

Research Development Fund – FALL FY17 Cover Page Template

SUBMISSION DEADLINE: September 12, 2016 at 12 noon CDT to rdf@tamu.edu

(All cover pages will be posted for the campus community to view at <http://rdf.tamu.edu/abstracts>)

Application Title: Big data storage infrastructure supporting the research in life sciences.

Lead contact for RDF Application:

Name: Ivan V. Ivanov

Department: Veterinary Physiology & Pharmacology

Email address: Ivanov@cvm.tamu.edu

Phone number: 979-739-4328

Key Participating Units:

College of Veterinary Medicine & Biomedical Sciences

College of Medicine

School of Public Health

College of Agriculture and Life Sciences

College of Science

College of Engineering

Center for Translational Environmental Health Research

Center for Bioinformatics and Genomics System Engineering

High Performance Research Computing

Anticipated Request Amount (\$): \$2,000,000

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

The explosion in new high throughput data acquisition technologies continues to generate huge data sets in life sciences research. This data deluge has been a concern for research faculty and organizations across Texas A&M University for the past several years. The growing need for mid-to-long term storage of big data has been observed across all of the life sciences research activities. The situation is clearly exacerbated by the federal requirements stipulated in the applications for federal funding that experimental data and data resulting from its analyses to be retained for at least three years after submission and acceptance of the project final closeout documents (OMB Circular A-110). To our knowledge, there is no currently available infrastructure and mechanism that would provide A&M investigators access to storage for their omics or image data despite the requirements by many granting agencies. The numerous cloud-based options that are available although relatively inexpensive for storage are cost prohibitive to the majority of the PIs when they have to move data from storage over the Internet. Therefore, an onsite A&M storage infrastructure for big data is an attractive option which will clearly facilitate the research activity and will ensure compliance with granting agencies regulations. This application proposes to fund a big data storage infrastructure (> 5 Petabyte) on campus for investigator use. The proposed system will be expandable as needed and the proposed capacity is based on the current needs plus a 20X amplification factor of raw data to processed data that need to be preserved. Additional high bandwidth lines will be required to connect the storage system to key locations across campus. The housing of the data storage system by the High Performance Research Computing (HPRC) will facilitate transfer of data to the primary computing clusters on campus used for data analyses, such as Ada. Importantly, the proposed infrastructure will provide a competitive edge to future applications for federal funding which will translate to more funded projects across the broad spectrum of life sciences research at Texas A&M University.