

LG-34-19-0055-19, University of Texas At Austin (Texas Digital Library)

Preserving Sensitive Data in Distributed Digital Storage Networks

Summary

Texas Digital Library seeks a \$82,707 planning grant to develop a service model for the first ever nationally distributed digital preservation service for sensitive data. Relevant to the category *National Digital Infrastructures and Initiatives* and in alignment with the IMLS Transforming Communities Strategic Goal of building capacity, IMLS planning grant funds will support the research and data gathering needed to model a nationwide distributed digital preservation service for private and sensitive content.

National Need and Rationale

Distributed digital preservation (DDP) services have been offered in the United States for over a decade, yet there is no distributed service offering for sensitive data. Personally Identifiable Information (PII) or Personal Health Information (PHI), as well as other sensitive data in the custody of libraries, health science centers, and archives is at an escalated risk of loss. Health science libraries, especially, face a growing backlog of digital PHI governed by HIPAA¹ which requires preserving. Additionally, university-held special collections and archives are likely to have materials governed by FERPA² requirements as well as valuable cultural heritage materials that contain personal identifying information such as social security numbers. Usually only stored locally, this data is at a high risk of loss because it is excluded from services which provide the essential components of digital preservation such as geographical distribution and technological infrastructure diversity. Nearly all archives inherently hold sensitive content yet every existing DDP network bars ingestion of sensitive data. Our project seeks to create a nationwide model for a DDP service that would close these gaps in current preservation offerings for sensitive data.

Proposed Work Plan

The Texas Digital Library (TDL) and University of California, San Diego Library (UCSD) both serve as collaborating nodes in the Chronopolis and the Digital Preservation Network (DPN) distributed digital preservation systems. TDL hosts digital preservation nodes at the Texas Advanced Computing Center (TACC), which offers secure HIPAA/FERPA compliant storage³ to local partners, and UCSD provides HIPAA compliant storage to its faculty and researchers at the San Diego Supercomputing Center (SDSC). TDL and UCSD will leverage these existing relationships and resources, as well as partnerships with interested parties and consultants listed in the Personnel section below, to create a service model to later implement and share among digital preservation practitioners in the U.S. over the 12-month performance period (8/1/19-7/31/20):

¹ The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is United States legislation that provides data privacy and security provisions for safeguarding medical information.

² The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99) is a Federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

³ <https://portal.tacc.utexas.edu/user-guides/corral#access-policies-category1>

- **Q1:** Convene grant personnel to discuss needs and outline ideas for service model; begin research investigating legal agreements required between all parties involved in order to create distributed PII nodes in a DDP network;
- **Q2:** Determine technical requirements for data security as it moves between secure locations; create preliminary outline of report;
- **Q3:** Draft report and legal agreement templates; consult with HIPAA and FERPA compliance experts; conduct system and cost modeling among grant partners;
- **Q4:** Hold final meeting to refine report; design dissemination and implementation plan.

Personnel

The main project leads are: Kristi Park, Director of TDL and Primary Investigator, Courtney Mumma, Deputy Director of TDL and co-PI, and Sibyl Schaefer, the Chronopolis Program Manager for UCSD. Experienced and trusted leaders in the cultural heritage community, UCSD and TDL are founding node members of DPN. In addition, UCSD has operated the Chronopolis DDP service for over a decade and TDL is a node of that network. The Massachusetts Green High Performance Computing Center (MGHPCC), the Academic Preservation Trust (APTrust), the Smithsonian Institute, Northeastern University, the Maryland Advanced Research Computing Center (MARCC), and the Johns Hopkins University Library have all expressed interest in this project as they also have a need to geographically distribute and preserve their sensitive data. These partners would inform the service model as well as attend virtual and in-person project meetings alongside student researchers and experts in HIPAA and FERPA technical and legal requirements.

Project Goals and Sustainable Outcomes

Libraries and archives have built robust community-driven networks for preservation of all types of content *except* sensitive data. The final grant deliverables are a report modeling the establishment of a DDP service in the United States for sensitive data, and templates for legal agreements, technical requirements for data transfer, and cost modeling. The sustainable service model produced will estimate resources needed collaboratively and across time. TDL and UCSD will enhance their current DDP offerings using this service model, make the report and all findings publicly available, and share it widely with librarians, archivists, and digital preservation practitioners working in higher education organizations of every size, and directly with health science centers and the greater cultural heritage community through various existing working and advisory groups, webinars and conference presentations.

Estimated Budget

Texas Digital Library respectfully requests \$82,707 for this planning grant. We estimate \$18,997 of personnel time, including student research assistance. Consultants in data privacy law and HIPAA/FERPA compliance are estimated at \$5,500, and travel for two meetings for project participants is requested for an estimated \$16,092. UCSD will be sub-awarded approximately \$19,344 for personnel, travel, and a federally negotiated indirect cost (IDC) rate of 48%. The IDC at The University of Texas is 38% for projects of this type, equaling an estimated \$22,774.