Research Article

What a man can think- machine can do!

Machine Learning and Artificial intelligence are the next gen technologies which will help us in evolving from conventional ways of exploring data specially, in the field of clinical research. The research and development programmers run by the giant Pharma companies can be optimized and time taken for the screening of blockbuster molecule considerably reduced.

The concept which I plan to present through this abstract will help us develop novel ways of designing and optimizing the clinical research. The data driven approach will govern us in strategizing the drug development and will be a stepping stone in driving research. The current research focuses in reviewing and analyzing the data "In Silo". The genomic scientist will look at the genomic data, clinical data manager will mostly look at the clinical data, and the toxicology expert will look at the preclinical data and so forth. With such rapid advancements in IT technologies and our statistical ways of modelling the data, it gives us an opportunity to integrate data from different sources and identify key link which may not be visible just by looking at a single source of data.

The idea is to develop an Eco system which will be a placeholder for data from multiple sources to sit and communicate with each

Anant Avasthi
FSP Head, GCE Solutions Inc, India

Correspondence: Anant Avasthi, FSP Head, GCE Solutions Inc, India, Tel: 918447779775; Email anant.avasthi@gcesolutions.com

other. This will provide a holistic view of the data across multiple and different sources. Once the data can communicate with each other, data mining can be performed to identify signals which can lead us to adaptive clinical trial design and identifying key elements in clinical research. (The below figure) Figure 1 depicts and summarizes the entire flow of data which has been conceptualized. When selecting the source of data between proteomic and Genomic, preference should be given to Proteomic data as proteomics is a study of proteins which are functional molecules in cell and represent actual condition.
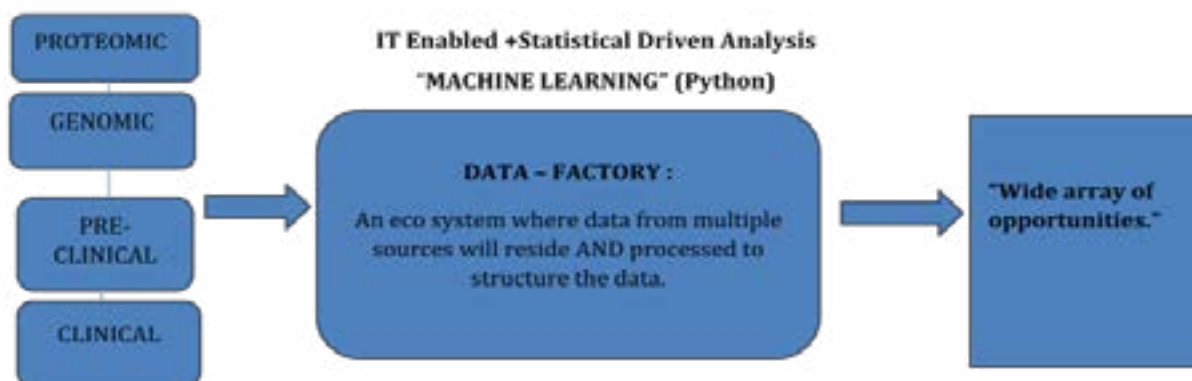

Figure 1 Data flow (Here is a summary of data flow from multiple sources and how the data resides in a single repository where it is processed using specific algorithms and programming languages such as Python, R, and SAS. The processed data yields structured data having enormous potentials to unlock mysteries which could never be identified looking at single source of data.)

So, while drawing statistical inferences from the data a high score should be should be given to proteomic Data base. The data standardization will be another key factor that needs to aligned, so that the data from multiple sources can communicate and we can seek crucial information from the data factory. The data standards such as CDASH, SDTM (for clinical) and SEND (nonclinical) will all come under single umbrella where there will be a uniform standard to understand integrated data from Data Factory. The next and the crucial stage will be developing key algorithms which can familiarize themselves with the integrated data and help to simulate a critical clinical research model. Future is exciting the question is how quickly we can embrace this change.

Opportunities

1. Simulating and modelling the trial results before executing them.

2. Increasing the probability of having positive outcome trials.

3. Devising novel approach to design and identify drugs using Genomic and preclinical data.

4. Centralized and holistic strategy to perform clinical research.

5. New gateways to research.

6. With huge chunk of data already available in the open source (PubMed, social media and NGOs, accessing data and performing quick analysis is a possibility.

Appendices

1. Data Factory: A cloud service for processing structured and unstructured data from any source.

2. Proteomics: Large scale study of proteins which are vital part of living organism with many functions

3. Genomics: is an interdisciplinary field of science focusing on genomes. A genome is a complete set of DNA within a single cell of an organism, and as such genomics is a branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

4. Machine Learning: is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data. The process of machine learning is like that of data mining.

5. Python/R: Programming languages used in AI. (Artificial Intelligence).\

6. IT: Information Technology

Challenges

1. Data from different sources need to communicate with each other and should have a common language or standard.

2. Emergence of new data standard to streamline and synchronize the preclinical clinical and genomic data.

3. Development of algorithms which can breathe in such ecosystem and educate themselves on clinical and preclinical data.

4. Evolution of Data Manager Roles from being a "Reviewer" to a "Data Scientist".

With such deluge of data, cloud based technologies and data science can provide deep insight and bring the data alive. Let's step forward and unlock these mysteries.

Acknowledgements

Conflicts of interest

None.

References

1. www.cdisc.org/standards/foundational/cdash

2. www.cdisc.org/standards/foundational/sdtm

3. www.cdisc.org/standards/foundational/send

4. https://msdn.microsoft.com/en-us/magazine/mt614253.aspx

5. www.journals.elsevier.com/genomics/

6. www.proteomicsdb.org/