7. Maximizing the value of traditional EDC

Overall, EDC was successful in displacing paper CRFs, speeding-up access to data from patient visits and streamlining the query process. Despite these great achievements, its full potential has not been fully

realized. For example, the challenges surrounding the transcription of site source data remains. Source data verification (SDV) functionalities have been implemented to better manage the symptoms, but not address the root cause which would eliminate the need to transcribe data altogether. Most EDC systems have not reduced the need for extensive, resource-intensive data checking processes, such as SDV and complex edit checks, and many implementations simply converted existing, inefficient paper processes to an electronic tool[1]. Additionally, its centricity is being challenged by the rapid adoption of direct-to-patient data captures (e.g., eCOA, sensors, wearables, etc.).

Furthermore, while the pharmaceutical industry concentrated on EDC, the healthcare industry invested heavily in its own electronic health record (EHR) systems, further increasing the technology gap between clinical research and medical care. Fortunately, sophisticated EHR systems, built on common healthcare data standards (i.e., Health Level Seven (HL7)) and interfaces, are now available. They can enable EHR-to-EDC data transfer, and ultimately eliminate the need for transcription and associated SDV[9]. While it is challenging to implement at large scale, this direct connection to EHR to ease clinical research is a real opportunity. Yet, CDM professionals must consider the increased level of complexity involved in integrating and managing these data streams, clearly documenting the data lineage from "collection to analysis" and addressing security and privacy requirements.

Regardless, EDC still has a critical role to play in today's clinical research and its use must be maximized. According to the 2017 Tufts study on the e-Clinical landscape[8], EDC remains the most prevalent clinical data application used, followed by randomization and trial supply management system (RTSM). It means CDM needs to be pragmatic about positioning EDC in a broader spectrum of data sources and maximize its value through a fit-for-purpose EDC Strategy while investing in future-proof solutions.

Below are lessons learned for identifying quick wins to maximize the value of traditional EDC.

EDC is not the optimum place to load external data: EDC systems were not designed to be the central study data repository and should not typically be viewed as the place to load all external data. CDM just used them this way, in the absence of any other viable solution. A total of 77% of companies surveyed by Tufts had issues loading data into their EDC application and 66% identified EDC systems themselves or integration issues as the primary reasons for being unable to load study data in EDC[8]. Additionally, EDC systems have been designed to meet specific requirements such as SDV or manual entry of data that do not apply to external data. It is therefore not surprising that less than 22.5% of the total volume of data in the EDC came from external sources, including but not limited to local lab data (5%), central lab data (5%), quality of life data (4%), and ePRO data (3%). Genomic data and m-Health data made-up the smallest portion of data in EDC systems at 0.4% and 0.3% respectively[8]. Lastly, as mentioned in the initial SCDM reflection paper[1], several companies informally surveyed during the SCDM 2018 Leadership forum stated that over 70% of the data volume (e.g., lab, eCOA, etc.) does **not** come from EDC.

In the absence of alternatives (e.g., intelligent CDMS allowing real-time data aggregation), CDM should limit the loading of external data in EDC to data critical to timely decision making. This may include investigator review, key safety reviews by sponsor physicians, site compliance (e.g., support drug reconciliation by monitor). Companies should investigate future-proof data integration and reporting platforms that are compatible with current and future data streams, including sensors and wearables.

Release EDC before First Patient First Visit (FPFV): The biggest challenge identified by most Tufts survey respondents was the journey from protocol to database release[8]. This often involves an extensive specifications process and, in some cases, the generation of a paper CRF. All this effort is to avoid mistakes in the build process, but at what cost? It takes, on average, more than 10 weeks to build a database. In some cases, it takes more than 14 weeks, with many companies blaming protocol complexity for delays. While complexity is a factor, misalignments within organizations which lead to multiple rounds of changes to specifications and protocol also play an important role. Overall, the time it takes to actually build an EDC is relatively short but much of the startup time is eaten up with the specification alignment process. As a result, more than 30% of companies said they often release the EDC after FPFV[8] preventing sites from entering data shortly after patient's visits. Most large companies stipulate in their processes that the EDC must be released before FPFV, or even before the first site is initiated.

While there is no regulation specifically requesting EDC to be released prior to FPFV, its availability before FPFV is critical, especially when EDC is to be used as an e-Source system for direct data capture. Sponsors may expose themselves to serious inspection findings if the EDC is determined to be a critical process enabler such as real-time safety assessment, verification of patient eligibility, drug reconciliation, etc. Companies releasing EDC after FPFV may also underestimate the downstream operational and potential data quality impact of such decisions. The study showed that, on average, companies releasing EDC post-FPFV observe an extra 5 days of delay from patient visit to the site entering the data in EDC, and up to 22 days longer database lock cycle times. Longer database lock cycle times are likely attributable to late data availability delaying data cleaning and increasing the volume of queries due to the lack of early feedback to sites from automated edit checks. Late and higher volumes of queries can also have a negative impact on site satisfaction. Perhaps most importantly, in early phase studies, this can delay the identification of safety concerns and expose patients to unnecessary risks.

Establish more flexible and efficient EDC build processes: Some reviewers of the 2017 Tufts study[8] challenged the need to build databases faster. They pointed to the frequency and prevalence of protocol changes prior to FPFV being a major risk. This has led some organizations to wait until the protocol is final before starting the EDC build process. As outlined above, the late release of EDC systems impacts the sites and overall study conduct. CDM ideally needs to develop processes that embrace flexibility, rather than adopt a process that starts with the final and approved protocol. With so many protocols subject to amendments in the age of adaptive designs, CDM has really no way out. Good practices include:

- Using standard and flexible EDC libraries accommodating multiple study designs
- Ensuring simple and user-friendly EDC design tuned to site workflows to foster direct data capture
- Leveraging industry standards (e.g., CDASH)
- Leveraging planning tools to manage dependencies and monitor activities on the critical path
- Targeting edit checks on what matters (i.e., being risk-based and focusing on critical data and processes)
- Using streamlined, fit-for-purpose and risk-based approaches to improve the costly and lengthy post-production eCRF change process

Leverage the value of all industry standards: Many companies still lack an end-to-end data standardization and integration strategy that considers all the dimensions of clinical data. It is critically important to understand that standards do not only apply to collection and transmission, but also to terminology and modeling. Failure to consider all data standardization dimension can result in process inefficiencies. Additionally, leveraging good standards not only facilitates the EDC and start-up process but also facilitates the creation of the datasets required for analyses and reporting.

Standards can be classified in four layers providing synergetic values[10]:

Data models: Conceptual, logical or physical semantic descriptions of objects and their relationships

Metadata standards: Representations of individual data concepts useful for organizing data in a database or data exchange file

Terminology standards: Representations of individual data values

Exchange standards: Schemas or file types for exchanging data between systems; the container for transporting data

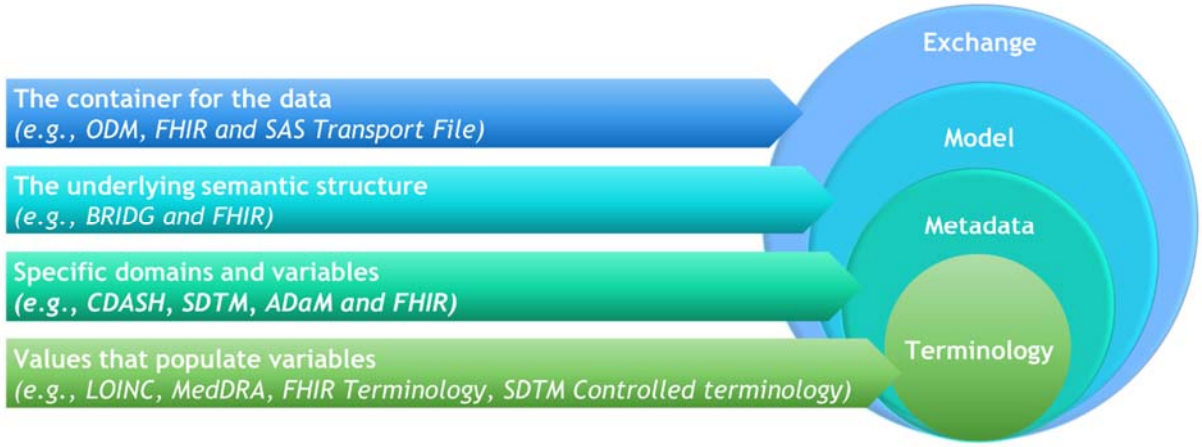

Fig 4. Layers of clinical data standards

As advocated and championed by SCDM, it is also worth noting that healthcare standards can be leveraged in clinical research to ease data mapping and exchange:

- Despite currently being at a low maturity level, the HL7 FHIR clinical research resources can be used to transfer EHR data to EDC. Aspects of these standards are interesting, as they support e-Source connections using modern technology and including the 4 layers outlined above.
- LOINC can be used to harmonize laboratory test terminology.

Few companies are exploring the use of metadata repository (MDR) systems to manage and apply all layers of standards to study-level processes to shorten the time to drug approval and drive automation efforts. Whether companies are using an MDR or other tools to manage standards, it is important to highlight that one of the key goals in managing set-up and post-production changes is to perform automated impact analysis. This allows for informed decision making, by providing more information on post-production changes and the impact of deviating from standards. Nowadays, much of the impact analysis is performed manually, which increases risk and extends timelines.