Data Science Services

Health sciences research requires access to high quality clinical & financial data, as well as Data Scientists with a sound understanding of the healthcare & technological processes that generated the data, knowledge of the healthcare system, and the tools & methods needed to analyze complex problems. Data Science involves the extraction of knowledge from data, and the UHealth Data Science Services (DSS) aims to provide such expertise to the UHealth research community. DSS staff provide data extraction & concierge services to investigators, help them navigate the tools & processes needed to obtain data, explore alternatives, and understand the limitations of these data. The Enterprise Data Warehouse (EDW) contains data from key systems including current and historical electronic health records (EHR), labs, radiology, financial and other ancillary systems. In many cases, EDW self-service query tools can provide between 50 – 75% of the variables needed in a study, while the rest require programming. DSS staff work with investigators to ensure that analytic datasets are prepared appropriately using a combination of self-service tools, Structured Query Language (SQL), and other techniques. Finally, in addition to the actual datasets, DSS staff also provide standardized description of the data extraction methods & metadata needed for manuscripts & grants, and consultation on data related questions for individuals and research groups. Below is a list of services currently offered by DSS:

- Self-service query tools – access, training, etc.
    - Clinical Universe (CU)
    - Clinical Research Universe (CRU)
    - "Warthog" Targeted Chart Review Tool
    - Human Subjects Recruitment Tool (HSRT)
    - Population Tool
- Data concierge
    - Study feasibility, study design & data availability
    - Alternate strategies & surrogates
- Pre-research
    - Grant applications.
    - IRB/RGE applications.
    - Sample data for pre-research data requests.
- Data Extraction
    - Counts and other forms of aggregate data (CU/CRU).
    - Raw/detailed data extracts / analytic datasets
    - Utah Population Database (UPDB) linkage for UUHC patients in research datasets.
- Prospective trials
    - Identifying patients for screening – HSRT and other approaches
    - Augmenting prospective data collection with extracts from EDW.
- Vocabulary & metadata
    - Semantic enrichment of older data.
    - Basic metadata, summary of extraction & data quality report.

Contact: datascience@hsc.utah.edu