

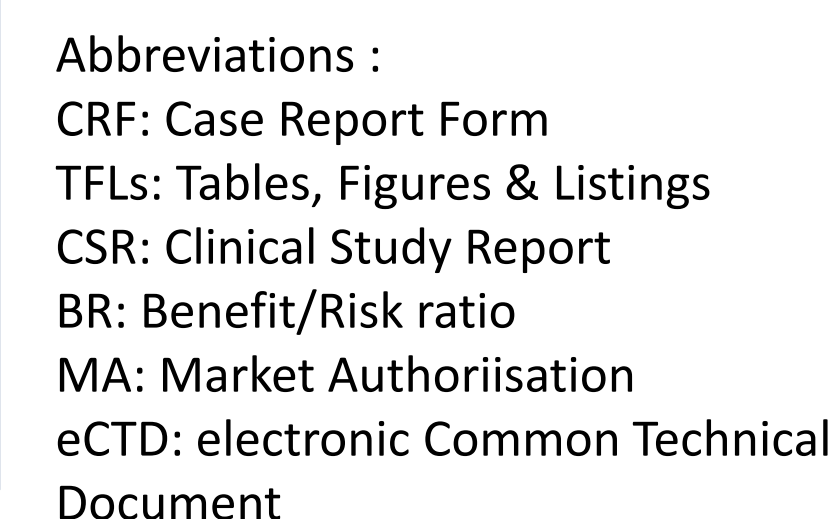


During the past decades, the inception of regulated clinical data standards such as MedDRA and CDISC, has facilitated the sharing, review and analysis of data collected during clinical trials, and enabled the realisation of efficiencies in the processes for handling and management of such data. Standardisation initiatives have taken place in a collaborative manner. Great efforts are made by the various standard organisations to integrate clinical trial data standards in order to enhance their usability and their spread, and significant results are progressively obtained. So, the question “why are clinical trial data standards needed, what is needed?” may seem like an obsolete question to ask nowadays. However, concomitantly to the major benefits obtained so far from clinical trial data standards, the landscape has also become more and more complex and this remains a challenge for those who are willing to use them. The lessons we learnt from our past experience at implementing a CDISC SDTM compatible data warehouse system since 2005, and of several full CDISC SDTM compliant regulatory submissions delivered successfully since that time, have confirmed to us that a holistic approach is a necessary basis for the successful implementation of an efficient clinical trial data management process. Is such a holistic approach possible yet or is this still the quest of the Holy Grail for data management organisations? This is what we are trying to evaluate here.

We propose a method based on the review of existing data standards against requirements and process components, allowing an analysis of available solutions and gaps that will be used in turn to drive the design of our future clinical data management system (Ipsen Advanced Clinical Data Center), define its underlying data model as well as our future data governance principles.

Clinical trial data management basically consists into:

- collecting data from patients who have consented to participate into clinical trials from a clinical development program
- ensuring trustworthiness of this data so that it can be used to generate scientifically sound conclusions
- making this data available in a suitable manner so that the information required to build/maintain the benefit and risk profile of the product for a given population of patients can be easily extracted for regulatory as well as public access



Existing Standards / Gap Analysis

#	Requirement	Acquisition (e) Recording (paper)	Collection (CRF)	Handling & Data Warehousing	Reporting Analysis Publication	Retention
1	To link clinical trial data elements to a standardized, consensus based, reference vocabulary from the point in time where the data is generated (source or origin) and to ensure this link is not altered during the data life cycle	NCI-Enterprise Vocabulary Services, LexEVS central EVS terminology server LexEVS API available for programmatic access to data made available by LexEVS				
2	To maintain a set of standard predefined data availability elements throughout the lifecycle of the data	Standard audit trail: no standard available		Five metadata levels from CDISC define.xml (CRT-DDS)		
3-a	To provide means to easily exchange data across organisations via electronic transfers without any loss of information or decrease in data quality (technical clinical trial data exchange, based on an open standard for the transport of data).	CDISC ODM (.xml), CDISC Lab (.xml), FDA Study Data Specifications based on SAS XPORT Transport Format (.xpt)				
3-b	Ideally to enable data interoperability in a more automated manner (i.e. enable syntactic and semantic interoperability across systems, on top of the technical data exchange capabilities)	CDISC Biomedical Research Integrated Domain Group (BRIDG) CDISC Study/Trial Design Model (.xml) FDA ISCR v2.0 for reporting of safety data Standard for eTMF (e.g. CareLex)				
4	To provide a library of harmonised and/or validated instruments to collect data, according to a pre-defined documented purpose	Validated DCI * CDISC CDASH * Data Collection Instrument	Standard (e)CRF – company/software specific			
5	To enable use of standard dictionaries and provide tools to classify the collected data as well as to normalise the data across studies where relevant	CDISC Controlled Terminology	Standard (e)CRF – company/software specific	ISO country codes, SI MedDRA, SNOMED CT, WHO-DD, etc...		
6	To provide a robust standard method to perform patient data de-identification (anonymisation)				HIPPA Safe Harbor Standard	
7	To enable production and submission of datasets from clinical trials according to the requirements from regulatory authorities (e.g. FDA)			CDISC Define.xml Define.pdf	CDISC SDTM & ADAM	

Based on the above tables, the main characteristics of our future Advanced Clinical Data Center (ACDC) regarding clinical data standards should be:

- a standard vocabulary (NCI-EVS) as a foundation to enable the system to be semantically driven
- an extensible core data model associated to functionalities to link and classify data using standard controlled terminologies and thesauri, among which regulated dictionaries
- CDISC CDASH, SDTM and ADaM metadata maintained centrally and version controlled to enable use of standard connectors between the core model and the desired SDTM and ADaM versions, ease the maintenance of these connectors and keep the full traceability E2E from the data entering into the system up to its reporting and analysis
- data governance principles based as much as possible on open standards and interactions with the Standard Development Organisations

The driver to implement clinical trial data standards has up to now mostly been to fulfill the submission regulatory guidelines, which - as it relates to clinical trials datasets - mainly consists into provisioning the agencies with raw and analysed data according to pre-specified data and metadata transport format, as well as to pre-specified data tabulation and data analysis models. The specifications of these standards have been produced with the gaps in mind (e.g. by lack of a unified approach to audit trail built-in the various data acquisition systems, there are no corresponding standard metadata within the CDISC define.xml). As a result, clinical data management systems (including eCRF applications) and warehouses used by the pharmaceutical industry still keep being insufficient to fully address the need for more efficient data management tools for the sites and sponsors, to leverage the quality, traceability and auditability of data and ultimately to support much powerful search engines for data to be used by scientists and clinicians.

More efforts will continue to be needed for data management organisations (whether from companies or not-for-profit organisations) to be a force that contributes to the vision of the required transformations and to lead the development of the necessary solutions.

The quest of the Holy Grail continues ... there is still a long journey before we can take some rest !...