<div align="center">**Research Development Fund – Cover Page**</div>

Application Title:

Texas Virtual Data Library (Tx-ViDaL): A Secure Data Infrastructure for Population Informatics

Lead contact 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 and 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
College of Science (Department of Statistics), TAMU
College of Liberal Arts (Department of Sociology), TAMU
Rural & Community Health Institute (RCHI), TAMU Health Science Center
Texas A&M Transportation Institute (TTI), TAMU

Anticipated Request Amount ($): 1,500,000

Executive summary of the intended application to utilize Research Development Funds.

The ***social genome*** is the collective footprints of our society captured in ever-larger and ever-more complex databases (e.g., government administrative data) about people in the digital society. ***Population informatics*** applies data science to social genome data to answer fundamental questions about human society and population health much like bioinformatics applies data science to human genome data to answer questions about individual health. To enable population informatics research at TAMU, 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) to meet the myriad of legal requirements for handling such data (e.g., HIPAA, Texas HB300) 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. 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 Texas A&M 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 compete competitively for external funding opportunities that involve sensitive personal level data. For example, the recently funded NSF ERC, PATHS-UP project, requires secure computing to be successful. 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 personal level data. We anticipate that the Tx-ViDaL will lead to more publications, external funding, and education of data science using person level data at Texas A&M University.