# CTEP Proposal – Samvit Section 5.4.2.4

## 5.4.2 System Integrity and Interoperability

4. Seek ways to enhance system interoperability and improve data integrity in support of

the research mission. For example, the Contractor shall ensure that standardized

terminologies (e.g., Logical Observation Identifiers Names and Codes (LOINC),

Clinical Data Interchange Standards Consortium (CDISC), NCI terms, NLM terms) are

implemented to promote data interoperability as per medical research standards.

As CTEP shifts the focus from the successful development of a comprehensive enterprise system that support the internal goals of CTEP and facilitating data submission objectives that are compliant with the Federal regulations, there is now a larger need to share the data from the enterprise systems with the other systems external to the CTEP-ESYS boundary. One of the objectives of this CTEP SOW is to seek support for integrating CTEP-ESYS to various systems from other NCI Clinical Trials Programs from the Cancer Research Infrastructure and establish new data flows and business processes. This will support access to CTEP data and enable the researchers to query across the fragmented data sets.

As new business processes are identified and established, it will be critical to define and document the data flows between the systems and then identify the metadata (data elements and the controlled terminology) that will be shared to support the research goals. The DIFZ/Samvit team has over 10 years of experience working in developing metadata standards and harmonizing semantics across various NCI and other NIH initiatives to enable interoperability. Our in-depth knowledge of the metadata associated with the NCI CBIIT programs like the CRDC and all its data commons nodes, the CTRP and the foundational semantic infrastructure of NCI caDSR and NCI Enterprise Vocabulary Service (EVS) will allow us to bring that to CTEP and support the system interoperability objectives in an efficient manner. Samvit team is recognized at CBIIT for its expertise in clinical research data standards analysis and development and being well informed about the ongoing efforts at CDISC and HL7. Below are some of the efforts where our team is doing extensive work in metadata standardization and harmonization across CRDC as well as other NIH initiatives:

1. CRDC Data Standards Team: Samvit team members lead the CRDC data standardization work at CBIIT for harmonizing metadata across the several CRDC nodes, such as Genomics Data Commons (GDC), Proteomics Data Commons (PDC), Clinical Trial Data Commons (CTDC), Integrated Canine Data Commons (ICDC) and few others. The focus of this harmonization is to establish CRDC level data standards to allow for data aggregation across the nodes. CRDC level data standards will support the [programs objective](https://datascience.cancer.gov/data-commons) of sharing, aggregating, analyzing, and visualizing cancer research data to drive scientific discovery. This involves working closely with the CRDC oncology community to identify and develop the appropriate standards and get feedback from the community. These data standards will then be registered in the caDSR as we work with the curators and use/re-use the NCI EVS concept codes and caDSR public IDs from the NCI Semantic infrastructure. One aspect of this effort is to ensure that CRDC is looking at the broader programs at NIH that have data flows for submission to CRDC and have defined metadata standards for data collection from the research community, for example, programs such as HTAN, APOLLO, etc. Aligning these with standards such as CDISC SDTM (as relevant) is also critical to support the consistent semantic representation of the data that is exchanged.
2. COVID CDEs and UMLS SemNet: Samvit team members are the metadata analysts in identification and development of COVID CDEs in the trans-NIH project to identify a single set of common data elements. This also involved identifying the LOINC codes associated with the terms. This work continues as the team works in aligning the COVID CDEs with the UMLS Semantic network (SemNet) and proposing extensions.

END OF DOCUMENT