

INTRODUCTION

Through the Surgical Critical Care Initiative Tissue and Data Acquisition Protocol (SC2i TDAP), we have collected clinical and molecular assay data for more than 1,000 patients from Walter Reed National Military Medical Center, Duke University Hospital, and Grady Medical Center. While each site follows the same protocol, interpretation and implementation can be vastly different, leading to significant differences in data collection. Here, we examine the process of transforming the TDAP dataset into the CDM.

METHODS

Assessment of the dataset was performed using the OHDSI tools White Rabbit and Rabbit-In-a-Hat. Variables are simultaneously being mapped to corresponding standard concepts and identified using Usagi and Athena. The transformation is being developed and stored in pgAdmin using PostgreSQL.

Transformation logic for each standard concept is determined in collaboration between Walter Reed and Duke University members of SC2i. This collaboration determines if a given TDAP variable is required and what transformation from the corresponding variable(s) is needed (e.g. one-to-one direct mapping of variables, many TDAP variables to one standard concept, value transformation, etc). This logic is executed using Python resulting in a PostgreSQL database for TDAP containing standardized CDM data.

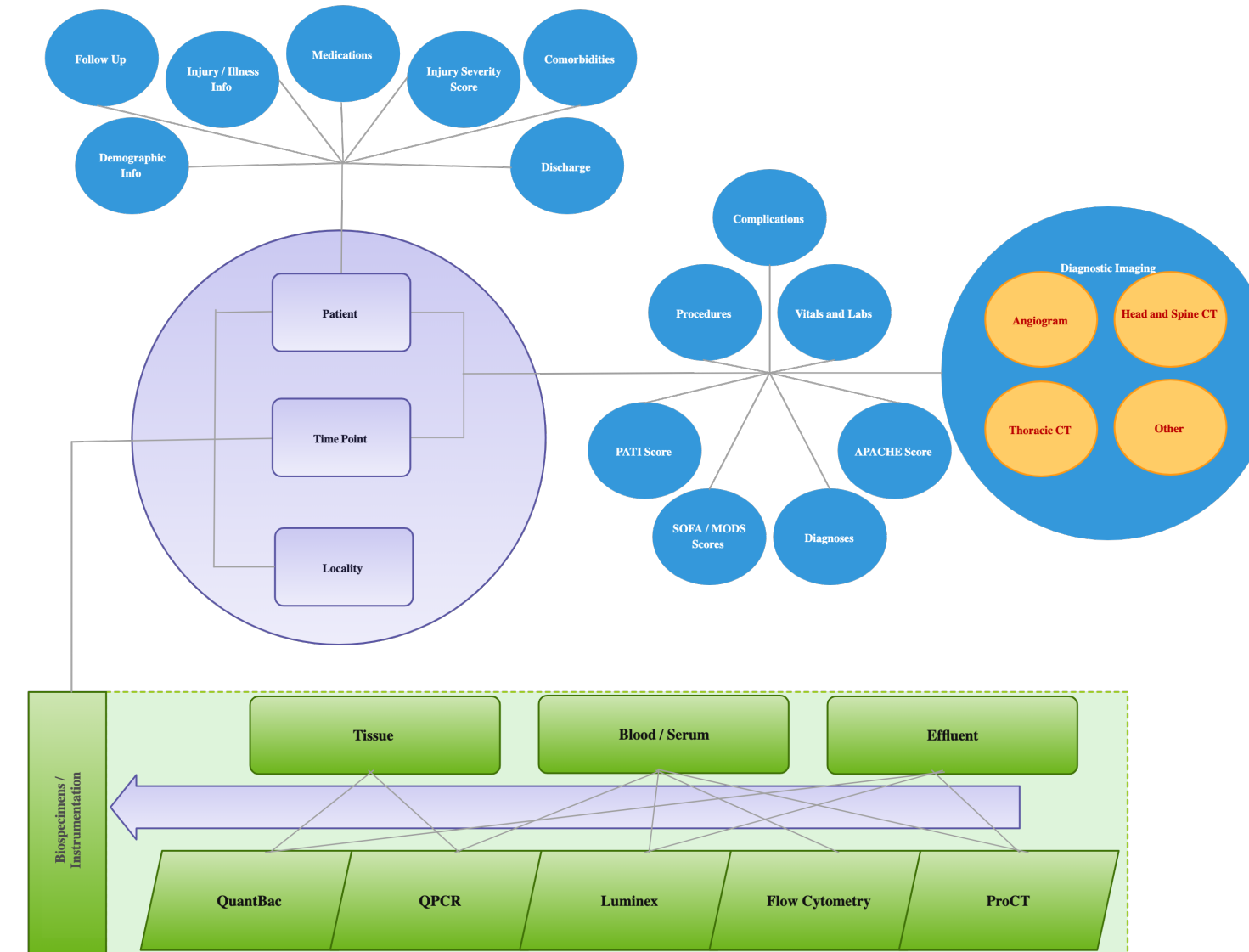


Figure 1. Schematic diagram representing the various data types and their relationships found in TDAP.

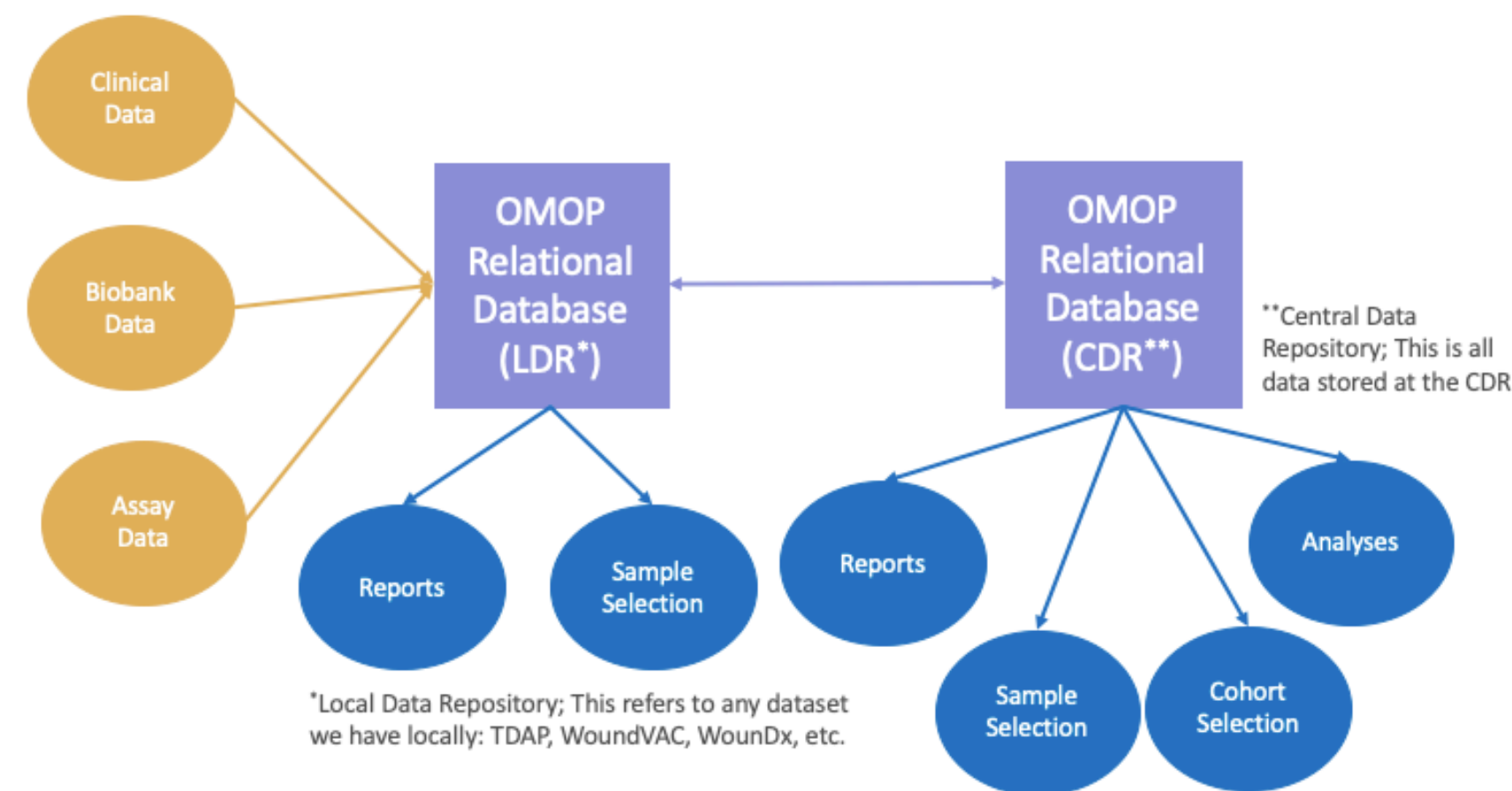


Figure 2. Schematic diagram representing the overall database goals.

RESULTS

The TDAP data dictionary contains over 11,000 variables. Our dataset as a whole comprises of over 48.7 million data elements across 1,071 patients. We have completed variable mapping for 4 tables: PERSON, VISIT_OCCURRENCE, OBSERVATION_PERIOD, and DRUG_EXPOSURE. We are in the process of identifying Standard Concept IDs for mapped variables using Usagi and Athena. This includes transformations of 31,855 RxNorm codes, 4,608 CPT codes, and 8,131 ICD codes, representing 2,300, 1,045, and 2,855 unique codes, respectively.

The Extract-Transform-Load (ETL) processes from the Central Data Repository (CDR) to the OMOP CDM database currently accepts all data as CSV files, which are downloaded from the CDR. Steps are underway to upgrade this process to allow data to be directly queried from the CDR instead, thus removing a step of human-intervention.

CONCLUSIONS

Standardizing the dataset will allow for ease of sharing of information among institutions and will lay the foundation for future collaborations. By transforming the TDAP dataset, we are one of the first institutions to provide cohesive healthcare data on military patients using the CDM, which enables us to expand collaborations, including model validations, to organizations who are also using this model.