A Client provided dataset consisted of a single fact table with 436 research project records, and dimension tables for the Researcher Profile, Diagnosis Description, Ages of study participants, Practice Settings, and ICF Relation.

The Researcher Profile consisted of Name, Researcher ID, Institution, and Underrepresented Group attributes. The Research Project included Researcher ID, Project ID, Start Date, Keywords, Principal Investigator, Project Title, NIH Funding (Division, Grant #, Direct Costs, Total Funding), Other Federal Funding (Division, Grant #, Direct Costs, Total Funding), Non Federal Funding (Division, Grant #, Direct Costs, Total Funding) attributes. In addition, the Research Project Details consisted of Diagnoses, Ages, Practice Settings, ICF Category, and Research Agenda Category.

During data analysis we have uncovered a problem that an AOTA staff member entered her own name instead of the researcher’s name if she was assisting in data entry. As a result, associations between each researcher (USR\_ENTRANTS\_NAME) and his/her institution (USR\_INSTITUTION) and respective research projects that she/he was involved with (USR\_PROJECT\_TITLE) were erroneous and had to be discarded. Since the USR\_ENTRANTS\_NAME was deemed to be fatally flawed we used the USR\_PRINCIPAL\_INVESTIGATOR field instead to connect project with researchers. Furthermore, we discovered records with missing profile id that corresponded to the Institution the project was associated with. Several duplicate project records were identified and removed. Lack of specific address for the research institution was reconciled by providing geospatial coordinates based on the name of the institution.

To overcome these discrepancies, we imported the dataset into MS Access input tool and normalized it by extracting lookup tables for Diagnosis, OCF, Agenda and Age categories. The data were thoroughly cleaned and referential integrity enforced using Project ID attributes. Initial statistical analysis was performed in MS Access by creating crosstab queries for PI and Diagnosis, Diagnosis and Agenda, and Diagnosis and ICF categories. Total funding for each project was calculated by summing up NIH Funding Amount, Fed and Non Fed Funding amount. Furthermore, the dataset was imported into Tableau to provide more granular filtering of the data and more comprehensive dynamic visualizations.

The initial data source file, the final MS Access input tool, and resulting visualizations are provided through the github.iu.edu.

The final dataset provides the following statistics as in Table 1.

Table 1. Occupational Therapy dataset statistics

Research Projects 426

Research Profile 55

Practice Setting 12

ICF Area 7

Diagnosis Area 29

Agenda Category 8

Age Bracket 10

Funding years 1982 -2016