships, trends and patterns.

Transforming data into knowledge is a challenge faced by the Clinical Data Scientist who is working with biometric data to support fact-based decisions. In an effort to prepare analysis ready data and elevate the level of understanding, there is a need for flexible data solutions to improve this process.

Today, many solution utilize a data model where data is structured into tables. This model requires joining and merging of tables every time a new use case is being explored. Tables are very powerful when working with a relatively small number of data points. However, as the number of data points increases, the ability to integrate multiple data points and data types becomes more complex and time consuming. Knowing where data points reside in the data base and building commands necessary to create data relationships becomes substantial. Often, programming skills are required simply to create and maintain these relationships and it's a very repetitive and time consuming work.

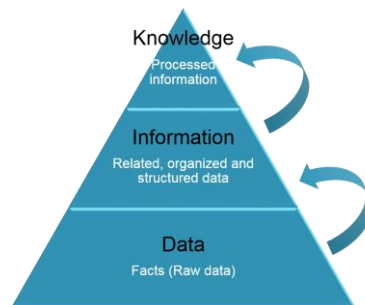


Figure 1. The pyramid shows the way of taking data to knowledge.

In contrast, using a more modern approach by starting higher in the pyramid (Figure 1) and accessing data points more directly, independent of traditional table structures, the outcome can be quicker, more productive and more easily repeated given the data is already related, organized and structured. This combination of data integrated into a flexible information solution enables the focus to be directed at the information itself rather than the way the data is organized. The traditional process of joining and merging tables using specific parameters becomes unnecessary since in the more modern structure, information is always accessible for analysis and explorative work directly by the scientist.

The focus of this paper is to describe how data from different knowledge domains is integrated and the ensuing positive effects that can result by using this modern approach. In addition, as data from domains outside of clinical data (e.g. healthcare, management, financial and metrological data) is also included and accessed via this single solution, the possibilities of understanding clinical data can be approached from a new perspective.

GAINING INSIGHT AND MAXIMIZING THE VALUE OF THE DATA

INTEGRATING DATA FROM DIFFERENT DOMAINS

Making data more accessible can be helpful in an effort to derive interesting and useful information from data that is spread out across disparate locations. Including data in a uniform format (Figure 2) provides the user with a broader understanding of all data in relation to e.g. treatment effects, safety signals and/or economic factors. All data included in the uniform format follows the same rules and terminology and is therefore easy to compare, search for and explore. Data integration will provide details of how data is related to each other but will also provide users with a unified view of this data. Integrating data from different knowledge domains and exploring data combinations in the same hypothesis expands the opportunity for identifying meaningful trends, patterns and differences both expected and unexpected. The possibility to test new hypotheses, without requiring users to have programming skills and do time consuming and repetitive work, enables theories to be evaluated very easily. The result is that the knowledge and the value of one's clinical data increases.



Figure 2. A uniform format for all kinds of data makes it easy to search, compare and explore data.

INFORMATION MODELLING

Many existing data models are designed based on how data was collected (e.g. CRF) or the type of ad hoc analysis that was requested. Often data is stored, accessed and analyzed using traditional relational or statistical databases using tables that are organized based on the type of data collected. Therefore, locating patient related data in different tables or across domains requires joining and merging tables of interest and using a complex query language. This process is often inefficient and thus typically costly both in terms of money and work effort required. A different way of enabling data of different types to be more accessible and easy to understand is to structure that data according to an information model consisting of small information entities related to each other. These small information entities can be seen as small messages representing the minimum amount of information a person with knowledge within that subject matter field can understand and use for decision-making. It is not until the information units have been linked together that the optimum value of the information model is established.

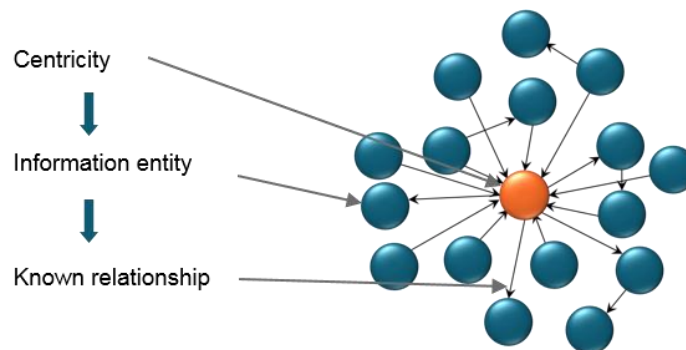


Figure 3. The created information model provide already linked and joined data ready to return direct answers to the user's questions when used via an interactive tool. The centricity e.g. patient or study is changeable dependent on interest.

This modern information model is designed to be readable and understandable by humans, as well as computers. The foundation is based on how the world can be divided into self-sustained but interconnected entities. This

granularity mimics how humans perceive the world. For example, it is not necessary to see the building blocks of a pen to understand that it is a pen. At the same time this model allows for amendments which may add breadth to the information but do not require restructuring of data relationships as is commonly seen with other models. Each entity holds data which is tightly associated (e.g., size and color of a pen) and these entities together might constitute yet another entity altogether. In other words, an entity is a whole in and of itself, but is at the same time also a part of another whole. This approach to data modelling allows for combining data as humans perceive it, yet providing a logic that computers can utilize. Adopting this method of data modelling removes the need for a user to have knowledge about the database structure in order to access and analyze data.

Having the information structured using a common information model alone is not sufficient to achieve full integration of data. Data labels and content may vary in numerous ways, though they may share exactly the same meaning. The implication thus is that a common terminology is also required to easily compare information from different data sources. This terminology coupled with the information model form an ontology. By utilizing an ontology the data will be easier to access, easier to explore, easier to share, and above all easier to understand.

VISUALIZATION - MAKING SENSE OF DATA

Data modelled for comprehension by the human brain in combination with an ontology gives a user more flexibility in analyzing data when using an interactive visualization tool. Exploring information entities that are by design already related to each other, enables user independence and self-service given the need to have any understanding of database structure is no longer required. This empowers the actual data consumers - subject matter experts and reviewers - rather than involving database experts.

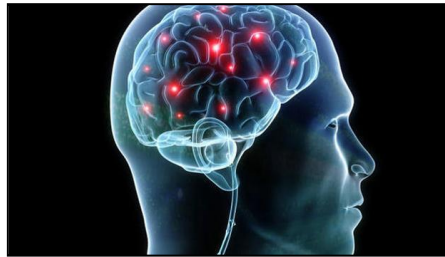


Figure 4. *Data modelled to be understandable by humans and not only computers gives more accessible data.*

Optimizing the understanding of data calls for solutions that are easy and intuitive for all users while permitting and enabling opportunities for collaboration among colleagues. The ability to share insights from data that can be reproduced and redefined in unlimited ways provides higher value and spawns data driven innovations.

The interactive graphical tool allowing any authorized user direct access to data should provide:

- Access to pooled, aggregated and all kinds of data that is possible to relate to each other
- Search both free-text and structured text on the entire information - The user can simply type a query in a search interface similar to a web search engine and quickly retrieve the data they are looking for. It is not only the data "requested" that is available in the result presentation, but all data for all patients are a click away.
- Create cohorts from data for further analysis
- Visualize the data from the whole patient pool down to individual patient information
- Analyze data from different angles to identify relevant details and get a fast result
- Explore data by navigating via the relations

PhUSE 2015

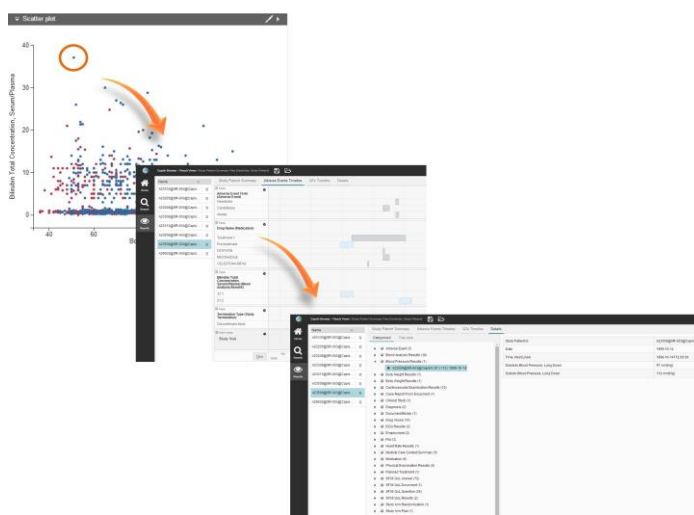


Figure 5. An example of the drill-down principle, where it is possible to identify data points of interest in aggregate graphs and tables and drill down to the relevant patient. Screen-shots taken from the Capish® applications.

BENEFITS

The combination of a metadata rich information model, based on the human thought process, coupled with a solution that provides the advantages mentioned previously, new opportunities to explore data become real.

Reusing data to reveal insight

Data can be used for many more purposes than for which it was initially intended. This facilitates an increase in the ability to explore existing data more creatively and thoroughly and potentially leads to higher returns. Information visualized using a solution that extracts the most data from a trial in a flexible manner that enables anyone in the organization to explore this data adds value to decision making. Data can be combined in many ways by experts in different knowledge domains which may lead to uncovering newly undiscovered correlations.

This information model constructs no limits to connecting different types of data. The data must just be related in a useful way. Data that has been prepared once is from then on always accessible. New and updated data is easy to add. This combination of integrating data into a uniform format, using an information model and terminology and providing data accessibility using a tool that takes advantage of the integration, is the cornerstone to revealing insights. Examples of how this combination provides value without for instance conducting new clinical studies:

- How/does the weather in different continents affect the illness of an asthma patient?
 - Metrological data for continents for the period of time the study proceeded may be added and related to the clinical data - making all clinical data and all added metrological data are available for any hypothesis to be tested. Analysts may then discover how the weather does or does not affect asthma patients in a positive or negative way.
- How sites across different studies have managed the studies?
 - After integrating data from different clinical studies collected at different times, the study data will use the same terminology and be stored at the same place. A reviewer can then evaluate all the sites separately or by country or by any other desired parameter. Sites not fulfilling necessary requirements can then be easier to identify.
- While exploring data for one reason or intention, outcomes could unexpectedly lead to new insights/patterns/trending in the data that no one previously expected to find.
 - The clinical data for a new molecule is evaluated. While analyzing graphs independently, users become interested in evaluating new hypotheses looking for new data correlations. A best case scenario would lead to identifying a new indication for the molecule that was previously undiscovered.

Collaboration

Collaboration between different departments in an organization is common during clinical trials. To further improve the outcome of these collaborations, if the same information could be accessed by many users, regardless of where they are situated, and reviewed collectively, decision making could be improved and expedited. Furthermore, independent data exploration that does not directly affect anyone else's previous or simultaneous work, is ideal. Organizations are most efficient and effective when ease of sharing results between colleagues and organizations is fast, easy and consistently uniform.

- If two or more companies are working together posing the same hypotheses but are responsible for different studies, collaborations will be smoother and less complicated if the organizations offer the same access to the same data at the same time allowing for the exchange of theories and interpretations using the same interactive solution.
- Licensing of a substance by a smaller company to a larger company could result with the smaller company not having easy access to the results from the larger company running the trials. Using a solution that enables easy access to all of the data, the smaller company could continue to use the data in their further research and possibly uncover new substances or indications.
- Researchers are given the opportunity to incorporate similar data sets, to gain a larger population base and evaluate new approaches using a solution where all integrated data is always accessible. By integrating external data with clinical data, other groups of users are included adding to opportunities to find new data correlations.

Knowledge

Running clinical trials require large investments and resources. A more modern way of integrating data and exploiting the accessibility of data in the ways described by this paper should provide a higher return on investment than was previously possible using traditional methods. The solutions must be easy for the user to understand and enable explorative searches independent of programming skills or knowledge of database structures. Creating knowledge from data, by using an ontology and a flexible solution, will help to build bridges between medicine and research across multiple different areas.

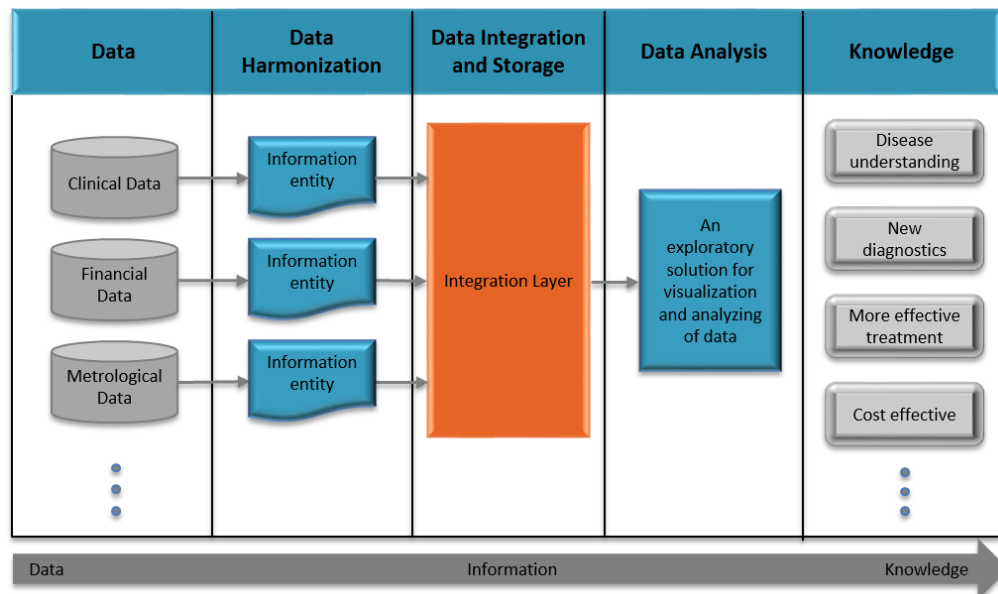


Figure 6. An overview of the process from data to knowledge.

CONCLUSION

The ability to interpret and understand all available information requires a unified view of all data in a solution that enables comprehensive visualizations and analysis. A solution that supports Clinical Data Scientists' efforts to gain deeper insights and understanding, coupled with an opportunity to share work collaboratively, needs to include integration of dissimilar data organized on a modern information model supporting unlimited explorations. By utilizing the possibility to combine different knowledge domains the data will provide more value to scientific findings and enable further investigative work without having an understanding of the underlying data structure and doing repetitive and time consuming work. Once data is integrated it is always immediately available for new explorations in concert with other knowledge domains.

CONTACT INFORMATION

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

Catharina Dahlbo
Capish Nordic AB
Stortorget 9
211 22 MALMÖ, Sweden

Work Phone: +46 (0)40 10 88 80
Email: catharina.dahlbo@capish.com
Web: www.capish.com

Brand and product names are trademarks of their respective companies.