Metadata Driven Automation in Drug Research

Hanming Tu, VP, Frontage Laboratories
John Lin, SVP, Frontage Laboratories

PhUSE Connect US 2019: Clinical Data Science Conference
Feb 24~27, 2019
Baltimore Marriott Waterfront Hotel
700 Aliceanna St, Baltimore, MD

phuse.eu/cs-working-groups.aspx
Agenda

• Introduction
• Standard Based Integration
• Metadata Driven Automation
• Efficiency Matrix
• Conclusion
Introduction: the Issues

- Adopting data and document standard is not just the trend but also the mandate
- Change is norm: standards and controlled terms
- More standards adopted, more data about the standard (metadata) are needed
- More intelligence needs to feed into the process to make it more efficient
Integration complexity:
  - No standard: exponential
  - With Standard: linear

<table>
<thead>
<tr>
<th></th>
<th>No standard</th>
<th>With standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems</td>
<td>3  4  5</td>
<td>6  7  8  9  10</td>
</tr>
<tr>
<td>No Std</td>
<td>3  6 10</td>
<td>15 21 28 37 46</td>
</tr>
<tr>
<td>With Std</td>
<td>3  4  5</td>
<td>6  7  8  9  10</td>
</tr>
</tbody>
</table>
Integration Issues

- **Data Issues:**
  - Diverse sources and in different formats
  - Different content / terms / variables

- **System Issues:**
  - Incompatible systems
  - Different means for data access

- **Organizational Issues**
  - Contract Research Organization (CRO): Multiple organizations
  - One company: Departmental standards
Two approaches
- Opportunistic approach is ad hoc way of finding or creating reusable codes
- Planned approach brings systematic code reuse.

Process of data integration represents in three steps: Extract, Transform and Load (ETL).

The ETL enables the collection of data from various sources into one data store or called data warehouse (DWH), ready for analysis.
## Comparison among ETL Tools

<table>
<thead>
<tr>
<th>Traditional Approach</th>
<th>OWB Approach</th>
<th>AutoDCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETL using custom programming such as SAS, PL/SQL, JAVA, Perl, etc.</td>
<td>ETL with User Interface</td>
<td>Web-based User Interface</td>
</tr>
<tr>
<td>High paid programmers</td>
<td>Users do not need to know the programming language – PL/SQL</td>
<td>No PL/SQL programming is needed</td>
</tr>
<tr>
<td>No audit trail</td>
<td>In an audited environment</td>
<td>In an audited environment and AutoDCD is validated product</td>
</tr>
<tr>
<td>No security</td>
<td>Built-in security: database and OWB security</td>
<td>Authenticated and authorized users only with audit trails</td>
</tr>
<tr>
<td>No consistency among coding</td>
<td>Consistency with all the users</td>
<td>Automatic and consistent coding</td>
</tr>
<tr>
<td>Difficult to manage and support</td>
<td>Easy to manage and support</td>
<td>Easy to manage and support</td>
</tr>
<tr>
<td>Scalability: silo and not scale; through adding more manpower</td>
<td>Scalability through hardware and software</td>
<td>Very scalable</td>
</tr>
</tbody>
</table>
Metadata is the key for reusability and automation
Metadata is the data that:
- Describes other data
- Provides information about other data

Data we collect:
- Preclinical studies: sample data
- Clinical research: subject and drug data

Data about the clinical research is critical as well.
Process repeatability is possible due to the metadata about the clinical process.

We could gain efficiency by building repeatable workflows.

A defined and repeatable workflow also ensure the quality of the work.
Conversion Overall Process

<table>
<thead>
<tr>
<th>Collection</th>
<th>Conversion</th>
<th>Validation / Load</th>
<th>Repository</th>
<th>Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDC</td>
<td>SDTM</td>
<td>Engine</td>
<td>Local CTR</td>
<td>Reports</td>
</tr>
<tr>
<td>Client Raw Data</td>
<td>Studies</td>
<td>CheckPoint</td>
<td>FDA CTR</td>
<td>OLAP jReview</td>
</tr>
<tr>
<td>Client SDTM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fuse EDC / CDMS | AutoDCD | CheckPoint/ OpenCDIDISC | Janus/CTR | jReview, OBIEE+ |
AutoDCD Data Flow Diagram

AutoDCD Users:
1. Upload Spec
2. Copy tables from a DB
3. Run data conversion jobs

Web:
- Use IE Browser
- Use SQL Developer

AutoDCD Database:
- SAS Service on a workstation
- OWB or other DBs

File server (drive C):
- PC SAS on each user's computer

QC Users:
- Develop SQL and Review Data

Computational Science Working Groups

Frontage

Phuse
AutoDCD Architecture and Features

- AutoDCD is a three tier web-based application: Oracle database, Oracle application server (Apache web server) and a browser

- Features:
  - Load and store map specification in relational database
  - Manage workspace with client, project, study, and specification hierarchy
  - Allow users to create and delete intermittent views and tables that we used in mapping
  - Run data conversion jobs by domain or by a group of domains or all
  - Link and copy tables from an Oracle database or use the tables created and loaded through SAS upload utility
  - Keep audit trails for each job
  - Track the performance of each job
Example in Clinical Research
Intelligent automation is a more advanced form of what is commonly known as robotic process automation (RPA) with contextual metadata.

The RPA is driven by predefined contextual metadata such as how to log into various systems, when to conduct pivot transformation, how to merge data from different domain, etc.

Intelligent automation brings fundamental changes to how drug research is conducted, how data is explored and how decision is made by individuals working with data.
Intelligent Automation: Conceptual Model

Conceptual Model:

Visual Data Exploration

Data

Element Classification

Transformation

Mapping

Visualization

User Interaction

Model Visualization

Model Building

Decision Making

Insight Reports

Models

AI

Modeling

Machine Learning

Data Mining

Metadata Collected and Intelligence Gained

Feedforward loop

Training data sets

Source: http://www.visual-analytics.eu/faq/
The conceptual model depicts how data are processed, classified, analyzed, modeled, visualized and used in gaining insightful knowledge and making intelligent decisions.

Four components:

- **Data Processing**: to make sure the data is clean, classified and ready to be used.
- **Machine learning**: to build meaningful algorithms or models such decision tree, deep neural network, convolutional neural network, recurrent neural network, etc.
- **Data Presenting**: to present data in different ways and different dimensions to allow us to understand the meanings and insights of the data and help us to make decisions.
- **Intelligence Building**: to feed more diverse data, collect more metadata, gain more insights and make better decisions.
Efficiency Matrix: Mix of All

- What is the efficiency matrix?
  - Standard-based systems allow for integration while metadata-driven systems enable automation.
  - The more intelligence collected about the clinical data, the more integration could be;
  - The more integration, the more metadata-driven automation is required;
  - The more standards are adopted, the more meaningful metadata could be applied in the process and the more efficient the process could become.

- How could we measure the level of the efficiency?
Efficiency Matrix of Intelligent Automation

Standard-based systems allow for integration while metadata-driven systems enable automation; As more intelligence is collected about the clinical data, a higher level of integration is achieved.

List of Efficient Levels:
1. No code reuse – double programming for every study
2. Code reuse at function level
3. Code reuse at module level and company code standard and reusable code library exists
4. Adopted standards in part of the process in silo systems
5. Adopted standards in some part of the process with some code reusability
6. Adopted standard enable the high reusability in the process
7. Adopted standard but have not build up metadata and no intelligence
8. Adopted standards and start learning the insight of the process
9. Adopted standard in all parts and with learned intelligence applied to the process
Efficiency Matrix: Key Factors

- **Standard adoption**: this could be the potential integration indicator

- **Code reusability**: to make it more efficient - efficiency indicator:

- **Process repeatability**: to make it more intelligent - intelligence indicator:
A key element of application-centric artificial intelligence is context.

Cutting-edge AI:
- Smart classification
- Smart recognition
- Smart predictions

Building a smart lab with various systems to acquire, process and analyze the lab data.
SMART Automation in Bioanalytics

➤ SMART Lab:
- Electronic lab notebook (ELNs): tools used for sample management
- Scientific Data Management System (SDMS): data acquisition and processing, reporting
- Laboratory information management systems (LIMS)
- Lab execution system (LES)

➤ SMART automation on existing robotic platforms
  - Simple to use;
  - Mindful of user errors;
  - Assay class automation;
  - Robust, rugged and reliable;
  - Tied into organization’s IT systems

➤ Bioanalytical wet chemistry kiosk (BWCK)

➤ Bioanalytical laboratory in a box (BLIB)

➤ Lab-on-a-chip technology
The premier community for people working in the biometric area

phuse.eu
@PhUSETwitta
/PhUSE
phusewiki.org