Data Standards White Paper

phuse.global

Contents

Revision History

Version	Date	Summary
1.0	2023-April-04	

1. Purpose of this White Paper

The biopharmaceutical industry strives to bring novel medicines and vaccines to patients faster to improve and extend lives. Data standards increase efficiency through 1) providing common definitions, syntax, and semantics from end to end (E2E) 2) defining the data described in a clinical protocol 3) how the data is collected, analysed and submitted to regulatory agencies. More data standardisation enables more automation and more data reuse for generating additional insights and optimising future clinical research. The digitalisation of healthcare is also driving evolution. Biopharma companies, regulatory agencies, academic institutes, technology vendors and standards development organisations must adapt to this new environment. Hence, a significant challenge for the data standards community is keeping pace with the sheer breadth and depth of disease areas for which therapies are being developed across the biopharmaceutical industry, and the myriad of data types (including biomarker, genomic and imaging) being collected.

This white paper will highlight current challenges in data standardisation across the biopharma industry and identify opportunities where we can work together to tackle them.

2. (i) Demand for More Standardisation

As our understanding of disease biology deepens, and thanks to advances in technology, the volume and complexity of data collected in clinical trials continues to increase. As part of the review and evaluation of new medications/vaccines, health authorities around the world are increasingly mandating for the submission of clinical trial and real-world data (RWD) in a standardised format. The availability of new and modified standards from standards development organisations (SDOs) and emerging technologies and digitisation are driving new ways of working. New trial designs, decentralised clinical trials (DCTs), data types including biomarker, genomic and imaging data and data sources such as eSource and electronic health records (eHRs) within clinical trials are essential for medicine and vaccine development and for changing the way we define and implement standards.

Readiness for Multiple Health Authority Requirements

Data standards teams have become well versed in submission requirements across multiple regulatory agencies and maintaining multiple versions of standards based on each agency's requirements. It is imperative the teams not only consider the CDISC Implementation Guides (IGs) and Therapeutic Area User Guides (TAUGs) but also include the regulatory agencies' requirements and guidelines as drivers in standards development.

Wearables, Decentralised Trials and Remote Patient Monitoring

In 2019 to 2020, it was estimated that approximately 29% of the US adult population uses a smart wearable device. Smartwatch and fitness tracker shipments continue to grow and are expected to reach 280 million units to be shipped in 2024.1 This massive volume of data brings major opportunities for healthcare, but it also comes with specific challenges such as the huge size of the datasets, data being generated in real time, the variety of formats, and the lack of standards and development of methodologies for summarisation of data and extracting insights from large data volumes.

The number of decentralised clinical trials (DCTs) has grown exponentially due to the COVID-19 pandemic. The need to reach patients outside of the healthcare unit during the pandemic has accelerated the adoption of solutions by the industry for trial continuity and data integrity. The adoption of DCT technologies will increase the number of data sources on a trial but also increase the complexity, as the scenario where data may be collected via EDC platforms or DCT technologies – depending on a patient's wish to go to a site or not – means that aggregation of data and monitoring for completeness of data will also grow in complexity and effort.

Evolution is two-fold: first in the collection of the data through electronic clinical outcome assessments (eCOAs) such as blood pressure monitors and glucose trackers or electronic patient-reported outcomes (ePROs) such as electronic patient questionnaires, and second in the development of solutions to support clinical trial regulatory requirements such as electronic consent and remote patient monitoring (RPM). New clinical trial operations and technologies bring huge opportunities to expand the diversity of the clinical trial population and new challenges in data collection, storage, analysis, privacy and security.

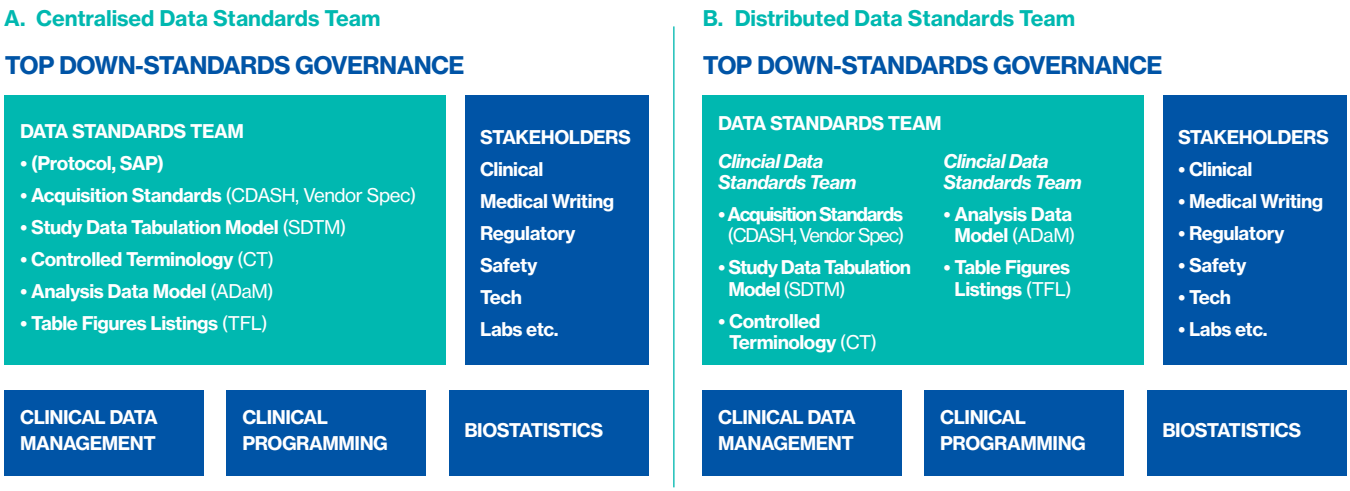
2. (ii) Implementation Challenges

Data Standards Operating Model

In surveying the pharma data standards community, we found most companies have created teams that are either centralised or distributed. Five out of ten companies have a centralised data standards team. Four out of ten companies have distributed standards teams, e.g. data acquisition standards teams and data analysis standards teams located within their data management and programming/biostatistics functions respectively and work together. One out of ten companies does not have a centralised team. In many companies, the scope spans data acquisition, data tabulation, data analysis and controlled terminology. In most cases, protocol, and statistical analysis plan (SAP), templates are owned by appropriate business units, typically clinical/medical writing (protocol) and biostatistics (SAP), which the data standards team review. One company has a Content Reuse and Automation office, which owns the protocol and SAP
templates (Figure 1).

Regardless of the structure, setting a clear data standards roadmap and prioritisation is instrumental to success. Embedding strong and efficient data standards governance requires experts who have a deep understanding of the complexity of historical and evolving data standards, of engaging and monitoring SDO activities and of collaborating with peers and stakeholders across the industry.

Figure 1: Centralised Data Standards Team vs Distributed Data Standards Team



Metadata Repositories (MDRs)

Most companies have invested significant resources in the implementation of a metadata repository (MDR) to manage the creation, update, versioning and linkage of data standards from end to end. These solutions should be capable of uploading historical metadata and automating the maintenance and relationships of our E2E data standards. They should link data elements that can trigger awareness to changes, be capable of managing multiple versions as new industry standards become available, and be integrated with other systems including digital protocol and clinical data management systems (CDMS, statistical computing environments (SCEs)). However, there have been significant challenges in implementing and adopting proper MDR solutions which demonstrate a return on investment (ROI). Via the discussions within our community, we have recognised the MDR vision has not been realised yet. To accelerate automation, there are now lots of considerations for the use of AI/ML in the data standards space, and metadata linked via biomedical/analytical concepts are even more increasingly desired. Robust and collaborative ways to develop and curate data standards content are also key to CDISC 360, CDISC Open-Source Alliance (COSA), and other ongoing activities, but these are not yet mature enough to implement.2, 3

3. Looking to the Future

There are various global collaborations, initiatives and SDOs, and collaboration is increasing with CDISC, TransCelerate, ICH M11, PHUSE, Vulcan HL7 FHIR and Observational Health Data Sciences and Informatics (OHDSI) to further leverage existing concepts and develop newer concepts which will start to address some of the above challenges. TransCelerate BioPharma's Digital Data Flow (DDF) is collaborating with CDISC to develop the Unified Study Definition Model (USDM) to provide a standardised way to represent study-level metadata.4–7

Like the emergence of an FDA mandate almost a decade ago and the requirement for sponsor companies to submit clinical trial data in CDISC's SDTM and ADaM format, other global regulatory agencies are mandating adoption of industry standards to accelerate medicine/vaccine development for patients. A similar mandate setting forth the requirement to use CDISC, Vulcan HL7 FHIR and OHDSI for certain use cases would mobilise the industry.8–10 Currently, the interactions between these SDOs are quite limited, focused on specific mapping efforts. AI/ML technology is evolving for accelerating automation and might support transforming the data between different standards models. While this is a start, the sheer scale of the challenge demands a complete paradigm shift in the way data standards and terminologies are developed in our industry.

The following calls to action are intended to offer tangible steps to make such a shift:

Global Collaboration

A strong collaboration is needed between SDOs, regulators and biopharma representatives to:

- Align on a standards-agnostic way to represent biomedical concepts in both human and machine-readable format which can be applied to current and future biomedical concepts used across clinical research and clinical healthcare.
- Conceptualise standards and solutions from end to end, which includes but is not limited to trial design, data collection, data tabulation, data analysis and reporting.
- Use the most accurate terminology available to describe biomedical data at its inception and carry this terminology through with the data for its entire life cycle, even if additional terminology is applied downstream that can pool or harmonise the data.

Global Biopharma Company and Health Authority Commitment

It is essential for global biopharma companies to have a common vision and commitment to accelerate the development of novel data standards. This commitment may require experts (resource commitment), partnerships, investing in proof of concept, co-developing tools/technologies with tech companies, etc. A unified voice regarding high-level requirements will enable our tech vendor partners to develop fit-for-purpose tools/technologies and provide efficiency gains. For the biopharma industry to have the most impact on patients and society, and as more health authorities (HAs) digitise their regulatory submission processes, it is imperative HAs work with the industry to understand pain points and align their regulations in terms of data standards.

In summary, the biopharma industry is not unique in terms of the opportunities presented by digital transformation and the increased focus on data. The breadth and depth of data available to biopharma companies will continue to grow both within and beyond the scope of randomised clinical trials while in pursuit of breakthrough therapies for patients. Putting fundamental steps in place to accurately define this data, starting at the biomedical concept level as a foundation for developing data standards and terminologies, will continue to be extremely challenging. However, this remains essential for achieving seamless data interoperability across the clinical research and healthcare ecosystem, maximising the use and reuse of data, automating data transformation and analysis, and fully realising the vision of a data-driven relationship between biopharma companies and health authorities around the world.

4. About the Authors

This white paper was authored by members of the PHUSE data standards community, comprised of leaders and experts responsible for the governance and implementation of data standards across the biopharmaceutical industry. There were many common themes during our journey that encompassed the governance and implementation of CDISC standards, and we have explored how best to manage our end-to-end data standards. We are an industry of talented and dedicated professionals, and we are confident we can collectively pool our resources to do something truly amazing.

References

1. Dhingra, L. S., Aminorroaya, A., Oikonomou, E. K. et al. Use of Wearable Devices in Individuals With or at Risk for Cardiovascular Disease in the US, 2019 to 2020. JAMA Netw Open, 6(6):e2316634. doi: 10.1001/jamanetworkopen.2023.16634
2. COSA|CDISC
3. OAK|CDISC
4. Vulcan FHIR Accelerator. https://hl7vulcan.org/
5. Data Standardization – OHDSI. www.ohdsi.org/data-standardization/
6. ICH. ich.org
7. TransCelerate – Digital Data Flow Asset – Clinical Trial Process. transceleratebiopharmainc.com
8. European Health Data Space – European Commission. europa.eu
9. DARWIN EU. http://www.darwin-eu.org
10. European Medicines Regulatory Network Data Standardisation Strategy. https://www.ema.europa.eu/en/documents/other/european-medicines-regulatory-network-data-standardisation-strategy_en.pdf