DESIGNING A FAST AND SCALABLE DATA PLATFORM IN THE CLOUD – A MODERN APPROACH

Prakriteswar Santikary, PhD
Vice President and Global Chief Data Officer
ABOUT ERT

MINIMIZING RISK & UNCERTAINTY, SO YOU CAN MOVE AHEAD QUICKLY WITH CONFIDENCE

Founded in 1977; privately held
2500+ employees
Operations in 12 countries
Supporting Pharmacos, Biotechs & CROs around the world
GLOBAL REACH & EXPERIENCE

SUPPORT FOR...

>50% of all FDA approvals over the past 4 years

550+ new drug approvals

13K+ trials

230K+ sites spanning 114 languages & 106 countries

3M+ patients

Boston, MA
Boulder, CO
Bridgewater, NJ
Cleveland, OH
Philadelphia, PA
Raleigh, NC
Rochester, NY
Scotts Valley, CA
St. Louis, MO

Brussels, BE
Estenfeld, GER
Geneva
Nottingham, UK
Peterborough, UK
Tokyo, JP
**CLINICAL TRIAL INDUSTRY CHALLENGES**

01 **Data Collection**
- Multiple data sources
- Multiple devices
- Multiple data types
- Multiple vendors
- Patient recruitment
- Site qualification

02 **Data Processing and Access**
- Real time data ingestion at scale
- Real time data processing
- Data silos – lack of governance
- Difficult to identify issues and resolve them in real time
- Lack of visibility into actions being taken to address issues

03 **Data Quality**
- Patient compliance
- Site compliance
- Query resolution
- Visit compliance
- Regulatory compliance

04 **Oversight and Monitoring**
- Oversight of trials and sites
- Oversight of vendors
- Oversight of systems
- Monitoring of trial risks
- Data security and privacy
SYSTEM DISPARITY IS GROWING WITH SPECIALIZATION

MORE SYSTEMS ARE INTRODUCED IN CLINICAL RESEARCH EACH YEAR

Established eClinical Systems

- EDC
- CTMS
- IVR
- mHealth
- Images
- Labs

eCOA

mHEALTH

HOME MONITORING

PATIENT IDENTIFICATION

PATIENT ENGAGEMENT

EMR
## TRIALS GETTING MORE COMPLEX

<table>
<thead>
<tr>
<th>Key Insight</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increase in Complex Trials</strong></td>
<td>Increasing interest in building protocols based on real-time outcomes; mitigating risk of protocol change mid trial</td>
</tr>
<tr>
<td><strong>Advancement of Precision Medicine</strong></td>
<td>Technology advancements driving highly individualized patient therapeutics administration.</td>
</tr>
<tr>
<td><strong>Patient Centricity</strong></td>
<td>Shift in more active patient engagement - improving access to trials via mHealth, virtual trials and focusing on outcomes.</td>
</tr>
<tr>
<td><strong>More Big Data &amp; AI Modeling</strong></td>
<td>An explosion of data from wearables, genomics, social, imaging, etc. – driving advances in data interaction and visualization.</td>
</tr>
<tr>
<td><strong>Blockchain</strong></td>
<td>The promise of Blockchain to help concerns over privacy, data sharing and reproducibility.</td>
</tr>
</tbody>
</table>

### Key Insights
- **Technology advancements** driving highly individualized patient therapeutics administration.
- **Ability to do smaller targeted trials** improving outcomes based on lifestyle, environment and genetic makeup.
- **Further investment** in clinical trials that meet patient needs – less burdensome, increased diversity and better outcomes.
- **Need to standardize data sets** - mine previous data sets, and better ingest, analyze and manage larger, more complex data.
- **Increased investment in blockchain technology** to improve data security – competitive differentiation.
EXPONENTIAL DATA GROWTH IN LIFE SCIENCES

**VOLUME**
More types along with more endpoints

**VELOCITY**
Digitization has increased the speed of information

**VARIETY**
Clinical, compliance, medical, RWE/commercial and actimetry

**VERACITY**
Ability to gain insight from aligned data sets

**Implications**
- Organizations must implement scalable infrastructures to deal with increased data volumes
- Leading customer experiences require contemporary insights that can be acted on
- Flexible data models are required to integrate customer information
- Need to enable multiple user groups with consumable insights
WAVES OF INFORMATION THAT AREN’T ACTIONABLE

CLINICAL TRIAL TEAMS ARE DROWNING IN DATA

› High administrative burden to extract and re-enter data into other systems
› Difficult to filter out noise to identify real issues that require attention
› More effort is spent confirming the accuracy of data than interpreting it for decision-making
› Cost overrun and delay to time-to-market

Our customers are crying out for help – “Please simplify the complex!”
MODERN PLATFORM ARCHITECTURE
PUT HIGH QUALITY DATA IN THE HANDS OF PEOPLE WHO NEED IT – A BUSINESS IMPERATIVE

Our Approach

01 Data Governance
- Who owns what data
- How data can be used
- Find data, understand it
- Find data that matters - fast
- Trust the data quality
- Make big data meaningful
- Framework for collaboration

02 Modern Data Platform
- Disparate data sources
- Multiple integration paths
- Dealing with unstructured data
- Dealing with binary data
- Dealing with streaming data
- Real time ingestion at scale
- Self service data access

03 Data Quality/Regulatory
- HIPAA
- GDPR
- Mitigate risks
- Avoid steep penalties
- Transparent data access

04 Master Data Management
- Single version of truth
- Lifecycle of master entities
- Technology, Process
- Stewardship, clarity
- Data Lineage

© Copyright ERT 2018
TRANSFORMING DATA INTO STRATEGIC BUSINESS ASSETS USING MODERN, SECURE, CLOUD-BASED DATA PLATFORM

Data platform architecture at a glance
- Lambda architecture (batch and speed layers)
- Modular microservices architecture
- Serverless Computing
- Master data management/data governance
- Data ingestion at scale
- Adapters and connectors for third party data integration
- Modern data visualization architecture

Customer benefits at a glance
- Enables real time decision making
- Enables self-service reporting and analytics
- Enables data monetization
- Reduces complexity and improves quality
- Enables Data-driven decision making
- Enable Speed to market of life-saving drugs
- Centralizes risk and performance management of trials
- Enables cost savings & operational efficiencies
- Enables AI and ML capabilities – anticipatory oversight

SCALE

CLOUD BASED

MODERN TECH STACK

GOVERNED

REAL TIME REPORTING ANALYTICS AND DATA SCIENCE
ERT’s APPROACH - MODERN DATA ARCHITECTURE AND SERVICES

Reporting, Analytics and Data Science Services
- Reporting
- Self-service BI
- Open Data API
- Data Science

Data Availability
- Data Quality
- Data Consistency
- Data Security
- Data Auditability
- Data Lineage

Master Data Management, Metadata Management and Data Governance
- Data Standards
- Data Policies and Procedures
- Business Metadata
- Technical Metadata

Data Architecture and Data Technology
- Relational
- Dimensional
- In-memory
- Polyglot

Data Integration, Data Services and Data Adapters
- Structured Data
- Unstructured Data
- Semi-structured Data
- Binary Data

Data Sources
- Internal
- External
- Third party
- Future M&A
Microservices Architecture

Characteristics

• Decentralized
• Independent
• Do one thing well
• Polyglot
• API-first Design
• You build it; You own it

Benefits

• Agility
• Innovation
• Quality
• Scalability
• Availability

Challenges

• Distributed systems
• Monolith->Microservices transition not easy
• Organizational issues (DevOps)
• Skillsets

Build scalable platform and applications
Serverless Architecture

MONOLITHIC APPLICATION

Microservice

Function

Function

Function

Function
Serverless Architecture

Characteristics
- Function-as-a-service
- Compute as a service (100ms interval)
- Stateless
- Ephemeral
- Event-triggered

Benefits
- Code without provisioning
- No server HW to maintain
- No server SW to maintain
- Increase productivity
- Scale your code with HA
- Pay-per-use CPU cycle
- Zero administration

Challenges
- Vendor lock-in
- Vendor control
- Monitoring and debugging
- Startup latency

Build applications without having to manage Server
LAMBDA ARCHITECTURE

SPEED LAYER
- Low latency data
- Supports complex analysis

BATCH LAYER
- Data processing
- Responds to queries
- Support large volume of high-velocity data
- Near real-time analysis and persisted history

SERVING LAYER
- Provides access to data
- Supports reporting and analytics
Data Science Platform at Scale

01 Data pre-processing
- Garbage-in, garbage-out
- Missing value
- Outliers
- Data quality (cleaning)
- Normalization
- Transformation
- Requires Platform that Scales

02 Feature Engineering
- Difficult step in AI/ML
- Time consuming
- Requires expert domain knowledge
- Feature selection
- Feature extraction
- Requires Platform that Scales

03 Algorithm choice
- Analytic sandbox
- Availability of integrated dataset
- Requires Platform that Scales

04 Model Development
- Productization
- Automation
- Dynamic training of model
- Requires Platform that Scales

Scalable Data Platform Enables AI
KEY TAKEAWAYS FROM TODAY’S DISCUSSION

Data Governance

To unleash the potential of data - Master data management, data quality, data profiling, data policies and standards – fosters cross-organizational collaboration

Modern Data Foundation

Ingest any data of any type of any velocity, Data processing at scale, API-first design, Self service, Real time and batch, enables advanced analytics capabilities including AI and ML

Data security and access

Governed data store, transparency, regulatory compliance, privacy and protection
QUESTIONS?

Prakriteswar Santikary, PhD
prakriteswar.santi@ert.com