Science For A Better Life

Leveraging Biq Data in Clinical Trials

08-10-2017 / Shital Desai and Basker Gummadi

Science For A Better Life

Disclaimer

The opinions expressed in the white paper are those of the authors and do not necessarily represent the opinions of PhUSE, members' respective companies or organizations, or regulatory authorities. The content in the document should not be interpreted as a data standard and/or information required by regulatory authorities.

Agenda

01 **Historical Collaborative Efforts (Transcelerate Database)**

02 **Big Data in Clinical Trials**

03 **Roche GLIDE Program**

04 **Next Generation of Uses:**
Real World Big Data
Clinical Trial Data Sources Flow

05 **Analytical Ecosystem as Next Generation Use of Big Data**

06 **Proposed Ecosystem as Next Generation Use of Big Data**

07 **Use Case: Data Mining of Genomics datasets for Non-Hodgkins Lymphoma**

08 **Use Case: Genomics datasets for Non Hodgkins Lymphoma Patient Cohort Analysis**

Executive Summary

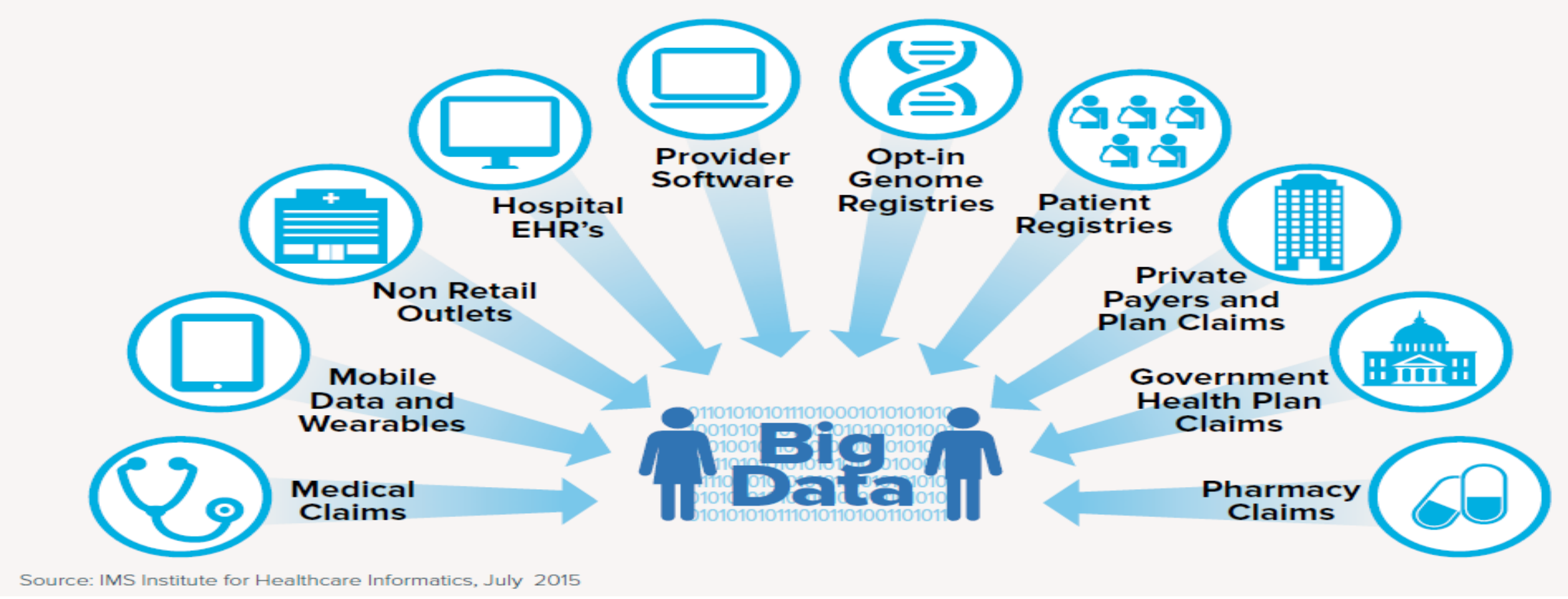
Executive Summary

Industry wide Clinical Trial collaborative efforts offers significant improvement over siloed individual databases in providing superior Patient Outcomes. The efforts however were still limited to Rare Disease categories and Data Sources resulting in limited Clinical Analyses and Insight. An industry wide Clinical Collaborative Data Repository utilizing Big Data will enable Pharmaceutical Cos to utilize new Analytic techniques and optimize Patient Journey from Drug Discovery to bedside treatment.

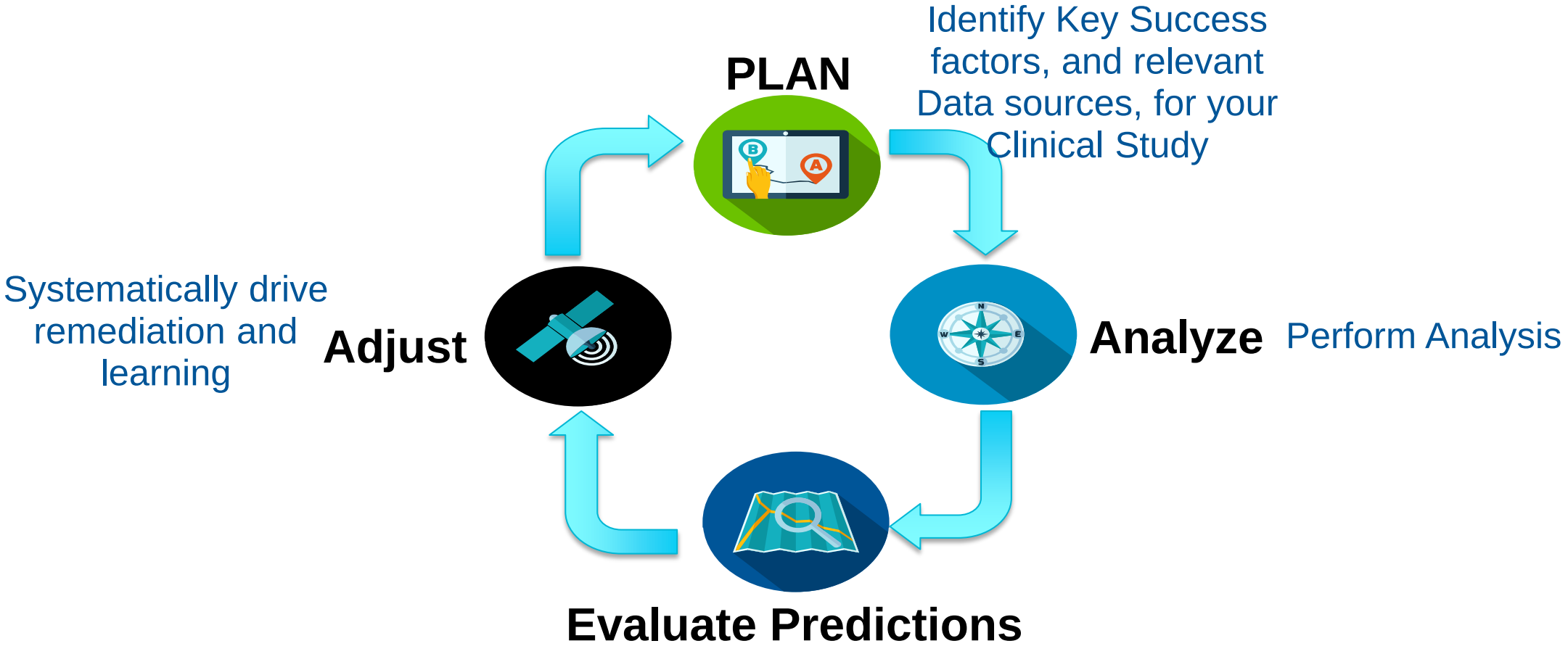
Mining Big Data Enables:

- Learn from historical data to optimize Study Design, Conduct and Analysis.
- Perform simulations to mitigate the risk of time delay for clinical trials.
- Perform predictive modelling with EHR and genomic datasets across numerous data providers.
- Glean insights from clinical data including unstructured patient's notes, scans and pathology reports.
- Empowers government agencies, payers, and providers to make decisions about drug discovery, patient access, and marketing

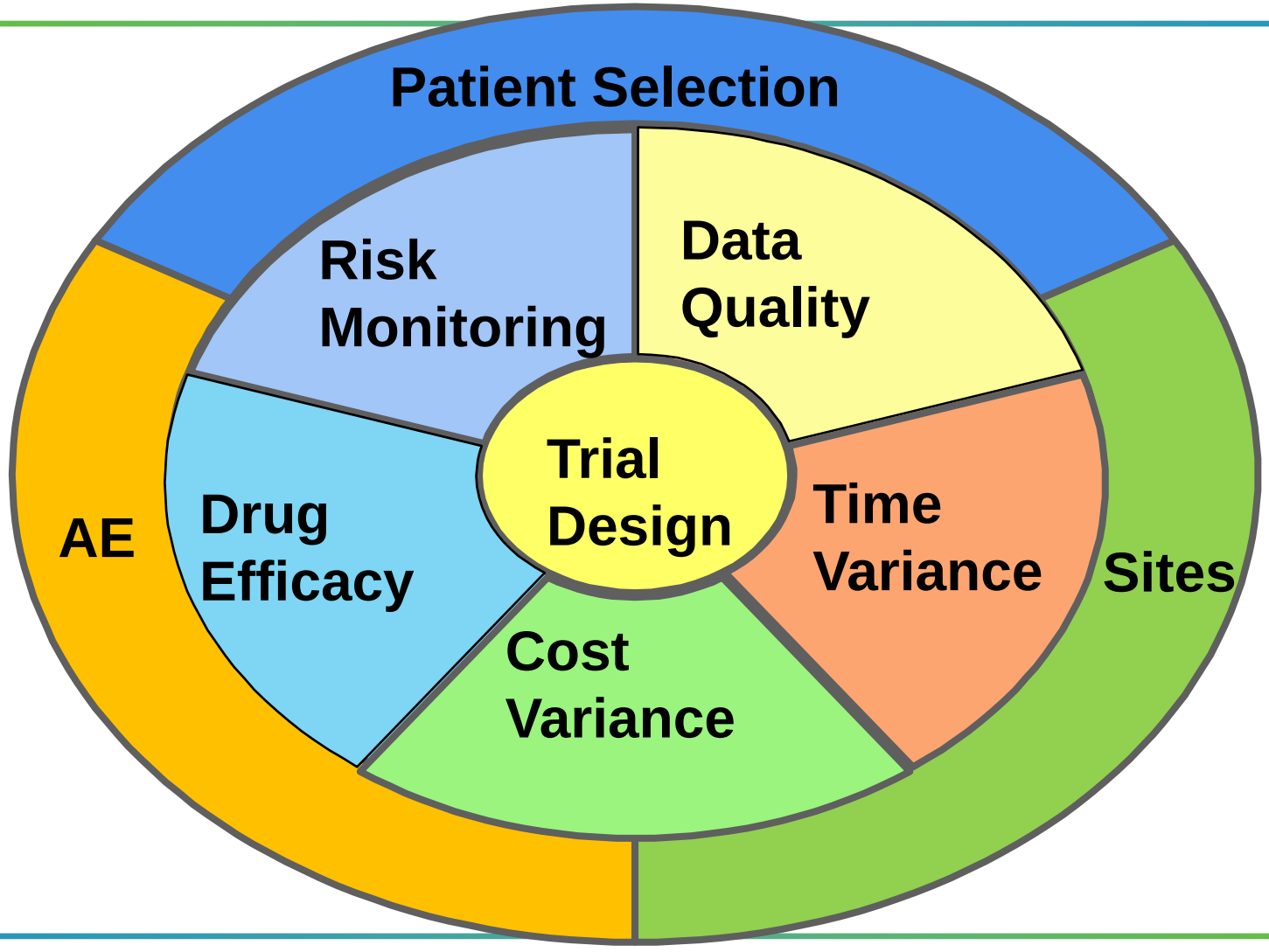
Potential Sources of Big Data for Clinical Trials



Leveraging BIG DATA



Key Success Factors for Clinical Trials



Case Study of Big DATA: Roche GLIDE Platform

GLIDE (**G**lobal **I**ntegrated **D**rug **D**evelopment **E**nvironment) analytical ecosystem, a next generation data architecture. Roche has identified Real World Data (RWD) - collected outside of a clinical trial - as a source of data type which can help to deepen the understanding of the disease area and ultimately improve access for patients to the medicine.

Multiple internal and external data sources are brought into the Teradata integrated warehouse, including:

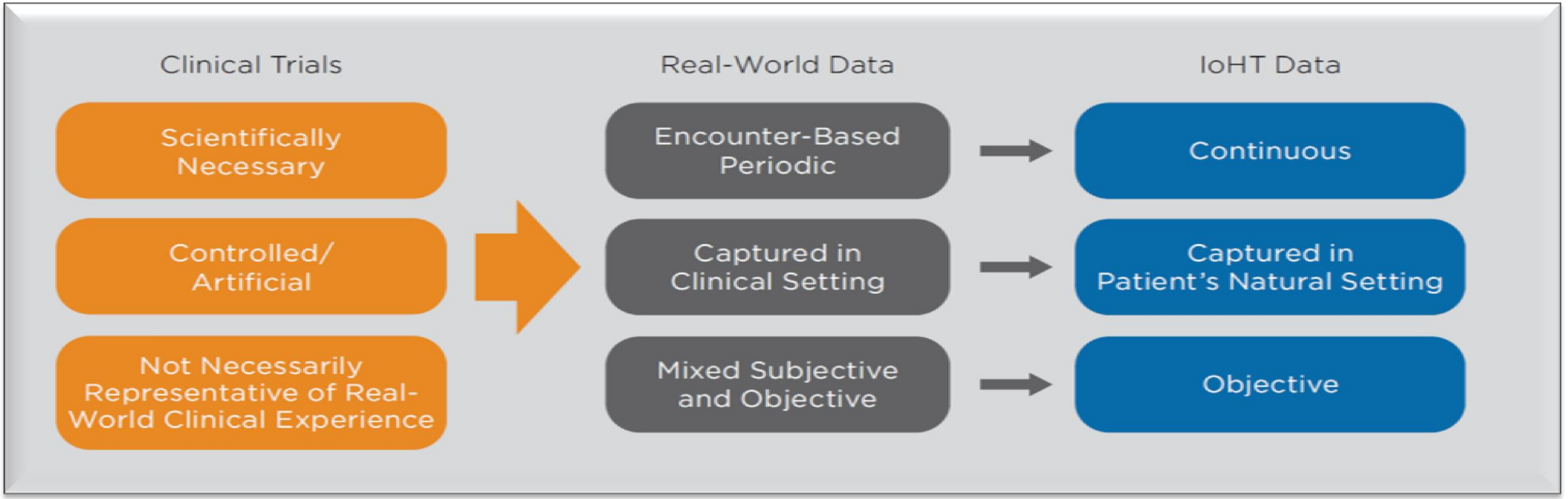
- Trial data/new treatment data (usually in a SAS data set);
- Lab data – blood, ECG, X-Ray;
- Genetics data;
- Electronic data records;
- Insurance claims;
- Medical data

Next Generation of Real World Big Data

Data from the “Internet of Healthcare Things” (IoHT) and Genomics datasets closes gaps in real world data

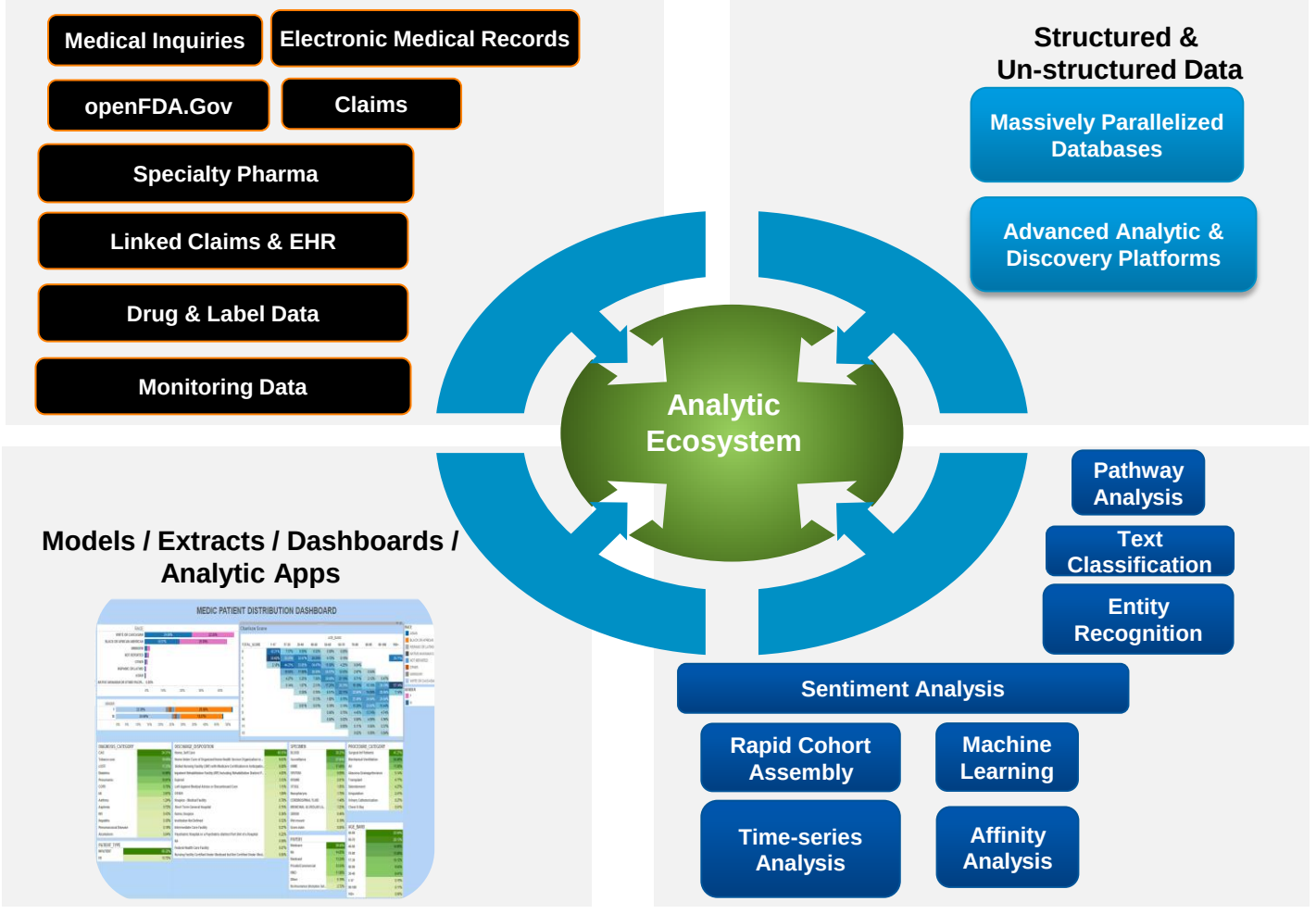
Gaps In Current Real World Data:

- **No / unreliable data between encounters**
- **Artificial setting can yield anomalous data**
- **Much of the data is subjective**



Example : Study of cancer genomes to reveal abnormalities in genes that drive the development and growth of many types of cancer. Understanding of systematic biology results in improved method of diagnosis and treatment.

Analytical Ecosystem as Next Generation Use of Big Data



Capabilities

- Plotting common pathways across billions of events
- Visualizing relationships across millions of entities
- Deriving insights with NLP, text mining, and sentiment analysis algorithms
- Detecting anomalies within large volumes of machine-data
- Accelerating next-generation predictive models (i.e., Machine-learning) with distributed computing

Use Case: Non Hodgkins Lymphoma Clinical Trial

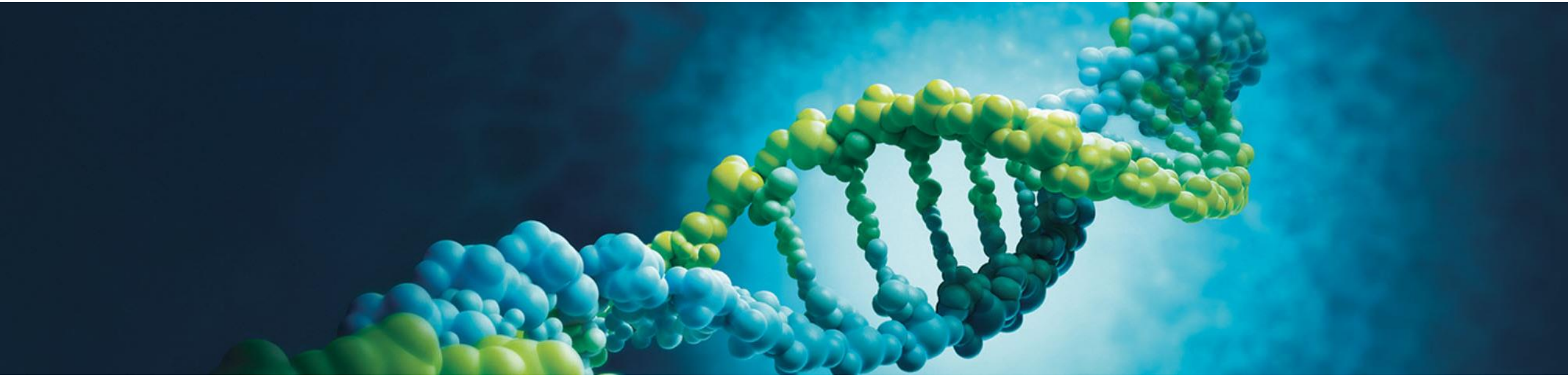
Stage I	Involvement of a single lymph node region (I) or localized involvement of a single extralymphatic, organ or site
Stage II	Involvement of two or more lymph node regions on the same side of the diaphragm (II) or localized involvement of a single associated extralymphatic organ or site and its regional lymph node(s), with or without involvement of other lymph node regions on the same side of the diaphragm (II)
Stage III	Involvement of lymph node regions on both sides of the diaphragm(III), which may also be accompanied by localized involvement of an associated xtralymphatic organ or site (IIIE), by involvement of the spleen (III) or by both
Stage IV	Disseminated (multifocal) involvement of one or more extralymphatic organs, with our without associated lymph node involvement, or isolated extralymphatic organ involvement with distant (nonregional) nodal involvement

Use Case: Non Hodgkins Lymphoma Clinical Trials

Data mining of public Genomics dataset identified the following Patient cohort insights during trail design

- The rearranged immunoglobulin heavy-chain variable region genes from both diagnoses were compared to each other. Twenty-six patients presented with both diagnoses.
- Twelve had NHL as the primary disorder (“HL first” group) and the majority of these (75%) presented with aggressive lymphoma as the second lymphoma.
- In contrast, in the 11 patients for whom NHL was the primary disorder (“NHL first” group), this was usually (82%) of low-grade histology. Three patients were diagnosed concurrently with both diseases.
- Leveraging sequence of heavy chain region resulting in identification of prospective patient cohorts for trial hypothesis of primary disorder patients and histology pattern.
- Other usage of data mining genetic region included confirmation of Second Line Therapy (bendamustine, bortezomib, rituximab)
- Administration of protease inhibitors concurrently with Chemotherapy and verify lack of interactions or changing to non-protease based regimens

Science For A Better Life



Thank you!
Q&A