保健医療科学　2019　Vol.68 No.3 p.194－201

Topics: Toward introduction of CDISC to clinical research and public health

＜ Review ＞

Clinical data standards and the new world of research science, technology, and data sources

David R. Bobbitt [1)], Bess LeRoy [1)], Amy Palmer [1)], Mike Hamidi [1)], Rhonda Facile [1)], Satoshi Ueno [2)], Sam Hume [1)], Peter Van Reusel [1)], Jon Neville [1)]

[1)] Clinical Data Interchange Standards Consortium (CDISC)
[2)] Center for Public Health Informatics, National Institute of Public Health (NIPH)

Abstract

More than 20 years ago, data scientists, statisticians, and researchers met to develop the data standards to facilitate the submission of clinical research results to the regulatory authorities; it is the beginning of the Clinical Data Interchange Standards Consortium (CDISC). The data to prove the safety of medicines and medical devices has wide-ranged requirements concerning the origin of data, methods of data collection, data tabulation method, consistency of contents. So CDISC has various standards such as SEND, CDASH, SDTM, ADaM, and ODM to support clinical research protocol. As a result of collaborating with other researcher groups and maintaining the standard, it has been endorsed by regulators around the world and has become the global standard for clinical research. The Pharmaceuticals and Medical Devices Integration Agency (PMDA) in Japan, and the U.S. Food and Drug Administration (FDA) in the United States make the obligation to the submission of clinical trial data with CDISC standards. Furthermore, in recent years, the CDISC Library in machine-readable format has been released so that the developers can automatically generate programs and data conforming to the CDISC standard, and implementation trials with CDISC Library are being conducted by volunteers of the CDISC 360 project. The scope of CDISC is extended to general clinical research now. Therapeutic Area User Guides (TAUGs) covers clinical research on tumors, vascular diseases, neurological diseases, infections etc. The National Cancer Institute (NCI), the world's largest funding agency, built the Cancer Data Research Commons (CDRC), the platform for sharing data submitted by researchers. The stored data is required to comply with CDISC standards. As the CDISC standard is the comprehensive standard for data quality control and research management, it is infiltrating all area related to medical research that develop new, safe, and effective medical devices and treatment methods. CDISC may contribute to further acceleration to research and development. Real Word Data (RWD) also tends to have low quality, but it is expected that quality will be improved by incorporating CDISC standards into management of these data. CDISC is taking into consideration the use of observational research.

keywords: standards, data science, automation, artificial intelligence, real world data

(accepted for publication, 5th August 2019)

Corresponding author: David R. Bobbitt
401 West 15th Street, Suite 800, Austin, TX 78701, USA.
Tel: +1 512.363.5826
E-mail: dbobbitt@cdsic.org

I. Introduction

More than twenty years ago, a group of data scientists, statisticians, and researchers began meeting informally to develop a common set of data standards to improve efficiency in the clinical research regulatory submission and review process. From this initial work CDISC (founded as Clinical Data Interchange Standards Consortium) was formed. CDISC brings clarity to data by convening a global community of experts to build data standards that can represent the full lifecycle of clinical and nonclinical research.

II. What are data standards?

The irony of data standards is that there is that there is no standard way to describe a standard. Indeed there is no standard way to generate a standard[1]. Standards generally describe metadata, the data about data. Standards attempt to organize in a somewhat predictable manner the who, what, when, where, why, and how of data, i.e. who collected it? When, where, how was it collected and to what ends? Of key interest in clinical research are disorders or diseases under investigation; the intervention that is (most frequently) the target purpose of a study; the frequency and results of interventions and vital assessments; and any other medical conditions—either those the patient already experienced or those that begin during or soon after the intervention begins and ends. CDISC implementers describe CDISC standards as a 'content standard' meaning that CDISC standardizes the contents of the data with minimal distortion of the data. CDISC metadata informs how data should be collected but not what data ought to be collected nor what research questions ought to be asked. CDISC standards also include data exchange standards which organize the metadata and standardize the way data sets are structured and exchanged to support data sharing and cross-system interoperability.

III. The CDISC Standards Metadata

Over years of iteration and development, CDISC has built an architecture to represent data that largely aligns with the structure in our core use case, the data sets built by the biopharmaceutical, life sciences, and medical device industries to represent to regulatory agencies that their product is safe and efficacious. CDISC core foundation standards include:

- SEND, the preclinical standard, represents safety data, data from animal models and tissues, and data from pre-human clinical trial work in an organized and coherent manner.
- CDASH, the data collection standard, standardizes the collection of data in a harmonized way facilitating the direct mapping of collected data into the tabulation and analysis segments of the model.
- SDTM, the tabulation standard, allows researchers to organize, format, and tabulate the data. SDTM supports data aggregation and warehousing.
- ADaM, the analysis standard, enables the efficient generation of results while maximizing traceability and reproducibility[2].

The core foundational standards are undergirded by:

- The Protocol Representation Model (PRM) developed in collaboration with TransCelerate BioPharma's Common Protocol Template, standardizes planning and designing a research protocol with focus on study characteristics such as study design, eligibility criteria, and requirements from ClinicalTrials.gov, World Health Organization, and EudraCT registries[3].
- Controlled Terminology (CT) is a shared lexicon of terms, concepts, and variables that is developed in partnership with the National Cancer Institute—Enterprise Vocabulary Services (NCI-EVS) of the US National Institutes of Health[4]. CT also includes standardization of commonly utilized questionnaires, ratings, and scales (QRS).
- Data Exchange Standards facilitate the transference of metadata and data across the various electronic systems throughout the clinical research lifecycle.
  - Define-XML is a data exchange standard that stores CDISC SDTM and ADaM metadata in a machine-readable format, enabling automation and making the data easier to understand and share[5].
  - ODM is a platform-agnostic format for exchanging and archiving clinical and translational research data, along with their associated metadata (i.e., administrative, reference, and audit information). ODM is one of the most widely-used of the CDISC standards[6] as it is commonly utilized for representing case report form (CRF) content in many electronic data capture (EDC) systems[7].

To facilitate the broad, consistent usage of CDISC standards, CDISC began developing Therapeutic Area User Guides (TAUGs) in partnership with the Critical Path Institute[8], the US Food and Drug Administration (US FDA), Japan Pharmaceuticals and Medical Devices Agency (PMDA), NCI-EVS, and a variety of other stakeholders that varies depending on the therapeutic area. TAUGs jumpstart trial standardization by highlighting the most commonly utilized or special data collected for that indication and by providing examples of CDISC standards within both scientific and clinical context. More than thirty TAUGs exist that

cover disease categories including oncology, cardiovascular disease, neurological disorders, and infectious disease. CDISC TAUGs will continue to be developed to support evolving science and regulatory needs[9].

IV. Volunteer Development of Clinical Data Standards

CDISC standards have been developed and continue to be developed by a global community of volunteer experts. CDISC was founded by volunteers in fact. These volunteers include executives, data scientists, statisticians, computer programmers, and standards experts to name a few. Some volunteers utilize their own time on CDISC projects and many volunteers are supported by their employers to work on team-based projects. Volunteer teams manage the scoping, launch, update, and development of standards as well as help identify new areas for standardization (such as the CDISC Pharmacogenomic and Genetics Standards, or PGx). Volunteer standards development work is organized to ensure it aligns with both strategy and conformance to the general CDISC model. The open process incorporates quality review at key touch points by a Global Governance Group which is comprised of expert data modelers including volunteers and staff. CDISC standards development work is supported by a small but mighty staff of content experts and project managers led by the Chief Standards Officer and two Heads of Standards Development.

CDISC volunteers contribute significantly to other areas of key importance to the global clinical data standards community. Volunteers ensure the accurate and professional translation of CDISC standards and support documents into languages other than English, including Japanese and Chinese. CDISC volunteers support the development of regional conferences called CDISC Interchanges as well as user groups scattered around the globe. CDISC volunteers develop training modules to help those who utilize the standards in order to understand and attain the full benefits of standardization. An all-volunteer Board of Directors sets high-level strategy and selects the nonprofit organization's Chief Executive Officer.

V. The Global Standard

CDISC standards have become the de facto global standard for clinical research because CDISC standards are broadly utilized to facilitate the critical conversation between—on the one hand—industry actors that sponsor the development of pharmaceuticals, diagnostics, and medical devices and—on the other hand—the regulatory agencies that protect public health through approving only those pharmaceuticals, diagnostics, and devices that are demonstrably both efficacious and safe. The regulators include:

- PDMA which has mandated CDISC standards in October 2016 with a three-year transition period. The transition period has recently successfully ended so most clinical trial data is now submitted in CDISC standards[10].
- US FDA which has mandated CDISC standards since December 2016 (December 2017 for new Initial New Drug Applications)[11].
- China National Medical Products Agency (NMPA, formerly China FDA) which strongly recommends CDISC standards. As of 2018, more than 70% of patient trial data submitted to NMPA was built to CDISC standards[12].
- EU European Medicines Agency (EMA) does not require patient-level trial data Nevertheless, EMA and the Heads of Medicines Agencies representing each EU member state explicitly recommend utilization of global data standards, naming CDISC specifically for clinical research data. EMA generally recommends against developing new data standards whenever standards currently exist[13].

Government funders of research also recommend and utilize CDISC standards including Innovative Medicines Institute[14] and the US National Cancer Institute (NCI) which recently launched its Cancer Data Research Commons, a cloud-based platform for data sharing, analysis, archiving, and data re-use and which is built utilizing CDISC standards from inception[15]. NCI is currently the world's largest funder of oncology research. Private funders of biomedical research, including the Bill and Melinda Gates Foundation and the Leona M. and Harry B. Helmsley Charitable Trust, are increasingly supporting CDISC standards development in areas of interest and are also utilizing standards in their clinical research programs[16].

Regulators tell CDISC that they choose CDISC standards because

- Standardization improves the quality and timeliness of reviews.
- CDISC standards are well-designed to support reviewers in finding data and understanding the claims of sponsors.
- Since so many multinational pharmaceutical companies, CROs, and technology companies serving these companies already utilize CDISC standards; have built CDISC standards into their enterprise architecture; and train their staff members to utilize CDISC standards, CDISC standards have become the de facto global standard for industry.
- CDISC ensures stability of the standards and of the model over time, including managing up-versioning of standards

due to CDISC's process and thanks to the organization's longevity and stability.
- Tools built from CDISC standards such as visualization tools and validation tools are available in the global marketplace.
- CDISC provides training to statistical reviewers and medical reviewers[17]

CDISC standards are mature, globally recognized and globally utilized. CDISC standards have become the global standard for clinical trial data.

VI. New Ways to Consume Data Standards

While CDISC standards are stable, they are also dynamic. New sources of data, new types of research science, and new technologies require the standards to be updated and require that new versions be developed. In 2017, CDISC launched an ambitious project to generate a single machine-readable instantiation of the CDISC standards. This metadata repository named CDISC Library launched in April 2019. CDISC Library was built using a new semantic technology stack, which is expressed in as a graph database via Resource Description Framework, or RDF. CDISC Library utilizes linked data and a REST API to deliver CDISC standards metadata to eSystems that can help automate standards-based processes. At launch, the CDISC Library contained more than 1 million resources, including all foundational standards and CT for the prior five years. A graph database allows resources to be linked at the conceptual level, so there are more than 6 million linkages (i.e., RDF triples) inside CDISC Library. CDISC standards now have a single source of truth for all implementations.

This first-in-class standards metadata repository is a robust substrate, a technology platform to support new innovative applications and tools. CDISC is beginning to work with tool developers that will help build the front-end user interface/experience layers, including the automation of repeatable standards-related activities to ensure easier and consistent implementation of CDISC standards.

New content will be added to CDISC Library each quarter including updates and new versions of foundational standards, TAUGs, QRS supplements, and additional CT[18].

In January 2019, CDISC launched an automation pilot project named CDISC 360 that builds on the new capabilities of CDISC Library. The goal of CDISC 360 is to demonstrate over a very narrow set of clinical research outcomes a complete, machine-readable automation of CDISC standards from beginning to end (and then from end to beginning, thus demonstrating end-to-end automation). This pilot project includes loaned talent from 26 member companies in Europe, Japan, North America, and China. When successful, CDISC 360 will provide a template to build multidimensional standards. CDISC's goal is to build standards for machines first, people second. Such new standards will contain the needed linked metadata to allow standardization and visualization of:
- Electronic case report forms (CRFs)
- Tables, figures and listings (TFLs)
- Other research artifacts supporting consistent implementations of CDISC standards
- Standardized machine-readable mappings to other standards such as HL7 FHIR

CDISC 360 is an ambitious effort divided into short-term sprints. Ongoing information on the CDISC 360 project will be available at the CDISC website[19].

VII. New Sources of Data

In the coming years, new sources of data will enter the clinical research enterprise. Like most standards development organizations, CDISC's work is inherently conservative: after science is settled, then standards can be developed effectively. Nevertheless, CDISC anticipates some areas where CDISC standards will likely continue to develop in the coming years, including:
- eSource and scaling up use of electronic health record (EHR) data. CDISC has a partnership with HL7 FHIR and TransCelerate Biopharma that focuses on improving clinical trial execution incorporating data sets from EHRs[20]. CDISC staff continue to see the challenges and limitations inherent in current EHR data sources, yet staff also recognize the potential trove of insights.
- While common wisdom suggests that artificial intelligence (AI) including machine learning (ML) and deep learning (DL) can change the clinical research landscape, there is notable caution that data quality is insufficient and expensive to improve[21] and that such applications require significant quantities of complete data, not merely more data, which is easier said than done[22]. CDISC staff believe that the CDISC Library and CDISC 360 projects, above, are critical to unlocking the potential of AI and ML in this space.
- Genomics data and personalized medicine continue to evolve. The CDISC SDTM PGx team is working to update CDISC standards in this space[23].
- Registries are of growing interest to CDISC. The CDISC Blue Ribbon Commission suggests CDISC should work with stakeholders to build tools such that registries are designed with CDISC standards from inception[24].

VIII. Observational Studies

Historically, CDISC standards have primarily been used for regulatory submissions of clinical trials data in support of approval to market medical products. However, recent expansion of CDISC standards through therapeutic area user guide (TAUG) development and an increase in CDISC visibility has led to the recognition of the value of data standards in other areas of medical research as well. The existing biomedical conceptual content of CDISC standards, described mostly in TAUGs, is study type-agnostic and aligns well with analogous concepts examined from limited comparisons of data collected in observational studies[25]. CDISC standards are utilized for example by academic researchers who hope to benefit from data standardization. These researchers perceive standards to offer a significantly faster and less costly avenue for generating evidence and performing robust analyses as well as for making research data available to share and for ensuring reproducibility of studies[26]. CDISC standards are utilized in global data sharing platforms including Project Data Sphere[27] and Vivli[28] among others. CDISC tools are also available for academic-focused research through platforms such as REDCap[29]. Training products and derivative products from CDISC TAUGs have been developed to support among other use cases, data capture in low and middle income countries field operations[30].

There are many challenges faced by academic researchers utilizing CDISC standards. Observational studies differ from randomized controlled trials in significant ways regarding study goals, study design, subject populations, clinical settings, regulatory/study oversight requirements, and data collection and data management practices. Many of these differences present challenges and are at least perceived to be barriers to the adoption of CDISC standards in observational research. Unlike a randomized controlled trial, observational studies do not involve an intervention and no attempt is made on the part of the investigator to impact health outcomes. When collected in an academic or government research setting, observational data are often of high quality; these studies are protocol driven and subject to oversight by an Observational Study Monitoring Board. Like randomized controlled trials, observational studies vary in the study design employed and can be generally categorized as case-control, cohort, or cross-sectional studies. The intent of a randomized controlled trial is to determine the safety and/or efficacy of an intervention. In contrast, observational studies seek to relate potential risk factors to disease outcomes. Because of the lack of randomization, observational studies are more prone to bias and thus potential confounding factors must be collected in order to control for bias during analysis[31]. An analysis of a trial of one molecule is relatively straightforward to standardize: did the intervention cohort receiving the molecule change from baseline and change in comparison to a control cohort over a period of several weeks or months? An observational study might follow a population over decades and consider whether lifestyle, behavioral, environmental, and socioeconomic factors contribute to health outcomes and to what degree. The CDISC model will require additional components to represent observational studies.

A second challenge is that it is enormously challenging to standardize protocols for academic studies. By their nature, academic researchers are building new knowledge. Protocols for even simple studies demand extraordinary attention to detail. It is most difficult to reflect nuance at this level in standards, so any data standards, not just CDISC standards, will be sorely pressed to standardize protocols with any granularity.

Beyond research driven studies, observational data may also be generated from RWD sources including EHRs, claims and billing, patient registries, and mobile devices. These data have generally not been collected with the intent of supporting research and thus may be less complete and of lower quality than data collected in a research setting[32].

CDISC staff have had limited interactions with the large and growing global community of academics who utilized CDISC standards due to staff size and budget constraints. Yet CDISC staff believe there is a vast opportunity to improve data standardization and extend the benefits of standardization to their work. Some areas for future exploration include:

- Understand how academic investigators can best utilize CDISC standards
- Develop a 'fit for use' extension of CDISC standards that can support academic investigators (but that does not apply to the sponsor-regulator use-case)
- Develop specific strategies to overcome existing barriers to data standardization
- Generate a better substrate for data sharing and support platform-agnostic data sharing
- Develop open-source tools and other affordable tools built on the CDISC Library
- Recognize good data standardization by these researchers and encourage journal article development

IX. Conclusions

CDISC standards collectively form a mature system of inter-related componential standards that represent the lifecycle of clinical research. CDISC standards are evolving

to leverage new sources of data available to the research enterprise, as well as new technologies. This evolution will make standards development more efficient and the standards themselves more rich and useful. We at CDISC believe that these changes will bring clarity to data and thereby speed the development and approval of new safe and efficacious treatments to patients who so desperately need them.

References

[1] Russell A, Vinsel L. The Joy of standards. The New York Times, 16 February 2019.

[2] CDISC. https://www.cdisc.org/standards (accessed 2019-08-02)

[3] TransCeleratebiopharmainc. https://transceleratebiopharmainc.com/initiatives/common-protocol-template/ (accessed 2019-08-02)

[4] NIH. https://www.cancer.gov/research/resources/terminology/cdisc (accessed 2019-08-02)

[5] CDISC. https://www.cdisc.org/standards/data-exchange (accessed 2019-08-02)

[6] Brix TJ, Bruland P, Sarfraz S, Ernsting J, Neuhaus F, Storck M, Doods J, et al. ODM data analysis— A tool for the automatic validation, monitoring and generation of generic descriptive statistics of patient data. PLoS One. 2018. https://doi.org/10.1371/journal.pone.0199242 (accessed 2019-08-02)

[7] Hume S, Aerts J, Sarnikar S, Huser V. Current applications and future directions for the CDISC Operational Data Model standard: A methodological review. Journal of Biomedical Informatics. 2016. https://doi.org/10.1016/j.jbi.2016.02.016 (accessed 2019-08-02)

[8] Critical Path Institute. https://c-path.org (accessed 2019-08-02)

[9] CDISC. https://www.cdisc.org/standards/therapeutic-areas (accessed 2019-08-02)

[10] PMDA. Basic principles on electronic submission of study data for new drug applications. https://www.pmda.go.jp/files/000153708.pdf (accessed 2019-08-02)

[11] Nelson B. FDA binding guidance: A pivotal milestone for CDISC standards. Applied Clinical Trials. 2016;25(12). http://www.appliedclinicaltrialsonline.com/fda-binding-guidance-pivotal-milestone-cdisc-standards (accessed 2019-08-02)

[12] Wang J. Presentation on NMPA Center for Drug Evaluation Utilization of CDISC Standards. CDISC China Interchange. 2018.

[13] EMA. HMA-EMA Joint Big Data Taskforce Summary Report. 2019. https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf (accessed 2019-08-02)

[14] Mansell P. IMI, CDISC hook up for standardised data management. PharmaTimes. 20 December 2011.

[15] NIH. https://datascience.cancer.gov/data-commons (accessed 2019-08-02)

[16] HELMSLEY. Helmsley charitable trust press release: CDISC to develop data standards for Crohn's Disease with support from helmsley. 2019. https://helmsleytrust.org/news/cdisc-develop-data-standards-crohn's-disease-support-helmsley (accessed 2019-08-02)

[17] Wilson, S. Statistical review and data standards: It's gettin' better all the time. Presentation to PharmaSUG Annual Conference; 2017.5.14-17; Baltimore, MD, U.S. Abstract. https://www.pharmasug.org/2017-proceedings.html

[18] CDISC. https://www.cdisc.org/cdisc-library (accessed 2019-08-02)

[19] CDISC. https://www.cdisc.org/restricted/cdisc-360 (accessed 2019-08-02)

[20] Kellar E, et al. Optimizing the use of electronic data sources in clinical trials: The landscape, Part 1. DIA Therapeutic Innovation and Regulator Science. 2016;50(6):682-696.

[21] Council J. Data challenges are halting AI projects, IBM executive says. The Wall Street Journal. 28 May 2019.

[22] New J. AI needs better data, Not just more data. Center for data innovation blog, 20 March 2019. https://www.datainnovation.org/2019/03/ai-needs-better-data-not-just-more-data/ (accessed 2019-08-02)

[23] CDISC. https://www.cdisc.org/standards/foundational/pgx (accessed 2019-08-02)

[24] CDISC. https://www.cdisc.org/system/files/about/brc/2018-2019_Blue_Ribbon_Commission_Insights.pdf (accessed 2019-08-02)

[25] Neville J, LeRoy B. Considerations for using CDISC standards in observational studies. PhUSE US Connect 2019; 2019.2.24-27; Baltimore, MD, U.S.. Proceedings.

[26] Jauregui B, Hudson LD, Becnel LB, Almario EN, Fitzmartin R, Pétavy F, et al. The turning point for clinical research: Global data standardization. Journal of Applied Clinical Trials. 2019. http://www.appliedclinicaltrialsonline.com/turning-point-clinical-research-global-data-standardization

[27] Project Data Sphere. https://projectdatasphere.org/projectdatasphere/html/home (accessed 2019-08-02)

[28] Vivli. https://vivli.org (accessed 2019-08-02)

[29] Yamamoto, K, Ota K, Akiya I, Shintani A. A pragmatic method for transforming clinical research data from the research electronic data capture "REDCap" to Clinical Data Interchange Standards Consortium (CDISC)

Study Data Tabulation Model (SDTM): Development and evaluation of REDCap2SDTM. Journal of Biomedical Informatics. 2017;70:65-76.

[30] WWARN. https://www.wwarn.org/working-together/malaria-standards-CDISC (accessed 2019-08-02)

[31] Neville and LeRoy, ibid.

[32] US Food and Drug Administration. Framework for FDA's Real-World Evidence Program, December 2018. Retrieved from: https://www.fda.gov/media/120060/download (accessed 2019-08-02)

臨床データ標準の新しい世界
―研究科学，テクノロジー，データソース―

David R. Bobbitt[1)], Bess LeRoy[1)], Amy Palmer[1)], Mike Hamidi[1)], Rhonda Facile[1)], Satoshi Ueno[2)], Sam Hume[1)], Peter Van Reusel[1)], Jon Neville[1)]

[1)] Clinical Data Interchange Standards Consortium (CDISC)
[2)] Center for Public Health Informatics, National Institute of Public Health (NIPH)

抄録

20年前以上前にデータサイエンティスト，統計学者，研究者が会合して臨床研究の成果を規制当局に提出するためのデータ標準規格の策定を試みたのが，CDISC(Clinical Data Interchange Standards Consortium)の嚆矢である．医薬品や医療機器の安全性を証明するためのデータは，発生源，データ収集，集計方法，内容（語彙）について厳密に管理される必要があることから，CDISC標準の中でも非臨床試験と臨床試験むけにSEND，CDASH，SDTM，ADaM，ODMなどの様々な規格が考案された．その結果，世界各国の規制当局に支持され，臨床研究データの世界的な標準規格の地位を占めるに至っている．日本では独立行政法人医薬品医療機器統合機構（Pharmaceuticals and Medical Devices Agency PMDA）が，米国ではアメリカ食品医薬品局（U.S. Food and Drug Administration FDA）が臨床試験のデータをCDISC標準で提出することを義務つけている．さらに近年はコンピュータで自動的にCDISC規格に準拠したプログラム，データを生成できるようにCDISC Libraryが公開され，CDISC 360プロジェクトの有志によって実証実験が進められている．そして，CDISCの対象範囲は一般的な臨床研究にも広がり，疾患領域別データ標準であるTherapeutic Area User Guides (TAUGs)という拡張によって，腫瘍，血管疾患，神経疾患，感染症等における臨床研究もカバーされつつある．また，世界最大のFunding Agencyである米国国立がん研究所（National Cancer Institute NCI）は研究者によって提出されたデータを共有するためのプラットフォームであるCancer Data Research Commons（CDRC）を構築し，そこに蓄積されるデータはCDISC規格に準拠するように求めている．以上のようにCDISC標準はデータの品質管理，マネジメントを含む包括的な規格であるが故に，医学領域の各研究領域に浸透しつつあり，新しくかつ安全で有効な医療機器，治療方法の開発の迅速化に貢献している．リアルワールドデータ（RWD）のデータも品質が低い傾向があるが，CDISCをマネジメントの中に組み込んでいくことで品質を引き上げていくことが期待される．また，CDISCは観察研究の利用も考慮している．

キーワード：標準，データサイエンス，自動化，人工知能，リアルワールドデータ

Appendix（日本語訳全文はこちら）