Standards and Ontology Driven Workflow Automation of Clinical Studies

PointCross Life Sciences

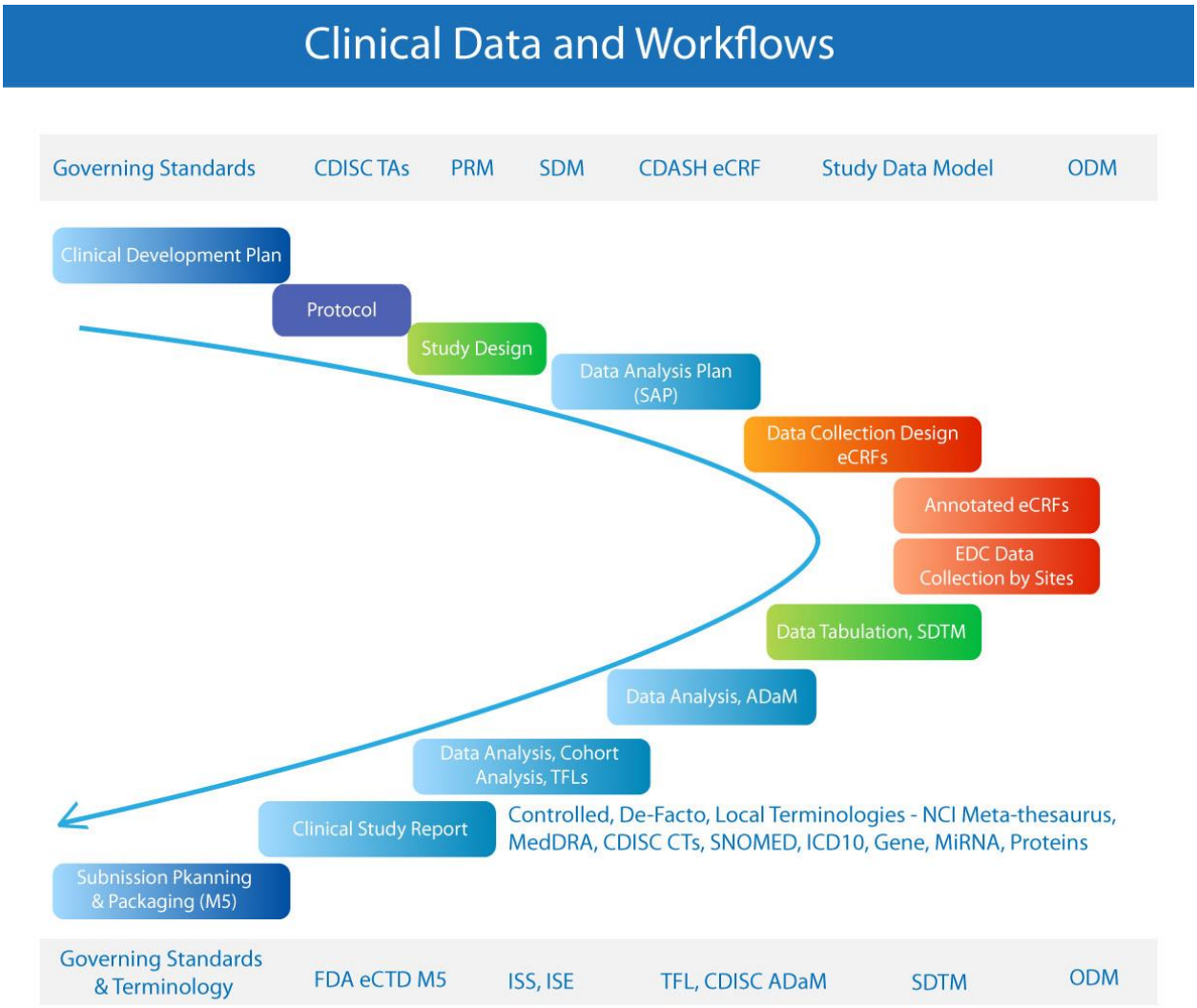
10 October 2017

Clinical trial costs have been rising to keep up with the complexities of drug development and the business need for better benefits to risk ratios. This 2014 paper, "Examination of Clinical Trial Cost and Barriers for Drug Development" written by the Assistant Secretary of Planning and Evaluation of **U.S. Department of Health and Human Services (HHS)** points out many of the cost drivers and barriers for drug development. At the time of writing this paper took into account various aspects including electronic health records, simplifying protocols and controlling amendments, wider use of EDC and other forms of testing including at-home or remote. A number of developments within the industry and regulatory environment now offer new opportunities to both reduce and streamline the clinical trial data and workflow process while improving the richness and value of the data being collected.

In December 2014, the FDA mandated that all submissions of studies, both nonclinical and clinical must be in machine consumable digital format. This was not just about requiring submissions to be in electronic format instead of paper, it was mandating that the data be published in a manner that is machine readable and usable so that the data can be instantly viewed, analyzed by programs and scripts, and otherwise re-purposed for review without the need for manual transcription or handling. The de jure standard of choice has been CDISC and its tabulation model SDTM for raw data and ADaM standards for analysis ready data extracted from the collected data. These were data "Exchange" standards meant for two systems to communicate data in a very coherent, clear and consistent manner.

The paper by HHS considers the possibilities of streamlining and reducing costs of clinical trials within the data and workflows through automation exploiting various standards and published controlled terminology lists to codify specifications for quantitative and qualitative data. Separately, the paper also covers the processes that touch site investigators, scientists, technicians, and regulatory affairs very well.

The lifecycle of a specific clinical study with regards to data begins with a protocol and ends up with a completed study report that is ultimately used to determine objectively by the drug developers and regulatory agencies if the drug is effective and safe. This lifecycle is punctuated by a series of stages of workflows to specify, collect, organize, analyze or package data. Most of these workflow stages require considerable manual effort involving data management, programmers, bio-statisticians and Study Data Tabulation Model
(SDTM) savvy data wranglers. These stages are designed to address questions about how the study should be conducted such as: what kind of data will be collected and when it will be collected from subjects recruited into the trial; how the data will be analyzed to generate an objective unbiased view of specific cohorts in the trial to the drug candidate; and how the data shall be modeled and reported for submission to the regulatory agencies.



As you can see in the workflow stages shown in the above chart, data is collected at the site through the Electronic Data Capture (EDC) systems where Case Report Forms (CRF) are available with the right information required to be collected for each interaction with the subject. The activities on the EDC are not considered here because that is usually well-managed and automated within the EDC software and the site investigators. What is important though is the EDC system's internal data model reflects the study design. While in the past data modelers needed to configure the databases to reflect the data models for each study and then spend considerable amount of effort to modify them in case there is a protocol amendment, today most EDCs have or will have the ability to accept an Operational Data Model (ODM) XML of the eCRFs and the Study Design. Similarly, adaptors can export the EDC data from its ODM to a standard exchange format such as SDTM. However there is a challenge that remains with the transition from the protocol to the data collection point and then from the EDC to the study report and submittable data sets.

The workflow diagram above lists a number of governing standards for how various types of data can be modeled and exchanged. CDISC has extensive coverage of the clinical development stages. There are some other relevant standards such as HL7 for patient records and the FDA has standards and

guidelines for eCTD gateways for data submissions and the structure in which the submissions information must be organized.

Managing the use of terminology throughout the lifecycle makes it easier to streamline and automate the processes. The controlled terminology version selected for the standardized exchange needs to be matched with the terminology in use by the sponsor, the site investigators. This requires that both the set of standards and the terminology must be managed in a robust repository and semantically enabled ontology engine.

The protocol for a clinical trial follows the scientific intent behind the study and that is influenced in great part on the phase of the trial and whether it is focused on safety or efficacy. However, the data that must be collected and the tests that must be conducted are influenced by the disease being targeted. Standards for Therapeutic Areas (TA – see https://www.cdisc.org/standards/therapeutic-areas and https://www.cdisc.org/system/files/all/standard/CFAST-TA-Project-Status.pdf) are being published by CDISC that can help Pharma companies create standard templates using the disease specific metadata and other implementation guidelines for each stage of the study lifecycle up until submission. These can be combined with legacy study protocols and study designs to accelerate the workflow for defining protocols and study designs.

One of the most expensive, time consuming and manual efforts goes into dealing with protocol amendments. In the past, these amendments mean changes in the CRFs that are to be distributed to the sites; changes in eCRFS where EDCs are used; and the most expensive workflows are in making changes to the EDC's study specific data models and migration of data. Protocol Representation Models (PRM) can accelerate the development of the CRFs and Electronic Health Record (EHR) for defining study attributes and design, subject eligibility and requirements from various resources such as ClinicalTrials.gov, World Health Organization (WHO) registries, and EudraCT registries. See https://www.cdisc.org/standards/foundational/protocol)

CDISC SDM-XML is an important extension to the ODM-XML for providing Study/Trial Design Model in an exchangeable format that is machine readable. This includes specifications of the study design, treatment plans, eligibility, visits, and events that must be monitored. CDISC SDM-XML can be used to define the Structure, Workflow, and Timing of a clinical study's design. See https://www.cdisc.org/standards/transport/sdm-xml

CDISC SDTM tabulation standard and the selected controlled terminology is the exchange standard for submitting data to the regulatory agencies. CDASH ODM XML offers a way to standardize the collection of data in the CRFs and annotating it for easy conversion to SDTM from the EDC. See https://www.cdisc.org/standards/foundational/cdash.

SDTM provides a columnar standard for organizing and formatting collected data from subjects participating in the trial and exchanging the data among contracting research organizations, sites, Pharma sponsors companies and regulatory agencies. See https://www.cdisc.org/standards/foundational/sdtm

ADaM defines analyzable datasets and metadata needed for clinical trial statistical analyses. It provides a traceable record of the SDTM source data, and the intermediate data sets to the final published results. See https://www.cdisc.org/standards/foundational/adam

Both SDTM and ADaM sets need to be worked on by programmers and bio-statisticians who prepare analysis scripts or use standard data visualization software to generate study results that are meaningful to medical and clinical researchers. Automation at this stage is not just about automating data flow but it also includes various search tools and cohort selection facilities that are seamlessly accessible without having to go to disparate data stores and use data management effort to pull out and make such data available. It is quite common to hear the complaint that as much 75% of data management and programmer time is spent on looking for data and only 25% on actually working with the data. That can change with an integrated environment.

The FDA provides guidelines for the structure and formatting of data and results that need to be submitted in the Integrated Summaries of Safety and Effectiveness (ISS/ISE) in Applications Submitted in the eCTD Format. See
https://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm163558.htm

The Electronic Common Technical Document (eCTD) is the standard format for submitting applications, amendments, supplements, and reports to FDA's Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER).
An eCTD submission has five modules: region-specific information, summary documents, quality-related information, nonclinical study reports, and clinical study reports.
https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM511230.pdf

Sponsor companies that conduct clinical trials can streamline all stages of the study lifecycle by using workflow automation that is end to end and supported by a standards repository that makes use of the XML specifications to drive data adaptors with semantic enabled ontology. It allows use of controlled terminology from standards organizations as well as the local terminology used by sites or contracting organizations with a curated terminology management system. Since the data is exchanged among so many disparate organizations it is also important at each interchange to have clear definition and specification of how data will be provided by one party to another. With rigorous attention to how data is expected and a automatic verification of data against these specifications ensures that the source data is constantly being verified and quality and consistency is ensured.

Clinical Study Workflow Stage	Types of questions addressed
Clinical Development Plan	What kinds of studies should be conducted to meet the needs of the therapeutic area and stage of the investigation?
Protocol	How to conduct the study? Who (what kind of subjects) should be recruited to the study?
Study Design	How to organize the trial? When should the subjects visit the clinic during the course of the study? What biological samples should be taken from the subjects and what data should be extracted from the samples? What data should be collected from the subjects during visits or at home? How should the subjects be dosed and when?
Study Data Analysis Plan	How will the data be analyzed and how will the subjects be grouped for analysis?
Data Collection Design (eCRFs)	How will the data be organized for reporting and analysis?
Annotated eCRFs	What terminology should the data be reported to the sponsor?
Data Collection	How will the data be captured in the EDC?
Data Tabulation (SDTM)	How will the data be validated to make sure that it meets the original intent and model? How will the data be transformed to meet the data exchange standards (SDTM)?
Data Analysis (ADaM)	How are the analysis data sets extracted? What analysis scripts are applicable? What tabulations, figures and listings should be reported (TFL)?
Clinical Study Report	Who generates the study reports?
Submission Planning and Packaging	Which studies will be packaged to support each of the planned submissions to the agency?

References:

ASPE, Examination Of Clinical Trial Costs and Barriers For Drug Development
July 25 2014
https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development