Clinical Data Standards in the Era of AI, ML and Digital Transformation

The need of the hour is to reach patients faster. By reducing both the drug development timeline and the overall cost of pharmaceutical research and development, we can achieve just that. COVID-19 has been a good example of just what can be achieved. Vaccines were developed within a year; pre-COVID, this was unheard of, with vaccines typically taking many years to develop.

For biometrics teams, this means that we need to be quick and accurate in data collection, processing and analysis. In order to cut short the timeline and improve efficiency, we need to automate many steps involved in the clinical data life cycle (planning phase, data collection, tabulation, statistical analysis, and exchange/sharing of data) and, in order to automate, we must have consistent metadata, standards, and technology.

This article highlights how clinical trial data standards have evolved so far, and how they are continuing to evolve in the era of artificial intelligence (AI), machine learning (ML) and digital transformation to meet future demands.

COVID-19 and the Wave of Digital Transformation
We can all agree that COVID-19 has acted as something of a catalyst, encouraging increased innovation, technology adoption, and a willingness to embrace digital transformation. As a result, we are witnessing a "new normal" – including an increased number of virtual trials being successfully designed, a move away from the more conventional clinical trials, in order to manage the global pandemic situation and ensure the continuation of clinical trials for many other, potentially life-saving, drugs. Virtual trials have helped patient recruitment, retention, real-time access to data, and better quality.

Digital data collection methodologies (mobile technology, wearables, electronic patient-reported outcomes (ePRO), electronic clinical outcome assessment (eCOA) etc.) have been instrumental, acting as game-changers to enable robust data capture in the era of COVID-19.

Why Data Standards?
As clinical research becomes increasingly complex, the opportunity to bring clarity to the data is more important than ever. A true measure of the data is the impact it has.

Science comes to life through data. Data doesn't mean anything if you have to struggle to understand where it is located, how it is organised, or how to analyse it. One cannot harmonise anything or combine without standards.

Standardisation helps in data aggregation, accessibility, interoperability, reusability and traceability. Ultimately, standardisation helps regulators to focus on their scientific review and make patient-centric decisions.

We need end-to-end data standardisation and integration strategy that considers all the dimensions of clinical data.

CDISC (Clinical Data Interchange Standards Consortium): There are many standards development organisations. When it comes to clinical trial data standards, CDISC has contributed significantly over the last two decades.

CDISC has played a significant role over the last two decades to achieve data quality (in order for us to trust the data to make credible and significant scientifically valid decisions) and also gain efficiency across clinical trial data life cycle processing.

Originally formed in 1997 as a volunteer organisation, CDISC has brought together experts in the industry to align on common data structures and data content spanning both nonclinical (animal) and clinical (human) studies.

CDISC standards are widely used across the biopharma industry and have become a requirement for data submissions to many health authorities. Some of the key CDISC standards are:

Foundational Standards – CDISC Foundational Standards are the basis of a complete suite of data standards, enhancing the quality, efficiency, and cost-effectiveness of clinical research processes from beginning to end, i.e. Protocol Representation Model (PRM), Standard for Exchange of Nonclinical Data (SEND), Clinical Data Acquisition Standards Harmonization (CDASH), Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM) and Questionnaires, Ratings and Scales (QRS).

Data Exchange Standards facilitate the sharing of structured data across different information systems, i.e. Clinical Trial Registry (CTR)-XML, Operational Data Model (ODM)-XML, Study/Trial Design Model in XML (SDM-XML), Define-XML, Dataset-XML and Resource Description Framework (RDF, provides executable, machine-readable CDISC standards from CDISC Library).

CDISC Controlled Terminology (CT) is the set of CDISC-developed or CDISC-adopted standard expressions (values) used with data items within the Foundational Standards and Therapeutic Area User Guides. CDISC Terminology provides context, content, and meaning to clinical research data and provides a consistent semantic layer across all operational contexts, enabling interoperability of the CDISC Standards.

Therapeutic Area User Guides (TAUGs) extend the Foundational Standards to represent data that pertains to specific disease areas. TAUGs include disease-specific metadata, examples, and guidance on implementing CDISC standards for a variety of uses, including global regulatory submissions. Therapeutic Area (TA) expertise becomes more and more important for clinical data scientists working in the pharmaceutical industry as it is crucial for the understanding of patients' needs and the interpretation of analysed data.

Originally focusing on common data domains in clinical trials (e.g. demographic information, adverse events, routine lab results, and subject status), CDISC has grown from PDFs to machine-readable standards, and from few safety domains to more Therapeutic Area specific standards (a great example is COVID-19 guidelines).

CDISC was designed to have built-in quality starting much early in the process. At this point, CDISC continues to evolve in evaluating a new source of data, referred to as "real-world data" (RWD), which includes data such as electronic medical records, insurance claims data, and wearable devices.

The Road Ahead:
Since the formation of CDISC, significant progress has been made in standardising the format of data collected, analysed and submitted to the health authorities. However, one of the challenges faced was that CDISC foundational standards were built in a two-dimensional model, with no significant CDISC support for the automation of foundational standards in the research enterprise.

One of the strategic goals for CDISC to address in the coming years is to "develop multidimensional standards in an open, transparent manner that allows community members to transition with as little disruption to their research as possible while unlocking greater benefits of standardisation. Engage in concrete steps to achieve end-to-end standardisation." Some of the initiatives in place to achieve this are:

• CDISC 360 (to demonstrate the feasibility of standards-based metadata-driven automation across the end-to-end clinical research data life cycle)
• Evolve the expression of foundational conformance rules to an electronic format to increase consistency and instantiate multidimensional model artifacts in the CDISC Library
• Initiate a process to build the model for machines first, people second
• Commit to develop only end-to-end TAUGs
    Another key strategic goal of CDISC is to expand and identify adjacent research areas that can benefit from data standardisation, i.e. with the evolution of the model, selectively extend CDISC standards to support new data types and/or new technologies.
• Expertise in RWD / real-world evidence (RWE)
• Consumer wearables
• Medical devices
• Augment/replace patient-reported outcomes data from consumer wearables and/or medical devices
• Device registry, likely via collaboration, that uniquely identifies devices and enables automated mappings to CDISC standards
• CDISC-compliant registry toolkit that is built on the CDISC Library API
• 'Mapping registry', which standardises the conversion of proprietary device data to CDISC standards

RWD has the potential to provide answers to important questions. RWD may come from multiple sources, including electronic health records (EHRs), medical claims and billing activities, product and disease registries, patient-generated data (including in home-use settings) as well as data gathered from sources that can inform on health status, such as mobile devices. RWE is the clinical evidence regarding the usage and potential benefits, or risks of a medical product derived from the analysis of RWD. To fully understand the patient experience, one needs to access the extended RWD/RWE (broader sources of data) in addition to EHR (electronic health record).

We must figure out how to align previously existing data into standards and how to make sure impending data is collected in a standardised way in order to future-proof. Standardisation helps with the integration of RWD into the drug development processes and patient safety monitoring.

Global regulators, such as the US Food and Drug Administration (FDA), Japan's Pharmaceutical and Medical Devices Agency (PMDA), the European Medicines Agency (EMA), India's Central Drugs Standard Control Organisation(CDSCO) and China's National Medical Products Agency (NMPA) are increasingly interested in leveraging the potential of RWD to complement randomised controlled trials by providing insights into efficacy, safety and post-market surveillance as a means of supporting regulatory decision-making across the product life cycle. Indeed, the US FDA is accepting observational data to support efficacy determinations, and the EMA is assessing the use of registry data for rare diseases.

HL7 and CDISC – CDISC is focused on clinical trial data standardisation, whereas Health Level 7 (HL7) standards focus on standards utilised in real-life healthcare services. With the advent of RWE or the inclusion of electronic health records in clinical trials, these two worlds now merge together. With the publication

EXTENDED REAL - WORLD DATA

Clinical · Claims · Billing · Financial · Omics · Clinical Specialty · Enterprise Resource Planning/ Human Resources · Patient Satisfaction · Health Information Exchange

Source: Healthcatalyst

complete traceability, enhanced innovation, improved data quality, facilitated data sharing, reduced costs, increased predictability, and streamlined processes.

Leveraging RWD CDISC, and other established and emerging standards (e.g., HL7-FHIR), have the potential to transform the healthcare industry and deliver a holistic, interoperable future state, which will foster greater efficiencies across systems and resources, and encourage end users to support higher-quality data exchange integrations both within and outside of research.

It is well known that because RWD is not collected with research as its primary purpose, there are significant challenges in using and representing the data. These challenges include bias, data variability and heterogeneity, which can make analysis of RWD difficult and resource-consuming. However, the benefits of connecting RWD to CDISC standards (i.e. improvements in data sharing, cross-study analysis and meta-analysis of data for all clinical researchers) may outweigh the challenges if efficiencies achieved expedite global regulatory reviews, contribute to the evaluation of new treatments for patients, and drive next generation discovery.

The next generation CDISC Library will provide machine-readable standards and support data becoming accessible across geographies and disciplines, interoperable across systems and studies, and reusable for research today and tomorrow. FHIR to CDISC project will streamline the flow of data from EHRs to CDISC submission-ready datasets.

We must think of developing innovations beyond our current boundaries. Innovation is driven by curiosity, often sparked by a curious question, and it's about having a creative mindset. There may be a long way to go, and the journey is likely to be challenging, but it is equally exciting.

REFERENCES

1. PhUSE (https://phuse.global/Education)
2. CDISC
a. https://www.cdisc.org/
b. https://www.cdisc.org/sites/default/files/resource/CDISC_2019_2022_Strategic_Plan.pdf

of the Fast Healthcare Interoperability Resources (FHIR) draft standard, a paradigm change was introduced. The alignment between terminologies is a known problem and groups such as the CDISC EHR to CDASH (E2C) are looking to come up with a shared semantic layer to bridge the gaps between the data standards.

Latest Developments:
CDISC and Microsoft come together to plan the next generation CDISC Library: "The next generation CDISC Library will provide machine-readable standards, standardized mapping to other data standards, and serve as a tool for community curation of standards that support data becoming accessible across geographies and disciplines, interoperable across systems and studies, and reusable for research today and tomorrow."

CDISC Partners with Gevity to Facilitate Use of Electronic Health Record Data in Clinical Research: The FHIR to CDISC project will leverage Fast Healthcare Interoperability Resources (FHIR), HL7's standard for exchanging healthcare information electronically and CDISC's standards for data collection (CDASH) and data tabulation (SDTM) to streamline the flow of data from EHRs to CDISC submission-ready datasets. The mapping and Implementation Guide will be available via the CDISC Library API (source of CDISC standards metadata).

Conclusion
As previously discussed, CDISC has played a significant role over the last two decades in achieving data quality that ensures we can trust data and make credible and significant scientifically validated decisions while also gaining efficiencies across the clinical trial data life cycle process.

CDISC has grown from PDFs to machine-readable standards and from few safety domains to more therapeutic area specific standards. CDISC is helping the entire field of clinical research tap into and amplify its full value. Key benefits of standards are fostered efficiency,

Shrishaila Patil

Shrishaila has done Masters in Biotechnology from Bangalore University & have more than 16 years of experience across Drug Development. He is currently working as Vice President at Navitas Data Sciences, the global Functional Service Provider of Navitas Life Sciences, heading the Statistical Programming Department, India. He is also working as "CDISC Volunteer" & "PhUSE India Membership" officer. He is also supporting "R Package Validation Framework" and "Open Source Technologies for Regulatory Submissions" Projects in PhUSE working group "Data Visualisation and Open Source Technology in Clinical Research (DVOST)". He has authored an International book "FDA Clinical Outcome assessments and CDISC QRS supplements" under "Clinical disciplines" category with LAMBERT Academic Publishing group. He has exposure to various Analytical tools like SAS (Base and Advance Certified), R, PYTHON and CDMS Tools like Inform (EDC), Medidata Rave (EDC), Clintrial and Oracle Clinical LSH. He is a passionate Speaker & active in most of the conferences in the industry. His hobbies are Reading, playing Table tennis etc.