Research Development Fund – SPRING FY18
SUBMISSION DEADLINE: *March 30, 2018 at 12 noon CDT* to rdf@tamu.edu
Applications that exceed page limits for any section or do not follow template will not be reviewed

Application Title: Texas Virtual Data Library (Tx-ViDaL): A Secure and Compliant Data Infrastructure

<u>Lead contact</u> **for RDF Application:**
    Name: Hye-Chung Kum
    Department: Health Policy & Management (HPM)
    Email address: kum@tamu.edu
    Phone number: (979) 436-9439

Key Participating Units:
    School of Public Health (Dept of Health Policy & Management), TAMU Health Science Center; High Performance Research Computing (HPRC), TAMU; Texas Federal Research Data Center (TX-RDC), TAMU; Center for Remote Health Technologies & Systems (CRHTS), TEES; College of Engineering (Computer Science & Engr, Industrial & Systems Engr), TAMU; Colleges of Agriculture and Life Sciences; College of Science (Department of Statistics), TAMU; College of Liberal Arts (Department of Sociology), TAMU; Mayes Business School, TAMU; College of Education & Human Development, TAMU; College of Nursing, TAMU Health Science Center; Texas A&M Transportation Institute (TTI); Public Policy Research Institute (PPRI), TAMU; Rural & Community Health Institute (RCHI), TAMU HSC

Anticipated Request Amount ($): 1,400,000

Executive summary of this application to utilize Research Development Funds:
    Many research at TAMU require secure and compliant computing facilities when research projects involve sensitive or proprietary data. Yet currently TAMU has no good options. We propose to **(1) develop a secure cloud computing data infrastructure to support data intensive research** that involve sensitive person level data (e.g., health data) or proprietary licensed data to meet the myriad legal requirements of handling such data (e.g., HIPAA, Texas HB 300, NDA) and **(2) accumulate good data sources** (e.g., HCUP, CMS, SEER), which often need to be purchased or processed to be fit for research, to be available to researchers with appropriate approvals and permissions. This project will extend the current Texas A&M High Performance Research Computing (HPRC) – a shared computing infrastructure used by many A&M investigators and students – capacity to support an even wider user base to include those that need secure compliant computing as well as good data sources. All units with interested faculty would benefit directly from the proposed virtual computing facility because remote access from their own personal computers would be the main mode of access. It will also extend the current TX-RDC infrastructure – a shared infrastructure to access federal data – to a much wider array of data sources including Texas state data. Such a research infrastructure will enable researchers across many disciplines (e.g., public health, public policy, sociology, remote health, transportation, computer science, statistics etc.) at TAMU to develop new research agendas and collaborative team science becoming thought leaders for data intensive research in their respective fields. The infrastructure is critical for researchers to be competitive for external funding opportunities that involve sensitive personal level data. For example, for the recently funded NSF ERC, PATHS-UP, this infrastructure will be essential for securing the sensitive data collected through remote sensing devices. Many institutions are working on developing secure computing facilities for handling sensitive data and it is becoming a necessity for many fields of science that involve person level data. We anticipate that the Tx ViDaL will lead to more publications, external funding, and education of data science at Texas A&M.