Building an Intelligent Biobank to Power Research Decision-Making

Lori Ball, Chief Operating Officer, President of Integrated Client Solutions, BioStorage Technologies, Inc.

Brian J. Brunner, Senior Manager, Clinical Practice, LabAnswer

Suresh Chandrasekaran, Senior Vice President, Denodo Technologies
1:00 - 1:05 pm  Workshop Introduction by ISBER

1:05 - 1:15 pm  Research Sample Intelligence: The Growing Need for Global Data Integration
Lori Ball, Chief Operating Officer, President of Integrated Client Solutions, BioStorage Technologies, Inc.

1:15 - 1:25 pm  Building a Research Data Integration Plan and Cloud Sourcing Strategy
Brian J. Brunner, Senior Manager Clinical Practice, LabAnswer

1:25 - 1:35 pm  How Data Virtualization Works and the Value it Delivers
Suresh Chandrasekaran, Senior Vice President, Denodo Technologies

1:35 - 2:00 pm  Panel led Workshop Questions

2:00 - 2:15 pm  Panel Solution Discussion

2:20 - 2:30 pm  Participant Q&A
Research Sample Intelligence: The Growing Need for Global Data Integration

Lori Ball, MBA
Chief Operating Officer, President of Integrated Client Solutions
BioStorage Technologies, Inc.
Biobank Sample and Data Stakeholders

**Patients and Physicians**
Collect samples from patients to utilize for testing, diagnosis and therapy decisions

**Research Scientists and Organizations**
Require samples to support translational science, biomarker development and research studies (ex. Industry, Academia, Foundations, Government, Hospitals, Research Institutes)

**Biobank Inventory Managers**
Manage sample inventories at biobanks to support research scientist requests

**Biobank Operations Personnel**
Support biobank managers in day-to-day management of sample storage, logistics, retrieval, materials and data management.
Development of effective data management systems is the #1 challenge to individuals managing biobanks today.¹

The most important part of data management

Analysis of sample data enables us to detect genetic differences in individuals which can improve therapy decisions and success.

- 30% of patients over-express the HER2 protein and are unresponsive to standard therapy
- 60% of melanoma patients who have the B-Raf protein were not effectively treated for skin cancer until vemurafenib was developed in 2011

Source:
¹ IQPC Global Biobanking Study, February 2015
Global Sample Data Integration Considerations

“As cloud computing extends its reach into clinical trials, data analysis, analytics and visualization are becoming key areas demanded by trial managers.”

- **Sample and Data Types** – Range of complexity
- **Data Sources** - >25B connected devices today and >50B estimated by 2020
- **Data Virtualization** (Data Integration) – Broad virtual data access with open APIs
- **Data Visualization** – Intuitive, flexible decision-support
- **Data Devices / Tools** – Need for standards
- **Data Analytics / Analysis Tools**

---

1Centerwatch, “Cloud Computing Expanding into All Areas of Clinical Trial Conduct”, August 2014
Value of Global Sample Data Integration

- Consolidate global sample inventory data
- Virtually connect global sample data to relevant research data
- Access sample data anytime from the cloud from anywhere in the world.

- Create visual dashboards and reports of sample data to support analytics and deliver business insights
- Assess the quality and value of samples stored for future research

- Select the best samples that are consented for research
- Dispose of samples not needed or compromised
- Determine which sample data and assets to utilize or share with research partners
Development of effective data management systems is the #1 challenge facing individuals managing biobanks today.¹

"Technology promises to deliver more value to clinical trials as it enables a real-time view of the patient experience and collects data that had previously gone uncaptured."²

Sources:
Building a Research Data Integration Plan and Cloud Sourcing Strategy
Development of effective data management systems is the #1 challenge facing individuals managing biobanks today.¹

There are 3 primary strategies in the market today for Sample Data Integration:

1. Systems Consolidation
2. Systems Integration
3. Systems Externalization

All of these strategies include an Integration Platform, Harmonized Data Store and Visualization to enable a single view into Sample Data.

These options will be discussed in the context of cost, complexity and maintenance. On a scale of 1 – 5 with 5 being the highest.

“Technology promises to deliver more value to clinical trials as it enables a real-time view of the patient experience and collects data that had previously gone uncaptured.”²

Sources: ¹ PharmaVoice, “The Internet of Things, March 2015
• Create a single, consolidated and integrated view to all sample data
• Integration can be either traditional ETL, ESB or Virtualized Database
• Integration requires interfaces with a wide range of systems
• Client and External source data is integrated as required
Global Sample Data Integration
Opportunities / Challenges

Systems Consolidation

Assume: Mass Systems Consolidation, ETL/ESB Integration, Internal Infrastructure

Cost: 5
- Cost to transition all systems
- Infrastructure
- SW Costs

Complexity: 4
- Challenge to harmonize and standardize during development
- Challenge for single platform

Maintenance: 2
- Lower integration cost ongoing
- Internal System Dependencies considered for system changes
Global Sample Data Integration
Opportunities / Challenges

Systems Integration

Assume: ETL/ESB Integration, Complex Master Data Management (MDM), Internal Infrastructure

**Cost:** 4
- Cost to integrate all systems
- Infrastructure Costs
- SW Costs

**Complexity:** 3
- Need harmonize internal data through integration platform
- Need to integrate ext data

**Maintenance:** 3
- Integration cost ongoing
- Ongoing Infrastructure and SW maintenance cost
Global Sample Data Integration
Opportunities / Challenges

Systems Externalization

Assume: Data Virtualization & Integration Platform, Complex MDM, Cloud Infrastructure

**Cost: 2**
- Cost to integrate systems
- SW & Infra costs are leased
- Data Virt lowers integrate cost

**Complexity: 3**
- Need to harmonize internal data through integration platform
- Need to integrate ext. data

**Maintenance: 2**
- Integration cost reduced
- SW & Infra are elastic with demand and growth
The Cloud Security Alliance defines cloud computing as:

Cloud computing (‘cloud’) is an evolving term that describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from the underlying infrastructure, and the mechanisms used to deliver them.

Other perspectives:
- Cloud computing is processing and storing of data as a service outside of the traditional processing that takes place inside an organization
- Enables organizations to treat data as a commodity that can be processed, transformed and delivered by outside organizations in a secure, compliant manner

Amazon Web Services approved by U.S. DoD as a Service Provider


Other Resources:
- EU Agency for Network and Information Security Risk Assessment Tool
- DoD Guidance on Cloud Computing

Adobe Acrobat Document

Laboratory Informatics

Cloud Sourcing Discussion
Selecting the Right Data Integration Strategy

- Consolidate Duplicative Systems, utilize COTS Platforms as basis for Core Systems
- Cloud Platform used for secure processing and storage; elastic based on need
- Data Virtualization enables agile, flexible data integration
- External Service providers can be used to process, harmonize and visualize data
Suresh Chandrasekaran
Senior Vice President, Denodo

How Data Virtualization Works and the Value it Delivers
Data Virtualization combines disparate data sources into a single “virtual” data abstraction layer (aka information fabric) that provides unified access and integrated data services to consuming applications in real-time (right-time).
DV Enables Agile Info Architecture – Many to Many

- Agile BI & Apps
- Faster Solution
- Any Data, Anywhere
- Integrate New Sources
- Optimized Performance
- Best practice
- Staging and pre-integration (mix of physical and virtual)
- Application format mappings (pure virtual)
- Data Services
- Consumers

- BI /Reporting
- Analytics
- Visualization
- Application
- Data Sharing

- Database
- Database
- Service
- Files
- Main-frame

- Canonical business models
Data Virtualization in Life Sciences

Business Solutions
Access Information-as-a-Service

Data Virtualization
Right Information at the Right Time

Disparate Data
Any Source
Any Format
**Data Virtualization Functions**

- Access **heterogeneous data sources**
- Create logical **abstraction layer**
- Deliver **canonical** business views
- **Data services** – access anywhere
- Agile **data integration** and quality
- Minimize replication (cost, time)
- Enterprise to cloud **scalable, secure**
Major Companies Find Compelling Benefits

CONCLUSION

While only certain scenarios in our PoC benefited from our current data virtualization solutions, the growing importance of implementing agile solutions for extracting business intelligence from big data and traditional enterprise data suggests that such methods will grow in importance and sophistication. Our findings provide initial guidance for the use of data virtualization at Intel and are driving our interest in investigating dedicated data virtualization solutions.

NNSA's Product Realization Integrated Digital Enterprise (PRIDE) program is built on a data virtualization fabric.

“Accessing and aggregating data from different NNSA sites would not be possible without Data Virtualization. It expedites projects by enabling fast and secure movement of data through different facilities.” - Stephanie Elsea, Pantex Plant, NNSA

“Data Virtualization is a “Game Changer” for Intel's Data
Guillermo Rueda, Abi Borber & Stacie Hall
Enterprise Data World 2015, Washington, DC
March 29 – April 3, 2015

Solutions that deploy quickly and provide a performance boost

U.S. Nuclear Warhead Lifecycle Phases

NNSA provides critical support to transporting weapons and weapon components.

SearchBusinessAnalytics

Genera said the new setup can store larger data sets and new types of information while freeing business units from having to manage data and prepare it for analysis themselves.
Top CIOs in Life Sciences Use Agile Data & Cloud Intelligence

Amgen, a $17.3 billion pharmaceutical company uses statistical analysis of data points collected in real time. The analytics system includes virtualized data warehousing tools from Denodo. Amgen can draw conclusions much faster than in the past, says CIO Diana McKenzie.

Now You Can Too!
WORKSHOP SESSION
WORKSHOP QUESTION #1
What is your current state of sample data integration?

Take 2 minutes and discuss as a table

Write down common issues on note cards
WORKSHOP QUESTION #2
What does an intelligent biobank look like to you?

Take 2 minutes and discuss as a table

Write down common issues on note cards
WORKSHOP QUESTION #3
What obstacles are preventing your success?

Take 2 minutes and discuss as a table

Write down common issues on note cards
Speaker Panel - Solutions Discussion
Audience - Q&A Session