PhUSE 2013

Paper PP11

Standards for the Management of Clinical Trial Data, why are they needed, what is needed ?

Isabelle ABOUSAHL-CHAUNU, Ipsen Innovation, Les Ulis, France

ABSTRACT

Ipsen has made the choice to replace its clinical data warehouse along with its underlying standard data model. The landscape of clinical trial data standards being quite complex, it appears critical, using the lessons learnt from our past experience at implementing Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM), to thoroughly assess what is needed and why this is needed, using an holistic approach.

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 is used in turn to drive the implementation and define our future data governance principles.

The system will be setup as a hub for clinical trial data, based on an extensible and flexible data model driven by semantic. It is intended to support quality as well as speed of execution, from the recording of clinical trial data to its regulatory submission and publication for scientific and transparency purpose.

INTRODUCTION

During the past decades, the inception of regulated clinical data standards such as MedDRA (Medical Dictionary for Regulatory Activities) 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 will first remind the aim of clinical trial data management and define at high level the scope of the involved process. We will then determine what the requirements are for clinical trial data standards, as a starting point to conduct the review of available solutions and perform the gap analysis. We will then use the results to draw the main conclusions regarding the characteristics of a data warehouse suitable to serve the needs for the management of clinical trial data according to the objectives of the underlying process.

1. DEFINITION OF THE CLINICAL TRIAL DATA MANAGEMENT PROCESS “END TO END” (E2E)

Clinical trial data management basically consists into:

- collecting data from patients who have consented to participate into clinical trials from a clinical development program

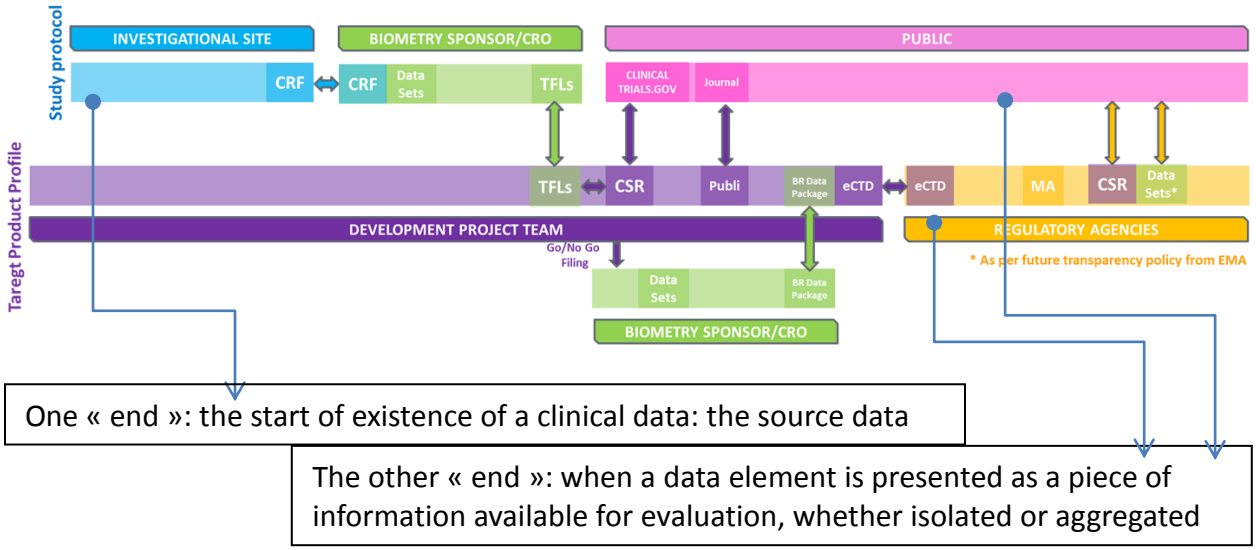
- ensuring trustworthiness of this data so that it can be used to generate scientifically sound conclusions

- 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

DATA LIFE-CYCLE

The main actors of the processes supporting the clinical study data life-cycle are the patients and investigational sites, the sponsor's biometry (i.e. data management and biostatistics) department and clinical development project team, the regulatory agencies and the public.

The process can be summarised as displayed in the following figure:



Abbreviations :
CRF: Case Report Form
TFLs: Tables, Figures & Listings
CSR: Clinical Study Report
BR: Benefit/Risk ratio
MA: Market Authoriisation
eCTD: electronic Common Technical Document

A PROGRESSIVE DATA ENRICHMENT

The data elements to be collected as part of a given clinical trial are defined within the Clinical Study Protocol and represented into the Case Report Form (CRF). The existence of a piece of data starts when it is first recorded (one end) and ends up when being used to produce a statistical table or a result which then is included in a Clinical Study Report, or is reported in a patient data listing appended to a Clinical Study Report, or is used as part of a Periodic Safety Report, or is included in an in-text table of a clinical summary of efficacy or safety, in a publication, in a submitted dataset, etc…

Whereas it looks like an accurate and simple approach to consider that the data exists first in a source document or electronic source system, then as a data collected as part of a CRF before finally being reported, data managers and statistical programmers could witness that the data they are handling have their origin from various processes. If we look at the data elements present into a data management study database and into analysis datasets, we see various "geological" strata, corresponding to different point in time where clinical trial data is generated throughout the data management process.

Some examples illustrating this progressive data enrichment are presented below:

- During the design of a data collection instrument, the designer of a blank CRF will define data points that are parameterised into the form (e.g. planned time-points for blood withdrawal) and which will be transferred

into the clinical database along with the data points recorded by the site in the CRF (e.g. the actual date-time when the sample was withdrawn).

- Data originating from the patient may be recorded directly by the investigational site, or indirectly as a result of a measure done by a third party from a specimen (e.g. a biomarker measured by a specialised laboratory from a sample of tissue).

- Data generated by the investigator or a central reader after having reviewed and evaluated patient data.

- Data added by the site as a result from a particular event (e.g. concomitant medications given at occurrence of an adverse event).

- Data derived by the sponsor's biometry or delegated CRO, from the collected CRF data (e.g. a visit number, a score, the patient's age, the patient's BMI (Body Mass Index), etc…).

- Data derived by the sponsor's biometry or delegated CRO, from the sponsor data management database, for data management, reporting or statistical analysis purpose (e.g. the treatment emergence flag for an adverse event, the number of days in the study at a given visit, etc…).

The resulting process is then clearly far from being a sequence of steps where a full set of data originally captured is transferred from a computerized system to another where it will undergo a set of processing steps. In fact, each system will enrich the data, on the basis of what has been received, and this happens on an on-going basis as patients visits continue to be conducted and data updated up to the end of the clinical study. In this context, maintaining the traceability and quality of the data, as well as representing the data relationships in an accurate manner, is not as trivial as it could look like at first glance.

2. REQUIREMENTS FOR CLINICAL TRIAL DATA STANDARDS (WHY ARE THEY NEEDED ?)

The requirements for clinical trial data standards in the processing of clinical trial data derive from regulatory guidelines and directives, as well as from the business needs. They can be regrouped into the following key principles that should guide any Data Management organisation:

DATA INTEGRITY AND DATA QUALITY

The two pillar principles for data management as defined by the Good Clinical Practices [1] (GCP) are data integrity and data quality, both contributing to the trustworthiness of the reported clinical trial results.

The data integrity pertains to the chain of custody of the data, in other words to the ability to trace the reported data back to its source. The reconstruction of the process from the reported data back to the source data must be possible through auditing or inspection and should be as easy and transparent as possible.

Clinical trial data standards can support the principle of data integrity:

- By ensuring that the definition of each of the data element collected has not been altered during its life cycle (see **requirement #1** in summary section below):
  This can be achieved by the use of standard data element definitions, as a way to ensure that the meaning of a data element and the context into which it has been collected remains consistent and properly documented across the entire process (E2E). Standard data element definitions constitute a consensus based referential to ensure that the meaning of the data can be shared unambiguously between systems and organisations. Standard data elements definitions are made available through controlled vocabularies.

- By ensuring the auditability and traceability of study data (see **requirement #2** in summary section below):
  If altered during its life-cycle, the history of changes should be captured and made available for auditing purpose. Standard approaches to capturing a set of auditing data such as - but not limited to - the origin of the data, the method for its capture, the computational or evaluation method or algorithm used when the data has been derived or normalised, the reason for its change, its date-time stamp and author of the change, should be defined along with a standard structure to link the history of changes to the data element in a complete and transparent manner.

- By ensuring that there is no loss of information when data is exchanged across systems and organisations (see **requirement #3a** in summary section below):
  This is the area of interoperability of systems (according to Wikipedia: "a property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future,

without any restricted access or implementation").
Communication systems that use - at minimum - a common, standard, vendor-neutral technical protocol, language and format to exchange data, in addition to secure data repositories and data transactions, will better support this requirement. This corresponds to the level where technical as well as syntactic interoperability are achieved [2].

The data quality pertains to the suitability and fit of the data collected for the purpose of the study.

Clinical trial data standards can support the principle of data quality:

- By providing data collection instruments standardised and validated to ensure a consistent capture and evaluation of data as a support to a robust analysis (see **requirement #4** in summary section below)

- By providing means to normalise or harmonise the data at the time of capture or later on in the data management process, which provides ways to monitor its quality as well as the quality of its evaluation by investigators and central readers on an on-going basis (see **requirement #5** in summary section below): This can be achieved by the use of standard thesaurus and taxonomies, where all the possible designations of specialised terms are organised according to the up to date knowledge that exists for the considered domain.

- To enable pre-defined meta-analysis (e.g. integrated safety data analysis), the data should be collected according to a set of harmonised content definitions and data collection rules (see **requirements #4 & #5** in summary section below)

COMPLIANCE WITH DATA PROTECTION PRINCIPLES

Another key principle that data management organisations must respect pertains to their role as controllers and/or processors of sensible patient data, and links in particular to the ability to de-identify data when reaching the process where data is published as part of the public domain. This is a specific requirement deriving from the data protection laws [3] that must be considered, in addition to the data integrity and data quality principles described above and which constitute also key pre-requisites to enable the protection of personal data. It has to be noted that other requirements must also be fulfilled but are not in the scope of the present paper (such as patient consent to the processing of their personal data). Standard approaches to patient data de-identification will be more and more needed as it becomes crucial to share clinical trial data (e.g. to re-purpose it for scientific research needs) and to provide transparency to the public as to the way it has been analysed and reported (see **requirement #6** in summary section below)

STUDY DATA EXCHANGE WITH REGULATORY AGENCIES

More and more regulatory agencies will require not only to be submitted documents where data is presented, and statistical results interpreted and discussed, but also the structured datasets (original data as collected into the CRFs as well as analysis datasets). The U.S. Food and Drug Administration has defined a set of standards (available from their Standards Study Data Resources web page) for submitting data, and issued a draft guideline on the use of these standards for regulatory submission [4] (see **requirement #7** in summary section below).

BUSINESS EFFICIENCY REQUIREMENTS

Beyond the compliance to various ethical and regulatory rules and guidelines, it is also important to achieve an efficient interchange of datasets between organisations (partners, CROs, public registrations) and to maximise the performance and speed of the various data processing steps (see **requirement #3b** below). This contributes significantly to reduce the duration of drug development, as established collectively by PhRMA, Gartner and CDISC in the Business Case for CDISC Standards [5]. We are touching here the levels of semantic interoperability, where communication systems do not require human manual processing to establish and execute data transactions. According to the complexity introduced by the progressive data enrichment described above, fulfilling this requirement is a difficult challenge to tackle.

SUMMARY OF KEY REQUIREMENTS FOR CLINICAL TRIAL STANDARDS (WHY ARE THEY NEEDED ?)

Requirement #1: to link clinical trial data elements to a standardised, 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

Requirement #2: to maintain a set of standard predefined data auditability elements throughout the lifecycle of the data

Requirement #3a: 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 non-proprietary standard for the transport of data). **Requirement #3b:** 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)

Requirement #4: to provide a library of harmonised and/or validated instruments to collect data, according to a pre-defined documented purpose

Requirement #5: to enable use of standard dictionaries and provide tools to classify the collected data according to standard taxonomies as well as to normalise the data across studies where relevant

Requirement #6: to provide a robust standard method to perform patient data de-identification (anonymisation)

Requirement #7: to enable production and submission of datasets from clinical trials according to the requirements from regulatory authorities (e.g. FDA)

3. MAPPING OF REQUIREMENTS TO CLINICAL TRIAL DATA STANDARDS (WHAT IS NEEDED ?)

On the basis of the requirements that we have now established, we are proposing to identify (see following table) what are the clinical data standards needed to fulfill each of the seven requirements, and to map them to the main data processing steps where they are required and which we are describing briefly below:

- Data acquisition and recording: capture of the original source data by an electronic device or recording of data by transcription into an electronic system or a paper document
- Collection of data: transcription or transfer of the acquired data into a electronic Case Report Form, could be by manual transcription into electronic forms, electronic data transfer or data entry from a paper CRF by the operator of a Clinical Data Management System (CDMS)
- Handling and warehousing of data: for the purpose of checking the data quality, integrating data from various sources (e.g. reconciliation of central laboratory data with eCRF demographic data), and producing a database ready for reporting and analysis
- Reporting, analysis and publication of data: could be in the form of data listings, data tabulations, statistical analysis, pooled analysis, publication of results in “clinicaltrials.gov”, sharing of de-identified data in public domain via publication, web company site, or regulatory agencies
- Retention: transfer of electronic data records and essential documents into a secure archive area enabling their retention according to regulatory requirements

The table below represents the mapping of clinical data standards to requirements by data processing step:

#	Requirement	Acquisition (e) Recording (paper)	Collection (CRF)	Handling & Data Warehousing	Reporting Analysis Publication	Retention
1	To link clinical trial data elements to a standardised, 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	**Controlled vocabulary**				
2	To maintain a set of standard predefined data auditability elements throughout the lifecycle of the data	**Standard audit trail**		**Standard metadata definitions**		
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).	**Technical & syntactic data exchange standards**				
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)	**Semantic interoperability standards** including Domain Analysis Models and System Implementation Models				
4	To provide a library of harmonised and/or validated instruments to collect data, according to a pre-defined documented purpose	**Data acquisition standards**	**Standard (e)CRF**			
5	To enable use of standard dictionaries and provide tools to classify the collected data according to standard taxonomies as well as to normalise the data across studies where relevant	**Controlled terminology**	**Standard (e)CRF**	**Conversion to standard SI units Standard taxonomies and thesaurii**		
6	To provide a robust standard method to perform patient data de-identification (anonymisation)			**De-identification standards**		
7	To enable production and submission of datasets from clinical trials according to the requirements from regulatory authorities (e.g. FDA)			**Transmission of metadata**	**Standard data tabulations and analysis models**	

Gap or draft standard available only

Open standards available

Proprietary standards available through license agreement or via copyright

Company defined specific standards

To overcome the complexity generated by the progressive data enrichment, the different systems used to acquire data, collect CRF data, handle and manage data into repositories to finally report, analyse and publish the study results should be working according to common standard data elements definitions (standard vocabulary), supported by shared metadata that allows to transport and use standard characteristics of the data reported such as provenance, versioning, standard naming conventions for variables, controlled terms according to shared taxonomies and thesauri, computational algorithms, etc…
The current schematic representation of an E2E implementation as it is generally emphasised (i.e. CDASH to SDTM to ADaM) is a good attempt to address the data traceability E2E, but is not sufficient to address the full holistic

picture, including eSource data (e.g. lab data), capture of data management algorithms (such as scoring of a scale via an eCRF system), nor does it allow to implement a lean process (as an example the complexity introduced by the issue of handling the SDTM/ADaM versioning for submission of supportive (non-pivotal) studies as part of an eCTD). The incurred complexity is perhaps manageable for large size organizations but represents a challenge for medium-small structures, thus the importance to focus first on what is important, available and what are the foundational elements that need to be implemented first to guarantee that the implemented system will be extensible and scalable.

4. REVIEW OF AVAILABLE CLINICAL TRIAL DATA STANDARDS AND GAP ANALYSIS

The table below represents the mapping of available clinical data standards to requirements by data processing step:

#	Requirement	Acquisition (e) Recording (paper)	Collection (CRF)	Handling & Data Warehousing	Reporting Analysis Publication	Retention
1	To link clinical trial data elements to a standardised, 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 auditability 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	Valida-ted 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 according to standard taxonomies 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	

Requirement #1:
The Enterprise Vocabulary Services from the National Cancer Institute (NCI-EVS) constitute a broad and rich referential for controlled vocabulary, providing search capabilities through the NCI meta-thesaurus into a large range of biomedical research standard terminologies and thesauri. It is open access and provides an API for programmatic link to the LexEVS central terminology server. CDISC controlled terms associated to CDASH and SDTM standards are fully maintained along with their definitions into NCI-EVS. CDISC CDASH and SDTM variable names are

generally referenced by their corresponding terms in the NCI-EVS. However, since there are still grey areas in the CDISC vocabulary (e.g. specific domains still open to custom implementations using one of the SDTM general observations classes: findings, events, interventions, findings about events or interventions), gaps exist that would need to be offset before CDISC can be used as a complete referential for standard vocabulary. From this standpoint, the CFAST (Coalition for Accelerating Standards and Therapies, a joint initiative of CDISC and the Critical Path Institute), as well as the CDISC SHARE programs represent a significant step expected to deliver via CDISC the necessary foundation for an E2E implementation of clinical trial data standards driven by semantic.
The off-the-shelve data acquisition, collection, management and warehousing systems currently available from software editors do not provide natively a programmatic interface to the NCI-EVS. As a result, the maintenance of an unambiguously defined and consistently shared vocabulary across the E2E data management chain relies on a manual resource intensive process. Alternatively, efforts can be invested to develop a custom programmatic interface to a referential such as NCI-EVS, to allow each system to maintain the link between the data elements that are built into them, and the concept unique identifiers from the NCI-EVS (C codes from the vocabulary of the NCI-Thesaurus or CUI codes from the NCI Meta-thesaurus). In addition, companies would still need to define and document unambiguously the CDISC domains that are still left open for interpretation or customisation and should better do that by referencing a standard vocabulary, maintained through a robust governance body and user community, such as the NCI-EVS.

Requirement #2:
Standard metadata have been defined by CDISC to support the traceability requirements across Annotated CRF, SDTM and ADaM datasets, however they would need to be expanded to capture the history and pedigree of the data up-stream, that is to say before it comes to the stage where it is being prepared for reporting and analysis. There are barriers to overcome in order to see the emergence of audit trail standards that spans from data acquisition to data archival. One is the existence of various and disparate eCRF technologies that would require significant efforts to align to common processes for data auditing and common representations of data relationships. Another one is the data acquisition process itself that still mainly relies on paper source data.

Requirement #3a:
The SAS® XPORT format is currently widely and successfully used by various organisations to exchange data. Other standards are available, such as the CDISC Operational Data Model (ODM, based on xml), as well as the CDISC LAB that constitutes an extension of the ODM, adapted to the specifics of laboratory data (and will soon include the formats for exchange of ECG and microbiology data).

Requirements #1, #2 and #3a, related to the support to **data integrity principle**, are rather well answered by existing and quite mature clinical trial data standards, even if some gaps must still be compensated by procedural safeguards or custom developments. Data management organisations in charge of small to medium volumes of clinical studies are not necessarily able to tackle those properly as the associated cost or effort may not be affordable for them.

Requirement #4:
A subscription can be made to commercial databases where libraries of validated data collection instruments (questionnaires, scales, etc…) can be searched and ordered under a copyright agreement for those that are protected.
eCRF standards, in the form of libraries of standard layout of forms supporting data entry as well as the underlying standard data relationships and models do not exist off the shelve into eCRF systems, but can be designed and maintained at company level within global libraries, to be picked and chosen from by the eCRF designer at study level. Libraries of eCRF layouts can be built on the basis of the existing content standard provided by CDISC Clinical Data Acquisition Standards Harmonization (CDASH), which are currently covering only a part of what is needed for clinical trials, especially in the efficacy domains. CDASH domains are harmonised with SDTM. They provide in addition to clinical trial data standards a set of comprehensive CRF Design best practices which should be applied to obtain a tool of quality for the sites to collect data.

Requirement #5: The quality of the protocol-driven data to be recorded for a clinical study can be controlled via the CDISC Controlled Terminology. Controlled Terminologies from CDISC are defined either as extensible or not, depending on the nature of the variable to be then reported (e.g. a severity qualifier for an adverse event is not extensible, whereas classification systems such as the test names for cardiac parameters are generally extensible). For data normalisation and harmonisation that are to be applied down-stream in the data management process, a variety of norms (e.g. ISO 3166 country codes), regulated standard dictionaries and thesauri are available via subscription (e.g. WHO-DD, MedDRA), and specific software exist to manage the data classification process and to create programmatic interfaces with data management or data warehousing systems. In the next future, some of these dictionaries may become accessible via the Cloud. As a result, their loading and versioning will be greatly facilitated.

Requirements #4 and #5 for standards supporting the **data quality principle** are not fully addressed, especially those that relate to the process of data acquisition are still incomplete, but are by the time being generally compensated by company specific data standard governance bodies.

Requirement #6:
In the frame of the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the U.S. federal law has produced a guideline [6] referring to the HIPPA Privacy Rule standard for de-identification of Protected Health Information (PHI). Two standard methods are proposed which could be applied in the context of personal data collected as part of clinical trials. More specifically the Safe Harbor method, recommended as the most conservative approach, is based on the removal, by default, of 18 pre-defined types of identifiers.

Requirement #7:
The main standards that answer this requirement are the CDISC SDTM and the CDISC ADaM (Analysis Data Model), along with the Case Report Tabulation Data Definition Specification (CRT-DDS also called define.xml or define.pdf). Define.xml is a machine readable transport format for the CDISC metadata used by the SDTM and ADaM models (define.xml can also be seen as an expansion of the ODM). It is also meant to be a metadata driven tool useful for data reviewers to navigate via hyperlinks through the study data tabulations (individual patient data records), the blank CRF, as well as the analysis datasets. Implementation of clinical data hubs supporting SDTM and ADaM models for submission of individual patient data and analysis datasets remains a big focus for pharma companies and the vendors nowadays. As a matter of fact, the way that has been opened up by the FDA in recommending the submission of clinical study data according to these standard models. The benefit of using them is to facilitate and accelerate the regulatory review and market authorisation for new therapies [5].

Requirement #3b:
The CDISC Biomedical Research Integrated Domain Group (BRIDG) has led the way to semantic interoperability in the area of protocol driven clinical research with the BRIDG Domain Analysis Model. Standard implementations having reached the level of semantic interoperability have been achieved so far in the area of safety data for expedited reporting (ISO Individual Case Safety Report (ICSR) standard, which was finalised in December 2011). Even if this represents a niche comparing to the paradigm of efficacy data, this shows however a concrete example as to how standards supporting semantic interoperability can be reached and there are lessons to learn from this example probably. Another area of implementation that is progressing in this direction is the Study Design Model that has reached an important step in 2010 with the Protocol Representation Model (PRM version 1.0) released by CDISC, along with a set of use cases. Finally, it should be noted that the CareLex group has produced an ontology based model for eTMF, and is working under the umbrella of the OASIS SDO (Standard Development Organisation) to produce globally available, open eTMF standard aiming at accelerating automated information exchange and interoperability of clinical trial information.
Finally, an important progress is expected from the CDISC SHARE project, which is built upon BRIDGE model and ISO21090 data types to facilitate the link of CDASH/SDTM standards with healthcare concepts.

It has to be highlighted that the standards fulfilling requirements #1 and #2 constitute also the foundation for a full semantic interoperability. It is clear from the gap analysis relating to those requirements that the foundation is not strong enough, but the foreseen implementation of CDISC SHARE is again expected to help filling in this gap.

5. CONCLUSION FOR THE IMPLEMENTATION OF IPSEN ADVANCED CLINICAL DATA CENTER (ACDC)

Based on our analysis, 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 define.xml, SDTM and ADaM metadata maintained centrally and version controlled to enable use of standard connectors between the core model and the desired STDM 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

CONCLUSION

The roles of clinical trial data standards are to provide technical, syntactic, and semantic interoperability. They constitute a consensus based referential for sharing the meaning of data in an unambiguous manner. They help ensuring that the content required for evaluation of a clinical study data is consistently obtained across studies to enable the «pooling » and meta-analysis of such data. They specify data models ready to be used by specific applications (e.g. loading into the FDA reviewers systems and tools). They allow classification of data that share same characteristics, and provide logical robustness and scientific accuracy for search purposes.

The driver to implement clinical trial data standards has up to now mostly been to fulfil regulatory guidelines, which - as it relates to submission of clinical trials datasets - mainly consist 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.
As a result, clinical data management systems (including eCRF applications) and warehouses used by the pharmaceutical industry still reflect requirements focusing on a narrow (yet highly critical!) area and thus keep being insufficient to address the much broader need to enhance the access to raw and analysed data for regulatory, research and scientific purposes, and to tackle the additional complexity it generates for data management processes.

If we think about standards in a broader manner, we embrace the potential they offer for transforming the site and sponsor processes by supporting much more fluid and flexible tools for the management of data they collect and report, 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.
Regulatory agencies are leading the industry towards this direction. The pharmaceutical industry willingness to share more data in a pre-competitive approach via public-private consortiums or by unilateral decision is increasing and needed for the benefit of patient care.

An increased reactivity from the providers of Information Technologies and from the SDOs will thus continue to be required in order to address the gaps that are currently encountered to fulfil the requirements we have listed. This means in a first place more efforts from the data management organisations (whether from companies or not-for-profit organisations) not only to continue adapting to the changing environment but more importantly to be a force that contributes to the vision of the required transformations and leading the development of the necessary solutions.

REFERENCES

[1] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use - ICH Harmonised Tripartite Guideline – Guideline for Good Clinical Practice E6 (R1)
[2] ETSI White Paper No. 3 Achieving Technical Interoperability - the ETSI Approach - Authors: Hans van der Veer (Alcatel-Lucent), Anthony Wiles (ETSI Secretariat) 3rd edition - April 2008
[3] Data Protection Directive – officially Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data
[4] U.S. Department of Health and Human Services - Food and Drug Administration Guidance for Industry - Providing Regulatory Submissions in Electronic Format - Standardized Study Data (draft guidance / February 2012 / Electronic Submissions)
[5] Business Case for CDISC Standards / Full Report -PhRMA-Gartner-CDISC Project - September 2006, Carol Rozwell (Gartner), Rebecca Daniels Kush (CDISC), Ed Helton (SAS), Frank Newby (CDISC), Tanyss Mason (CDISC)
[6] Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule November 26, 2012

RECOMMENDED READING

- CDISC: Adoption Trends, Benefits and Addressing Barriers, By William Friggle, Feng Li, Shannon Labout, Rebecca Kush; CDISC Journal; October 2010

- U.S. Food and Drug Administration ; Information from Public Meeting : Solutions for Study Data Exchange Standards, November 5, 2012 - available from the FDA website

- U.S. Department of Health and Human Services - Food and Drug Administration Guidance for Industry - Electronic Source Data in Clinical Investigations - DRAFT GUIDANCE - November 2012

- John S. McIlwain - Data Standards Harmonization - A look at the past, present, and future of interoperability across the clinical research spectrum, Applied Clinical Trials, Volume 19, Number 5, May 2010

- Enhanced EDC Service, A Case Study – CDISC and caBIG Standards-Based CRF Library. By Yun Lu, Ph.D., Manager of Clinical Data Management Systems, KAI Research, Inc.; Patti Shugarts, Chief Operating Officer, KAI Research, Inc.; Wenli Zhang, Ph.D., IT Project Lead, KAI Research, Inc.; Kristy Miller, MPH, Clinical Research Manager, KAI Research, Inc – Data Basics; A Publication Supported By and For the Members of The Society for Clinical Data Management, Inc.- Volume 20, Number 1; 2013 Spring

CONTACT INFORMATION

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

Isabelle ABOUSAHL
Ipsen Innovation
5, avenue du Canada
Les Ulis, 91940 France
Work Phone: +33 1 60 92 94 16
Fax: +33 1 60 92 94 64
Email: isabelle.abousahl@ipsen.com