



PhUSE US Connect 2019

Paper ML10

Metadata Driven Automation in Drug Research

Hanming Tu, Frontage Laboratories, Inc., Exton, PA, USA
Zhongping (John) Lin, Frontage Laboratories, Inc., Exton, PA, USA

ABSTRACT

The industry has adopted CDISC standards for data and FDA eCTD standard for document for many years but still faces many challenges. The solutions based on these standards enable integration among solo systems but the integration needs to be based on business requirements and provides the end-to-end intelligence for the business. The more standards are adopted, the more meaningful and timely metadata are needed to manage the change of the standards and need to be applied in the process. Automation using artificial intelligence (AI) and machine learning will be the next game changer in the industry to provide the data with higher quality and more efficiency.

This paper will discuss the challenges in managing the standard adoption, the potential approaches for automation through utilizing robotic process, artificial intelligence, and the maturity model for metadata driven automation in clinical researches.

INTRODUCTION

With the standard adoption become a trend and a norm in life science industry, more pharmaceutical and service (CRO) companies demand more efficiency and higher quality from the standard based solution. Automation using artificial intelligence (AI) and machine learning will be the next game changer in the industry to provide the data with higher quality and more efficiency.

The industry has adopted Clinical Data Interchange Standards Consortium (CDISC) standards for data and FDA eCTD standard for document for many years but still faces many challenges. The solutions based on these standards enable integration among solo systems but the integration needs to be based on business requirements and provides the end-to-end intelligence for the business. The more standards are adopted, the more meaningful and timely metadata are needed to manage the change of the standards and need to be applied in the process. These would need more intelligence to feed into the process and to make the process more efficient.

This paper will discuss the challenges in managing the standard adoption, the potential approaches for automation through utilizing robotic process, artificial intelligence, and the maturity model for metadata driven automation in clinical researches.

STANDARD BASED INTEGRATION

Delivering quality data is the core of clinical trial and data management in drug research. The key for data quality is to adopt a common standard, stick to the standard practice and process, and use industry strength technology to reduce human errors. Adopting a standard not only increases the interoperability and efficiency but also provides a foundation for data integration and increases degree of code reusability.

SYSTEM INTEGRATION

Data are stored and processed in various systems. System integration is an engineering process of bringing together the component subsystems into one system to deliver the overarching functionality and ensuring that the subsystems

function together as one system¹. In the real world, system integration involves integrating existing, often disparate systems to increase the value or to reduce the cost to the customer or to improve product quality and performance.

The effective integration of data and knowledge from many disparate systems or sources will be crucial to future drug discovery. We often face the following 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

The following diagram shows with the number of systems increases the number of integrations needed are growing exponentially; when we have data standard, the number of integration needed is linear.

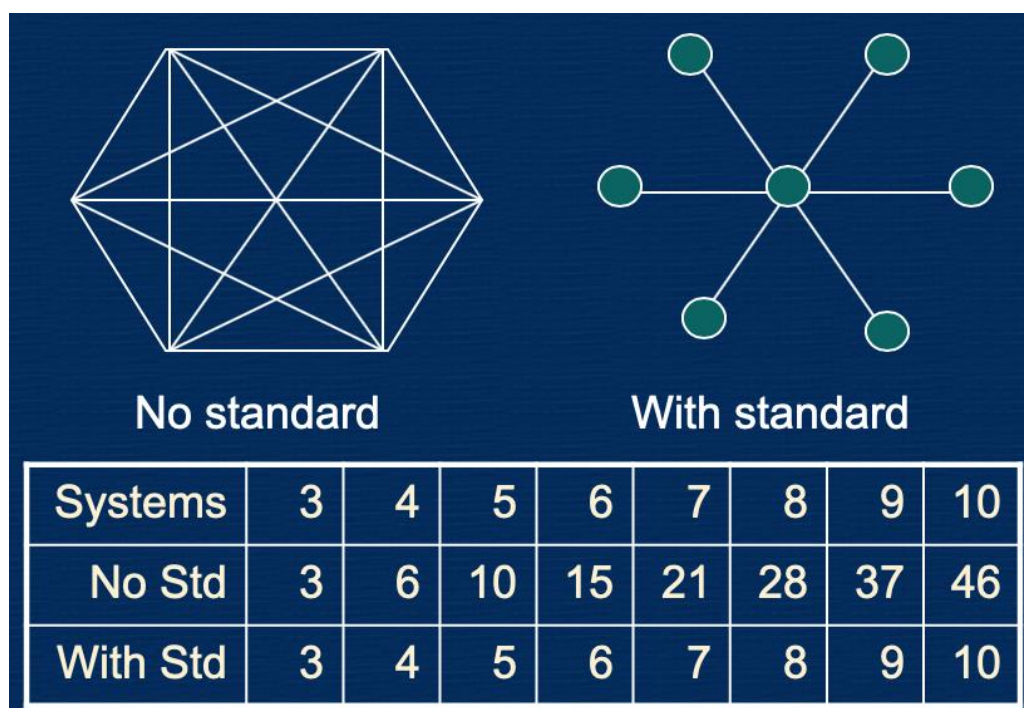


Figure 1: Comparison between Standard and No-Standard Integration

Data integration has to be based on standards. The Food and Drug Administration (FDA) in US has recommended Study Data Tabulation Model (SDTM) developed by the Clinical Data Interchange Standards Consortium (CDISC) as data standards for structure, terminology and code sets in its Federal Register, Volume 71, No. 237, December 11, 2006 and mandated all sponsors to have their study data in the data formats supported by FDA and listed in the FDA Data Standards Catalog if the studies started after Dec. 17, 2016². It took 10 years for FDA to mandate the data submission must be in SDTM format.

Adopting data standards not only reduce the difficulties of data integration from an exponential issue into a linear one but also has many practical benefits:

- Provides pre-built data models, structures and processes to standardize data to CDISC SDTM

¹ https://en.wikipedia.org/wiki/System_integration

² Providing Regulatory Submissions in Electronic Format — Standardized Study Data, December 2014 Electronic Submissions, <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf>.



- Eliminates data processing and transferring errors through pre-built and customizable standards adherence checking
- Make data build for internal analysis and regulatory submission at lower cost and higher quality
- Enables multiple users to access legacy data and perform real-time analysis of multiple/adaptive trials for multiple reports
- Automates migration of acquired data assets through data standards and make cross-study integration easier

Data integration is a key element of conducting scientific investigations with modern platform technologies, and it is a starting point for managing increasingly complex drug discovery and research, and fully realizing economies of scale in large enterprises. Such economies of scale relies on reusability of codes being developed in the process of data integration.

CODE REUSABILITY

There are different ways to build reusable codes. It could be opportunistic or planned. An opportunistic approach is ad hoc way of finding or creating reusable codes. A planned approach brings systematic code reuse. The codes developed for the datasets in one phase of a clinical study could be used for all the phases or from one study to be used for all the studies in the same therapeutic area. It is a strategy for increasing productivity and improving quality in the data transformation and standardization. Although it is simple in concept, successful code reuse implementation is difficult in practice.

The 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. The ETL could encompasses a cleaning step as a separate step. The sequence is then Extract-Clean-Transform-Load (ECTL). The ETL process seems straightforward and linear but it is quite complex and diverge. It often requires staging places to hold the intermediate data outputs so it could be restarted from where it fails. Many ETL tools were developed by technology companies and some of the well-known tools are IBM InfoSphere DataStage, Informatica, Oracle Data Integrator, SAP Data Integrator and Ab Initio. There are several open source ETL tools such as Apache Airflow, Apache Kafka, Apache NiFi, Apatar, CloverETL, Pentaho, Talend, etc. We can buy or adopt one of these tools or build your own ETL tools in your data integration. We use Oracle Warehouse Builder (OWB) and a custom built tool to compare these approaches.

Traditionally code reuse is done through leveraging the code libraries or adding custom codes into the libraries. Oracle Warehouse Builder (OWB) is a single, comprehensive tool for data integration and it provides data quality, data auditing, fully integrated relational and dimensional modeling, and full lifecycle management of data and metadata. It has simple graphic user interface to allow users to drag and drop lines to build up the source to target mapping. The design center in OWB is visual but take a lot of time and requires big screens to allow users to draw lines to link the sources and targets after we have built mapping specifications. After using OWB for three years, we built a much faster web-based system to automate the data conversion development (DCD). Here is the comparison among the traditional approach, OWB approach and AutoDCD approach:

Traditional Approach	OWB Approach	AutoDCD
ETL using custom programming such as SAS, PL/SQL, JAVA, Perl, etc.	ETL with User Interface	Web-based User Interface
High paid programmers	Users do not need to know the programming language – PL/SQL	No PL/SQL programming is needed
No audit trail	In an audited environment	In an audited environment and AutoDCD is validated product
No security	Built-in security: database and OWB security	Authenticated and authorized users only with audit trails



No consistency among coding	Consistency with all the users	Automatic and consistent coding
Difficult to manage and support	Easy to manage and support	Easy to manage and support
Scalability: silo and not scale; through adding more manpower	Scalability through hardware and software	Very scalable

Figure 2: Comparison among ETL Tools

Here are a few important considerations for increasing the code reusability:

- Standard adoption is the key for code reusability
 - Train people to understand the standards
 - Define standard templates
 - Build public libraries for code snippets and public transformation: Custom functions, procedures and packages; public data rules; and public Experts
 - Group code snippets and functional transformation into modular mapping and transformation: pluggable maps
 - Define workflow to govern the process: Workflow Manager and Process Flows
- Metadata-driven process is the key for automation
 - Metadata makes data meaningful
 - Metadata is machine readable
 - Metadata is the base for automation
- Replication and automation are the focuses
 - Use or create utilities to replicate the process: OMB+ for Project Set Up, Mapping Specification, Mapping Creation
 - Use analytics tool to identify the areas for replication and automation: Data Profiling & Data Rules for Source Data Review / Edit Checks

When we have identified the components that can be reused, we can follow the following steps to build reusable modules or even a base project which can be used to start a new project.

- Extract common components:
 - Build a public code library
 - Transformation
 - Utilities: functions, procedures, packages, pluggable maps, workflows
 - Build metadata repository:
 - SDTM data model
 - Controlled terminologies
 - Specification lookup tables: mapping intelligence
 - Create a base project
 - Common modules
 - Public locations (database links)
- Build subsequent projects
 - Create location linking to metadata repository
 - Import public utilities: transformation, data rules and experts
 - Copy the base project and modules

There are many ways that we can increase the code reusability of an ETL process in the data integration. However, viewing data integration as simply as data issue or technical issue underestimates the novel and serious scientific and management challenges it embodies - challenges that could require significant process reengineering, methodological and even cultural changes in our approach to data, particularly metadata.

METADATA DRIVEN AUTOMATION

DATA ABOUT DATA

The key of the reusability and automation is the metadata. Metadata is data that describes other data³ and that provides information about other data⁴. In clinical research, we focus on collecting subject and drug data and in preclinical studies, we focus on analyzing sample data but the metadata about the clinical research is critical as well. For instance, the purpose of the clinical study, the means of creation of the data, who created the data, when the data was created, in what type of system, the security of the system, the quality of the data, etc. According to the National Information Standard Organizations (NISO)⁵, metadata can be categorized into three main types:

- **Descriptive metadata:** describes a drug research for purposes such as drug discovery and identification for finding or understanding a protocol. It can include elements such as protocol title, inclusion/exclusion criteria, process of conducting a trial, keywords, etc.
- **Structural metadata:** indicates how analytical results are put together, what data or document standards are used, and how all the parts in the study are related to each other.
- **Administrative metadata:** provides information on the process of conducting a drug research and on managing a clinical trial. The information includes when and how the study was conducted, what type of data are collected, what format of file type and other technical information are used, and what analytical method is used, etc. There are several subsets of administrative data as follows:
 - **Rights management metadata**, which deals with intellectual property rights.
 - **Preservation metadata**, which contains information needed to archive and preserve a resource.
 - **Technical metadata**, which is used for decoding and rendering files or potentially is used for automation.

In the NISO primer paper, markup languages integrating metadata and flags for other structural or semantic features within content is also included as a type of metadata. The importance of collecting these type of metadata has been recognized by both the regulatory agencies and the industry not only for transparency, traceability and quality but also for building a foundation for automation. The role of metadata for Life Sciences & Healthcare has been changed⁶. It is not only increasing the desire for integration across verticals and patient-centric business processes but also driving the automation and collaboration within the drug research and cross sectors in the life sciences industry.

PROCESS REPEATABILITY

Another important aspect of the metadata driven automation is the process repeatability. We could gain efficiency by building repeatable workflows. A defined and repeatable workflow also ensure the quality of the work. The following diagram shows the overall process of data conversion. In reality, we have so many isolated systems and single purposed codes. How could we link reusable silo codes into a repeatable process?

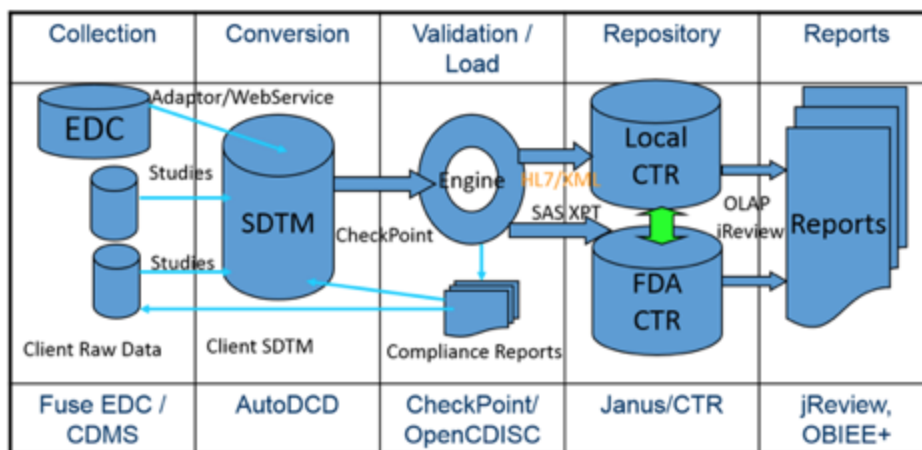


Figure 3: Conversion Overall Process

³ <https://whatis.techtarget.com/definition/metadata>

⁴ <https://en.wikipedia.org/wiki/Metadata>

⁵ "Understanding Metadata: What is Metadata and What is it for?" by Jenn Riley

(https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf) 2017

⁶ Healthcare, Life Sciences, September 30, 2015: <https://xtalks.com/webinars/metadata-for-life-sciences/>



In the AutoDCD project, we built an automatic data conversion system to link the pieces in data integration and standardization process. AutoDCD provides the workflow to link maps to form a controlled data flows, from vertical code reusability to horizontal process repeatability. The AutoDCD portal provides a web-based software tool for the Data Integration and Standardization (DIS) Department to use. The portal improves and accelerates the Data Conversion Development (DCD) process.

The input to AutoDCD is a Data Conversion Specification (DCS) that describes the source to target migration process. AutoDCD extracts and loads into Oracle the SQL statements from the DCS and sequences these statements into a "stored procedure". The 'stored procedure' is then executed to convert source data into domains compliant with the CDISC SDTM model.

AutoDCD is a three tier web-based application: Oracle database, Oracle application server (Apache web server) and a browser. Here is a list of features being implemented:

- 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

In AutoDCD design, a SAS service is used to integrate with SAS scripts including upload (import) and download (export) SAS macros. See the following data flow diagram.

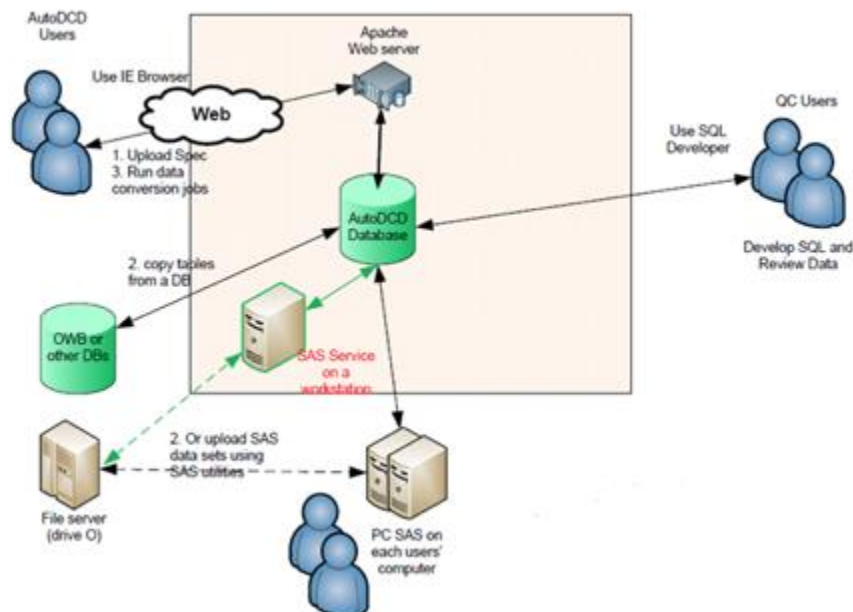


Figure 4: AutoDCD Data Flow Diagram

The AutoDCD project is an examples using source and target metadata to build reusable codes and drive repeatable process automation. This type of automation can be seen in the various phase of drug research. The following figure shows the possible ETL points in the clinical research.

Metadata Driven Automation in Clinical Research

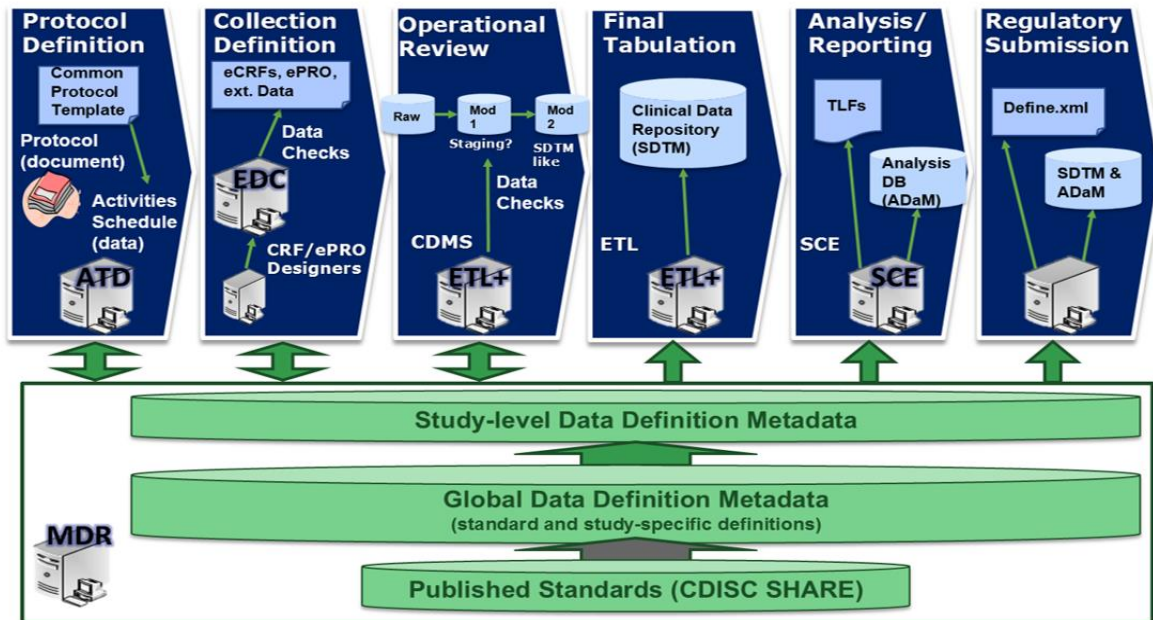


Figure 5: Metadata Driven Automation in Clinical Research

EFFICIENCY MATRIX

Standard adoption and metadata collection about the clinical study and the process of conducting the study enable the automation; the metadata drive automation not only improve data quality but also increase the efficiency. Collecting and classifying the study related metadata is the first step to build artificial intelligent about clinical study automation.

INTELLIGENT AUTOMATION

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. This type of operation may be overwhelming to end users but machines have different strengths and capabilities that complement their human supervisors. Together, they're changing what's possible.

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. The following diagram show the model for automation with intelligence.

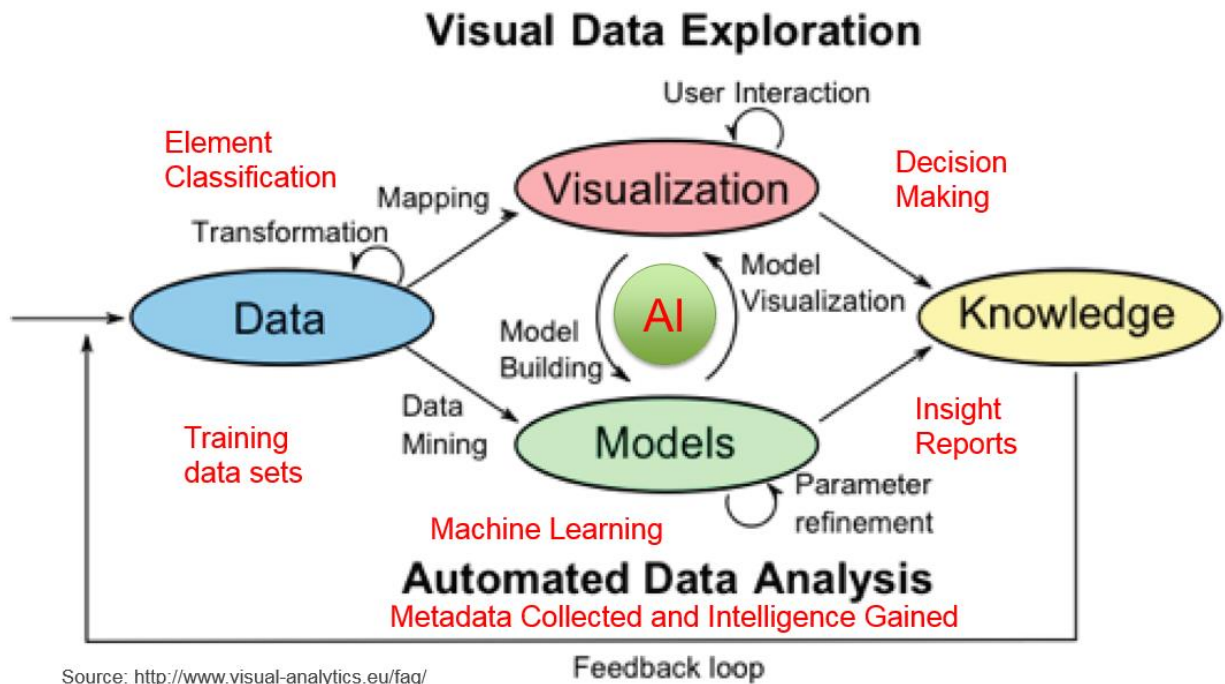


Figure 6: Conceptual Model of Automation with Intelligence

The conceptual model depicts how data are processed, classified, analyzed, modeled, visualized and used in gaining insightful knowledge and making intelligent decisions. It has the following components:

- **Data Processing:** to make sure the data is clean, classified and ready to be used.
 - Data are extracted from various sources, cleansed, mapped and transferred to a standard format or model, then loaded into a repository;
 - Data may be classified, transformed and aggregated for visualization and analysis;
 - Various training datasets may be built, mined and fed into various data models or algorithms. .
- **Machine learning:** to build meaningful algorithms or models such decision tree, deep neural network, convolutional neural network, recurrent neural network, etc.⁷
 - Use the training datasets to build a baseline of the model and to identify patterns using machine learning and computers by unearthing boundaries in datasets.
 - Evaluate the model with more real world data and by refining parameters.
 - Optimize the model with the highest-scoring classifier and refined algorithms to handle more outliers.
- **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 decision. It is a combination of data science, computer science and art of data visualization. Kelly Lauth at Microsoft created a term for it in 2007 - data presentation architecture (DPA) which is a skill-set that seeks to identify, locate, manipulate, format and present data in such a way as to optimally communicate meaning and proper knowledge. It requires much broader skill set than data visualization. Data visualization is just one element of DPA⁸. It has two main objectives:
 - To use data to provide knowledge in the most efficient manner possible (minimize noise, complexity, and unnecessary data or detail given each audience's needs and roles)
 - To use data to provide knowledge in the most effective manner possible (provide relevant, timely and complete data to each audience member in a clear and understandable manner that conveys important meaning, is actionable and can affect understanding, behavior and decisions)
- **Intelligence Building:** to feed more diverse data, collect more metadata, gain more insights and make better decision. It is a process and has the following characteristics:
 - It is a loop for automated data analysis and visual data exploration.
 - Artificial intelligence (AI) is in the heart of the whole data and process automation.
 - Once the loop being built, it should provide a productive and cost-effective human-computer

⁷ [What is Machine Learning?](#) December 21, 2018, published by Daniel Faggella.

⁸ Data Visualization: https://en.wikipedia.org/wiki/Data_visualization

interactive environment through optimization of four basic elements: data including metadata, models (standards), visualization, and knowledge, and the interrelationship between them.

MATRIX LEVELS

As it is shown in the efficiency level 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.

Further automation could be achieved through “Intelligent Data Flow”. The code reusability and process repeatability makes the data conversion very fast; the compliance check ensures the quality is good; the intelligent data flow makes the whole clinical data lifecycle smart! So it makes the overall project relatively cheap.

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.

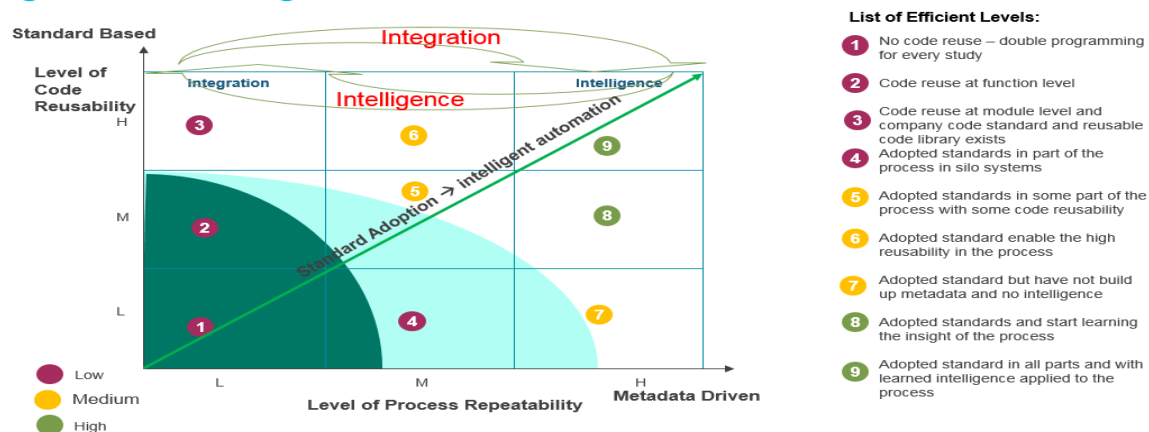


Figure 7: Efficiency Metrics of Intelligent Automation

There are 9 levels of automation efficiency based on the levels of standard adoption, code reusability and process repeatability:

- Level 1:
 - No common standard is adopted;
 - No codes are reused, double programming for every study;
 - It is difficult to reproduce the process in another study.
- Level 2:
 - Departmental standards exist in the company;
 - Small set of functions being developed and codes are reused at functional level;
 - Processes used in different studies are evaluated but process repeatability is not addressed.
- Level 3:
 - Coding standard and styles exist but no common data and process standard being adopted.
 - Common functions and macros are developed, code reuse at module level and companywide code standard and reusable code library exist;
 - Process repeatability is fragmented and very manual.
- Level 4:
 - Started educating people with standards; adopted standards in part of process in silo systems;
 - Still conduct double programming for every study with small percentage of codes being reused;



- Data flow in some phases of the study have been defined and about half of process can be repeated cross studies in the same therapeutic areas.
- Level 5:
 - A data standard coordinator is pointed and data standard is adopted in some parts of whole study life cycle;
 - Some reusable functions and modules are built and put into code library and are used cross studies;
 - Some common workflows are defined and used cross studies.
- Level 6:
 - Dedicated standard personnel has been hired to manage the standard adoption;
 - Most functions conducting data transformation, standardization and classification are put into shared common library and can be used for most of the studies;
 - Some common workflows are defined and used cross studies in multiple therapeutic areas.
- Level 7:
 - Standard governance body is formed and standard department is established to coordinate all the standard implementation in the company;
 - Double programming for every study is conducted with some code reusability; some algorithmic metadata have been collected but no metadata registration system (MDR) is built;
 - Collected metadata help to automate the most of the workflows being defined and used cross studies.
- Level 8:
 - Standard governance body is formed and standard department is established to coordinate all the standard implementation in the company; a MDR system is implemented;
 - Some reusable functions and modules are built and put into code library and are used cross studies; some visualization tools are used to explore the data;
 - Collected metadata help to automate the most of the workflows being defined and used cross studies.
- Level 9:
 - Standard adoption is championed by the senior executives in the company; the governance body is embedded in all parts of the process; training on standard adoption is conducted periodically; a metadata repository system is implemented to manage all versions of the standards, code libraries, workflow definitions and their usage in various studies;
 - 100% code reusability achieved through building modules, models, and services.
 - 100% process repeatability achieved through collecting metadata, using metadata to drive the whole process.

It may not be able to use quantitative measurements to clearly define each level but the following suggested key indicators might be helpful in evaluating a company's level in the efficiency matrix of intelligent automation:

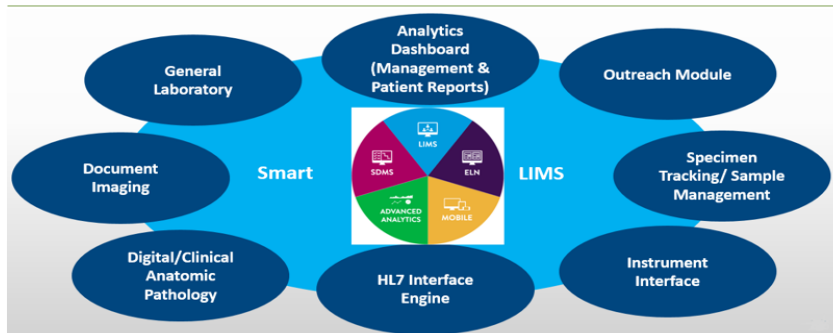
- **Standard adoption:** this could be the potential *integration indicator*: you can get three levels with number of positive answers to the five questions below:
 - Is there at least one dedicated resource for standard adoption in the company?
 - Is there periodical training to people involved in data processing and analyses?
 - Is there a permanent governance body for standard adoption and implementation?
 - Is there metadata registration system implemented to manage the standard adoption?
 - Is there any robotic process automation (RPA) and artificial intelligence (AI) technology being used to gain insights and help with decisions?

Note: Level 1 with 2 yeses, Level 2 with 4 yeses, and Level 3 with 5 yeses.
- **Code reusability:** to make it more efficient - *efficiency indicator*:
 - Is there a common function library?
 - Could the codes be modularized?
 - Are code metadata being collected?
 - Level 1: one yes from the above three questions with 30% code reusability
 - Level 2: two yeses with 30–60% code reusability
 - Level 3: three yeses with above 60% code reusability
- **Process repeatability:** to make it more intelligent - *intelligence indicator*:
 - Is there defined workflow?
 - Could workflow being group and classified?
 - Are the process metadata being collected?
 - Level 1: one yes from the above three questions with 30% process repeatability
 - Level 2: two yeses with 30–60% process repeatability
 - Level 3: three yeses with above 60% process repeatability

AUTOMATION TESTS

A key element of application-centric artificial intelligence is context. As Zeichick in a recent article pointed out⁹: smart classification, smart recognition, and smart predictions are three big buckets that encompass many cutting-edge AI and machine learning capabilities.

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

Figure 8: SMART Automation in Bio-analytics

Many contract research organizations (CRO) like Frontage laboratories have turned to technologies to build a smart lab with various systems implemented to acquire, process and analyze the lab data. The commonly deployed systems are

- **Laboratory Information System (LIS)** for processing and reporting data related to individual patients in a clinical setting.
- **Laboratory Information Management System (LIMS)** for process and report data related to batches of samples from biology labs, water treatment facilities, drug trials, and other entities that handle complex batches of data. It is a repository of structured data.
- **Electronic Laboratory Notebook (ELN)**: for documenting lab research work with electronic notes that are searchable, shareable and having safeguards of security and backup.
- **Lab Execution System (LES)**: for directing users to follow specific steps to ensure the rigidity of the test method and the process' end results, including tasks like enforcing standard operating procedures (SOPs), validating calculations and instrument interfaces, and acquiring or importing procedural data from other systems into one common system. It is for quality control, quality assurance and compliance in laboratories.
- **Scientific Data Management System (SDMS)**: for acting as a document management system (DMS), capturing, cataloging, and archiving data generated by laboratory instruments (HPLC, mass spectrometry)

⁹ Want A Bigger Bang From AI? Embed It Into Your Apps by Alan Zeichick
(<https://www.forbes.com/sites/oracle/2018/11/27/want-a-bigger-bang-from-ai-embed-it-into-your-apps/#66d5384d4e2d>)



and applications (LIMS, ELN) in a validated, compliant, secure environment. It is a repository for handling structured, unstructured data, mostly heterogeneous data; a seamless wrapper for other data systems like LIMS and ELN in the laboratory; a gatekeeper, serving platform- independent data to informatics applications and/or other consumers¹⁰.

These systems are designed to gather, store, and analyze large volumes of data, and some drug discovery units in particular have taken the lead in automating and robotizing their research laboratories. It may take a long time for a fully automated lab being developed but it is possible to have micro-robots in a chip to do very complicated experiments and analysis using microelectromechanical system [MEMS] techniques, according to Rodney Brooks, director of the Artificial Intelligence Laboratory at MIT¹¹. There are some existing smart kiosk or box being developed and in use such as bioanalytical wet chemistry kiosk (BWCK), bioanalytical wet chemistry in a box (BWCIB), and the bioanalytical laboratory in a box (BLIB)¹².

In general, lab automation refers to the use of technology to streamline or substitute manual manipulation of equipment and processes. This field of lab work comprises different automated lab instruments, devices, software algorithms and methodologies used to enable, expedite and increase of efficiency and effectiveness of scientific research and test analysis. The bioanalytical wet-chemistry automation on existing robotic platforms is defined as the Simple to use; Mindful of user errors; Assay class automation; Robust, rugged and reliable; Tied into organization's IT systems (SMART)^{13 14}. More importantly have these systems run in a validated environment and in compliant with 21 CFR Part 11 and good laboratory practice (GLP)¹⁵. As pointed out by Dr. Simon Bungers, smart lab in the future will be made of flexible, digital integration, automation and robotics, integrated functional surfaces and modular systems¹⁶.

CONCLUSION

CDISC SDTM is stable and recommended data model for interchange between sponsors and FDA. Standard based approach enables code reusability and process repeatability to gain greater efficiency and consistency. ETL tool like AutoDCD provides security, scalability and audited environment. High data quality relies on standard process, mature technologies, and trained people. The conceptual model of automation with intelligence shows the key components and their relationships and how data become knowledge and insight through collecting metadata, testing algorithm and building intelligence. The 9 levels of efficiency metrics provide a framework of measuring the intelligent automation. The metadata drive automation with artificial intelligence creates SMART labs.

¹⁰ Scientific Data Management System https://www.limswiki.org/index.php/Scientific_data_management_system

¹¹ Laboratory Technology Trends, <http://www.sciencemag.org/site/products/robotfinal.xhtml>

¹² Automation in the bioanalytical laboratory: what is the future? By Ming Li, Bioanalysis (2013) 5(23), 2859–2861

¹³ Li M, Chou J, Jing J et al. MARS: bringing the automation of small-molecule bioanalytical sample preparations to a new frontier. Bioanalysis, 4(11), 1311–1326 (2012).

¹⁴ Li M, Chou J, King K, Yang L. ASPECTS: an automation-assisted SPE method development system. Bioanalysis, 5(13), 1661–1676 (2013)

¹⁵ Zhongping (John) Lin et al, Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations, Edited by Wenkui Li et al, Wiley, 2013 ISBN 978-1-118-15924-8; Chapter 10: Current Understanding of Bioanalysis Data Management and Trend of Regulations on Data Management

¹⁶ "smartLAB 2017: For Tomorrow's Intelligent Laboratories", Simon Bungers, <https://www.technologynetworks.com/tn/news/smartlab-2017-for-tomorrows-intelligent-laboratories-209685>.



CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Hanming Tu
Company: Frontage Lab
Address: 700 Pennsylvania Drive
City / Postcode: Exton, PA 19341
Work Phone: 484-202-6479
Fax: 610-232-0101
Email: htu@frontagelab.com
Web: <http://www.frontagelab.com>

Author Name: Zhongping (John) Lin
Company: Frontage Lab
Address: 700 Pennsylvania Drive
City / Postcode: Exton, PA 19341
Work Phone: 484-248-4821
Fax: 610-232-0101
Email: jlin@frontagelab.com
Web: <http://www.frontagelab.com>