

Does the concept of a “Sentient-Enterprise” apply to the clinical trial execution process?

Peter Grolimund, Teradata, Zurich, Switzerland

ABSTRACT

In the world of big-data and data-driven organization, the data-analytics is moving from static reporting, business intelligence, predictive analytics to the concept of “Sentient Enterprise”: This means decisions are supported or even taken by reduced interaction by the business users. This presentation is going to explain the concept of the sentient enterprise and how this could be applied in the context of the execution of clinical trials, from the protocol drafting, writing of the protocol, setting up the clinical trial, the statistical plan, and as well the analytics itself. It includes the future challenges of device based data and the forecast of outcome-related reimbursement concepts.

INTRODUCTION

Machine learning, deep learning, becoming digital, cognitive computing, knowledge management, business intelligence are new headlines in many industry domains and different concepts are behind those. But there are a few elements which are in common. All those concepts try to bring ‘knowledge’ management from the past to a next level. It is all about the efficient use of the common knowledge and decision experience within an organization to be used to improve the situation for the future. *Figure 1* does include an overview of this concept and the different stages.

Another common element is data. The data volume available is growing and gets more diverse. But new technologies, as big-data, data-lakes, new analytical tools, etc. do provide tools to master these challenges.

These approaches are evolving in many industries. By adding / using more sensors and hence data, documents, voice streams, the predictions becomes more reliable and can prevent downtimes of production lines, or help to streamline traffic and even help to identify new opportunities etc.

In this paper, the question will be discussed, how far a heavily Research driven domain as the clinical development is suitable for such approaches.

SENTIENT ENTERPRISE

The concept of a ‘sentient enterprise’ developed by our company is a vision, how companies can drive there analytical capabilities along a framework. It should allow an enterprise to make full use of the available internal and external data by applying analytical capabilities in a progressing way. To get this vision a reality it is not about purchasing tools nor does Teradata cover all functionalities and systems to realize the vision. It needs many other elements in particular as well so called soft-factors as skills, people, and an according culture in an environment which is data-driven. The sentient enterprise does include more technically or analytically driven elements as e.g. machine learning, cognitive computing etc. These approaches are part of the sentient enterprise.

In order to explain it the approach is translated into some the key processes in the drug development. It might be in some cases disruptive but we do see as a company a high momentum in other industry domains, moving in this direction. In many cases, the existing processes to leverage IT systems doesn’t match the expectation of the user community either in the speed it gets delivered nor in the functionality expected. The user community itself is looking for their own solutions which might be apps, new self-service applications etc. which does show a growing maturity and capabilities and means there is a cultural shift. Data-silo’s are overcome by enabling a self-service approach to integrate and analyze data. It is not a big bang implementation.

ATTRIBUTES OF THE FIVE core CAPABILITIES of the SENTIENT ENTERPRISE

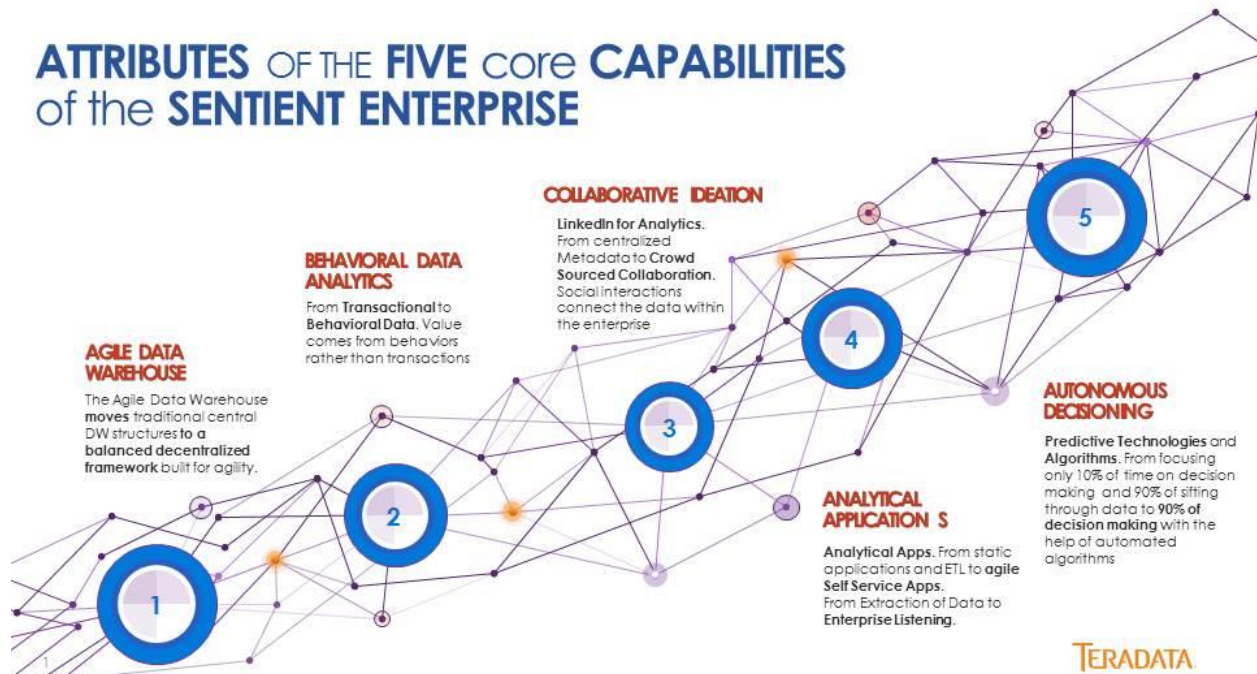


Figure 1: Sentient enterprise with the five different domains to drive analytics and decision making

The realization of the vision is a journey and the 5 different stages are not put in place in a logical sequence. It is the enterprise and its culture which sets the priorities which domain does bring them forward. The bi-modal IT (Rethinking Information Management for Bimodal IT 2015) approach in many white papers can be an approach to support the realization. In the bi-modal IT approach we have one part which is driven by clear defined basic processes and needs which are quite stable and clear (e.g. financial data aggregation and reporting). It is still changing but not as fast as the rather unknown domain, where the requirements are not as clear, or might change quickly. This will be covered by agile approaches which can cope with the uncertainty and the changes. In IT project methodologies this is already used as there is the so called waterfall approach and the SCRUM agile approach.

Some technologies or methods are already well established in many companies (e.g. KPI's, repetitive reporting, collaboration platforms, knowledge management etc.).

AGILE DATAWAREHOUSE: INTEGRATED DATA WAREHOUSE, VIRTUAL DATA MARTS, AND FEDERATION

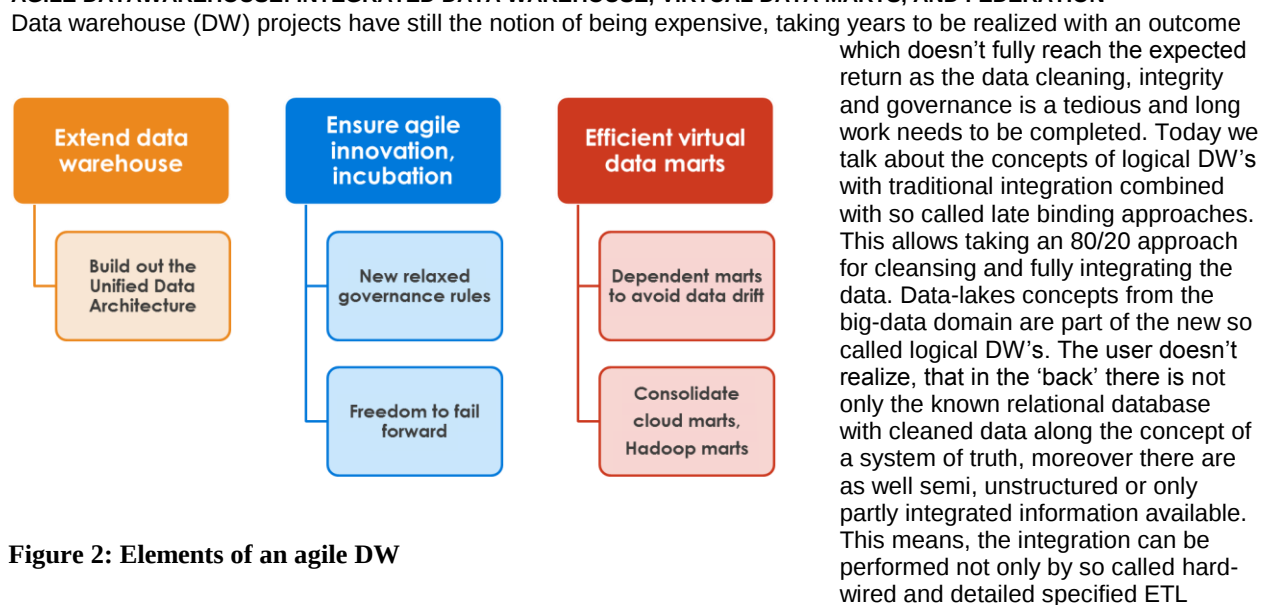


Figure 2: Elements of an agile DW

processes, but as well soft-integration approaches driven e.g. semantic connections which do link not only structured information but as well documents and raw data elements or so called key value pair tables.

To give an example how this can work is the fact, that the frequent use of code (how many times a SAS code is used in the context of a study design) can link the study design to those frequently used SAS code elements (pattern and similarity analytics) to establish a first proposal of the codes which might be helpful to perform the study analytics. This links SAS code to the study design, and may be some other documents which allows taking the knowledge from the past and the logical DW brings those elements more and more together. But it needs to be added, that in case of the study execution, the former traditional data-warehouse approach can still be of tremendous value, when it comes to the study data itself, as this is in its majority structured information. We are still in a position to demonstrate to organizations the value of an integrated clinical DW, where data pooling across trials into a common model (e.g. CDISC BRIDG) can make a big difference. But as well in this case, a stepwise approach can be taken, as the ETL process could be driven by mappings performed by the users in the past. Teradata has developed as well some techniques and functionalities as e.g. the data-lab, the Aster platform, to enable the agile DW approach.

BEHAVIORAL DATA ANALYTICS: FROM TRANSACTIONS TO INTERACTIONS

In the former chapter, some examples linked to the agile DW have been mentioned. But the concept of the behavioral data analytics is the fact, that we are using the collaborative know-how embedded in the systems and in the usage of it (structured, semi-structured or unstructured). We all are aware, that e.g. in data management data

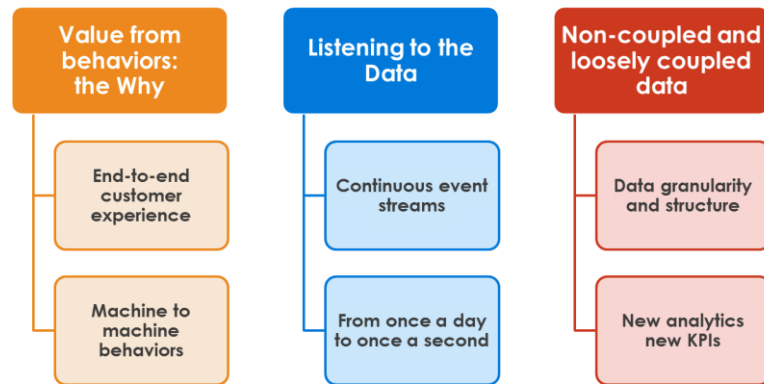


Figure 3: Elements of behavioral data analytics

entry related issues might be related to the set-up of the entry screens in the EDC systems. It might be even so, that depending on the country the sequence of the data-entry etc. is not reflecting the normal processes the doctors are used to in a hospital setting. More and more the trend is looking after an approach, where data which is collected anyway as part of the hospital system (e-Health Records) are simply taken over and not collected in a redundant way in an electronic data capturing (EDC) system provided by the sponsor of a study. A behavioral data analytics approach could reveal data entry issues patterns and propose adjustments as we go forward. It might give immediately by country or region feedback, when data is entered not correctly. The analytics of the user behavior, or the result of the use of the functions of an application in a certain way (in the case of EDC it is e.g. the data entry), can give the data manager and the application designer important information how the application can be adjusted to get less of data-quality issues and increase the user satisfaction or as well the adherence to a process. It includes the click sequence on the interface and how many times a user might move back and forward. This allows then to realize some other domains of the sentient enterprise as e.g. autonomous decision making. There is a strong link to the analytics application as well.

A domain, where behavioral data analytics might become even more relevant are devices and apps-based data collections, which are more and more used in the context of clinical trials. In such cases behavioral is related to behavior of the patients which could be revealed by looking at the data in stream and continuously. This means, the behavior of the patients can be seen and used to change the approach, correct issues as we go. The concept of adaptive trial designs might become a new notion and might end up in concepts, where the change of the approach is part of the approved trial design similar to the concept of individualized drug production. This means the protocol amendments are incorporated and do not involve separate approval steps. This means based on the 'behavior' observed (analytics) the steps of the changes have been agreed before.

COLLABORATIVE IDEATION: CROWD SOURCING ANALYTICS

Knowledge management and collaboration platforms are the applications/systems which are targeting this domain. There are many tools and technical solutions in place to foster collaboration within an organization. Most of us are working with one or several of those tools and frequently face some challenges. Either the tool doesn't offer the function we are looking for leading to the abundance of it or to declining of its use. But not only technical shortcomings are leading to a lack of leveraging the full potential of the common know-how of an organization. It's even more a cultural problem and the way we work. The number of 'lookers' into the tool is usually much higher than the contributors into a knowledge management tool is. That's why there is a wealth of articles, trainings, courses, change-management approaches etc. to overcome this hurdle. Knowledge management is something we do not something we do buy. Crowdsourcing from everyone in the analytical community is crucial in order to grab the full potential of analytics. In case of analytics it includes not only documentation of projects and initiatives but as well transparency what kind of code was used, what was the analytical product, what kind of data (data lineage), what techniques did work what didn't work, or simply who has done this before. The adherence of certain minimal

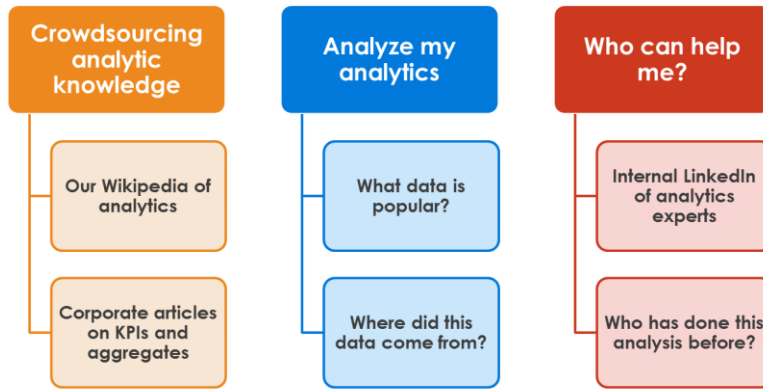


Figure 4: Elements of collaborative ideation

documentation, sharing and can only be improved, when there is a strong culture and leader who do support but in some cases as well impose the culture. It can be done by rewarding the sharing behavior.

Of course there are some technical approaches which can at least support the information search and connection. Semantic Web, smart search and link algorithm can be established to establish links to the information sources. This doesn't really lead to a change of the culture, but it might help to support the process, or even more, might give as well some analytics of the analytical processes to track the improvements of the sharing culture.

ANALYTICAL APPLICATION PLATFORM: THE APP'S

In the past analytical skills and the use of data was driven by pre-cooked reports or as well BI tools, which did allow performing drill-downs along predefined dimensions (e.g. geography, organization layers etc.). The extensions of those reports, or functions was tight to a change-management process and frequently was too heavy, in order to be reactive on what the user community was looking for. More and more tools are coming with a broad self-service

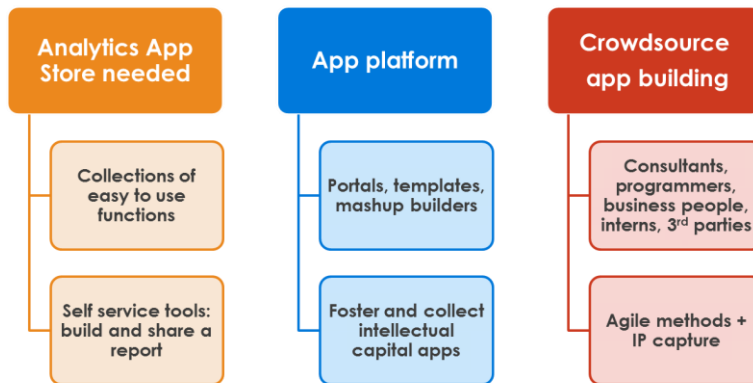


Figure 5: The elements for an application platform

visualization which gives more and more business uses an extended workbench to perform analytics in a self-service mode. Data which is not structured or more difficult to handle (e.g. weblogs, machine-logs, data produced out of apps and devices) are not as easy to be consumed and need frequently a lot of additional steps to get the analytics work for a broader audience. Sometimes, the common tool available is missing certain capabilities or there are new visualization techniques and approaches available in other sources (e.g. open source libraries). The difference of the approaches could be summarized by the fact that Teradata is looking for a solution, where the customer is keeping control of the analytics, algorithms etc. used to drive the decision making. But the scope is similar.

Existing know-how should help to drive the future decision making and processes.

The business is looking after to perform such new analytical functions and capabilities on a self-service mode. The so called shadow IT, meaning the more and more available skillset in the business to write scripts, who can use library can be used in a targeted way to contribute and manage an app-library which can provide analytical capabilities for a broader audience. There is place as well for collaboration with universities. The app platform does provide a framework to keep a certain governance and unified look and feel. It allows the business user to install an analytical app and use it along his needs. This could be a simple app for the data manager in the field, to identify event sequences which might lead to particular time-consuming and tedious data correction activities. In such a case he might be in a position to act before the effort is going to be too big or even worse, patient might need to be dropped out of the trial.

AUTONOMOUS DECISION PLATFORM: AUTONOMOUS DECISIONING

Many of our decisions we take or need to take are quite repetitive and could be handed over to a machine. In the industry robots driven by software, or software trading on the stock market are nowadays usual. Effective warehouse managements, supply chain control etc. are other examples with some still in the explorative stage as autonomous / self-driving cars. Of course, complex, science based and flexible decision making is not possible but machine

learning algorithm we can test about, give at least a certain situation-driven flexibility where the rules are embedded in the approach and driven by actual data.

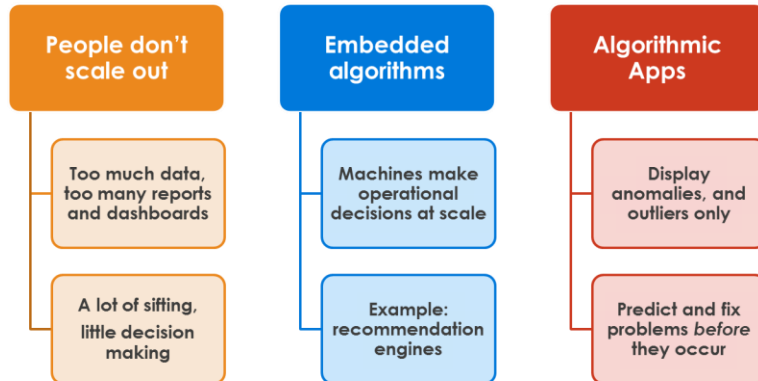


Figure 6: Elements for autonomous decisioning

setup folders, templates, codes, etc. and applies this to the document mgmt environment such, that when the study starts everything is in place. It could even adjust agreed latest practice or the best practice by pointing the algorithm to it so it could according to user assessment pick the best way to do it. Or in case of an EDC platform, it adjust the positioning of the entry elements on an ongoing basis, that takes the click-streams and how user work on the interface into consideration and aligning it with the data-management issues such they get minimized. We should keep in mind, that the skills to apply such algorithm and as well the open source libraries are growing and it will become a commodity to apply it in cases, where we would think today it might be an overkill in the context of effort vs. benefit.

THE STUDY SETUP

The design and authoring the protocol is a key activity and can even impact seriously the success of a study. This process is a controlled one involving many parties as medical affairs, the study team, the biostatistics department and even the financial control unit. The complexity of the process and tasks involved, the crucial subject matter expertise needed shouldn't prevent us to think out of the box which steps of the process could be supported to make them easier or even replaced by applying the concept of a sentient enterprise. The focus of the human resources and effort needed should be on those steps, where the real innovation and new concepts do make a difference for the study execution. In this case, the look for knowhow should go beyond the own organization and should include latest developments and successes in the competitive landscape as well.

Most probably most of people ever involved in the process have faced some common questions as e.g.:

- Did we ever execute a similar study and where are all the documents stored as e.g. protocol, the statistical plan, the cost calculations etc.
 - What kind of study design and endpoints did any competitor selected who was executing a similar study in the same disease area?
- Where were the patients recruited in similar cases internally and externally?
- What were the study costs of similar cases?
- How can I get the most recent numbers (cost-per-patient in the countries, or by CRO) and how I can model get immediate cost numbers when I do change the study design, the recruitment strategy etc.

These questions have something in common. They are mainly related to the steps of the sentient enterprise concept (see Table 1, and Figure 1).

Most of the information does exist somewhere and should be available in an easy way. The heterogeneity of the systems involved in the process (e.g. EDC system, biostatistics platform, document management systems, cloud solutions, outsourcing models with different information exchange approaches) doesn't make it easy to perform an collaborative platform easy and to capture the existing information.

The approach which could be taken for the future might be less driven by a knowledge management platform imposing major work practice than the right mixture of existing systems and big-data based analytical concept to further push the boarders into what we all are looking for.

The questions of course would be, how much resources might be needed in order to get such a vision to work and are there ways to make not easy to measure improvements visible. The answer from our experience is, it might be much less than perceived and the effort is shrinking as we speak. The key element in getting it working is a stepwise, explorative approach which helps to get the vision real.

The internet of things (IoT) is going to accelerate the process of autonomous decision making, as the information/data to drive decisions out of it, is much richer and bigger. The more sensors in a car, to more accurate the self-driving car will find its way in complex situations and the lower the risk of accidents.

The immediate reaction, to apply this is of course, the clinical trial process is not the best applying this. But if we might have a closer look, there are many steps and work we do perform along the process which can be considered repetitive. The set-up of all elements when a study is planned, prepared, and executed involves a lot of such decisions. If a machine learning algorithm takes the former clinical trial

THE STUDY INITIATION PROCESS

For this step, having all existing similar protocols (internal and may be even external) at hands is a big step. The study design for a certain indication, phase of study, may be even on a high level the endpoints (e.g. safety, efficacy as key words) could be retrieved from existing documents or even from external data sources. E.g. the trial design is partially incorporated in the information available on clinicaltrials.gov. But manual extraction takes time. It would be a great thing, if it simply extracts the important information of those designs as e.g. high level categories like double blind etc. in addition to it the endpoints, and even the arms and enrolled patients can be extracted.

This could be at least let focus people on the content rather on work of manual extraction.

If the protocol is drafted, or even during the drafting process, there might be even a continuous small affinity analytics performed about the content or part of the content which gives already a hint, which of the potential existing protocols come closest to some key features. The search should be performed as a distributed query on all internal and if possible as well on external sources (e.g. PubMed or others). It might even already put all related documents of those studies into some subfolders of the new study documentation to be used later (e.g. SAS programs written, etc.).

Such a concept includes already several steps of a sentient enterprise. Behavioral in the context, that it takes existing work (behavior) and provides information out of it. Collaborative as it does help and support the exchange of existing work performed. Apps could be used in the context of giving visual analytics capability like affinity, graphs etc. in order to identify document relationship, or even people in the organization which could be used to collaborate. Decisioning could be performed by automatically pull the SAS programs together which could be linked to the study.

TRIAL EXECUTION

Some activities during the trial execution could of course as well profit from a sentient-concept. The setup of the EDC system, which could be driven by taking existing form-libraries, the crosscheck with CDISC standard elements in the forms which does immediately reveal potential gaps in the standards. The potential of dynamic adjustments of the forms based on user behavior was already mentioned. Automated tracking of completion of the documents and their lifecycles in the document management system and put it into an overview. But as well pattern in drop-outs, selection criteria violations, or other frequent data cleaning steps which might need to be followed up. The diverse outsourcing models in place makes the situation even more complex. But as well here could be a logical data warehouse be the answer. In the context of the lab values one could get a more precise approach by using age or other condition related norm-ranges as the ranges today are applied across all people which might not be correct in the sense, that some values might be normal in such a population or even abnormal but they fall into the generic ranges. Such analytics could be performed on hospital records or patients from former trials (e.g. healthy volunteers). There are almost endless opportunities. Let's take a closer look into work of the statistical programmer and bio statisticians.

THE BIOSTATISTICS AND PROGRAMMING ACTIVITIES

One of the first and most probably as well most important step is the choice of the optimal statistical approach. Questions of costs, needed power to underpin the endpoints, former experience in this disease about data quality, but as well together with the medical community the assessment if com-medications are chosen rightly. Adverse events etc which are common in the disease domain need to be assessed, and additional information of the competitor product we might compare against is crucial. As pointed out in the former chapter, the programs used in similar studies might be available. There might be numbers from former studies (even external) which can give an indication of what was done outside (e.g. number of patients, arms, countries enrolled etc.). Now we can automate a more and more used approach to involve Real World Evidence data into the study design. It could be, by writing the inclusion exclusion criteria, the indication etc. the data available on the RWE platform in the different systems is just shown. First it might simple statistics how many patients in those samples would fit the criteria. It would immediately as well show the usual co-medications, the outcomes, the treatment costs etc. whatever is available. You might do it already today, but do you have a system which does run automatically those 'selection' criteria of the query against all available data and spits you out graphically or in tables what is inside?.

Beside that, you could see as well dropout rates from former studies internal and external which can give the opportunity to optimize the patient numbers to reach to power, in order to minimize exposure and costs.

Finally you could go even a step further in the work. If it is about data pooling, a data warehouse approach would give you all the data of your compound, or project, or similar compounds out of your data in a common format and accessible. Virtual marts might help e.g. late binding concepts, to handle the differences among the studies particularly in the domain of efficacy. But safety and demography would be available at your finger tips.

Another interesting aspect would be to perform analytics of your analytics. You could reveal how many programs and when they were running against which clinical studies, how many similar copies of the data might be around in your system to illustrate the risk of data shift. It might even be able to reveal how far in those copies the same number of patients were used in case SAS programs were used to strip down the patient number for certain purposes.

In case of outputs (listings, graphs) the SAS program can reveal (e.g. if you use a system like entiMICE) the connection of your outputs, the study report and the programs. You can see which outputs or similar outputs were used in how many studies etc.

PhUSE 2016

SUMMARY

All this may sound far to visionary. That is most probably the conclusion you draw, in case you really made it down to here. However, we see this concept stepwise becoming true within our customers. Of course, you are not going to realize such ideas in a big bang. But if there is the right spirit, support, and as well the right enabling technology, you can get step-wise into this. We did observe as well an acceleration of the collaborative mindset, and innovation spirit as people want to do interesting work and not repetitive work. Another interesting observation is the fact, that managers tend to underestimate the potential of a crowd-sourcing approach, the impact of a collaborative and innovation driven spirit.

We strongly believe that the future will be more and more data driven, and the use of former experience is in our world of fast change a key element. Another important part is the quite high-level and simplified understanding we <http://blogs.teradata.com/sentient-enterprise/means> you simplify a lot. The implication of a high performance part in a process and a line unit to another is many times not visible. If the organization is strongly bonus driven, this might end up in overall negative performance of the enterprise overall. Sentient enterprise, as it just expresses in words, is an approach where the 'data-nervous-system' allows the status of all the organs of an enterprise. From machinery, processes, finance etc. to even the mood and innovation power of the people working there. This again might sound scary. Indeed, the ethical aspect of going this path along is the most important one. We do watch and support our customers, to comply with ethical principles, data privacy, and protection and adherence to legal and regulatory aspects. If this is not given, the attempt to get a sentient enterprise sustainably running will simply not work out.

REFERENCES AND READING MATERIAL

<http://www.teradata.com/sentient-enterprise/?LangType=1033&LangSelect=true>

<https://www.youtube.com/watch?v=DIt1URig6qQ>

<http://blogs.teradata.com/sentient-enterprise/>

CONTACT INFORMATION

Peter Grolimund
Teradata
Postfach
8301 Glattzentrum
Switzerland
Work Phone: +41 79 619 37 20
Email: peter.grolimund@teradata.com
Web: www.teradata.com

Brand and product names are trademarks of their respective companies.