Preserving Sensitive Data in Distributed Digital Storage Networks

The Texas Digital Library, in collaboration with the University of California, San Diego Library, seeks a one year planning grant to gather requirements and create supporting documents for a service model for the first 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, this planning grant will support the research and data gathering needed to model a nationwide distributed digital preservation service for private and sensitive content.

Although distributed digital preservation (DDP) services have been offered in the United States for over a decade, 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 HIPAA which requires preserving. Additionally, university-held special collections and archives are likely to have materials governed by FERPA requirements as well as valuable cultural heritage materials that contain personal identifying information such as social security numbers. Our project seeks to propose a nationwide model for a DDP service that would close these gaps in current preservation offerings for sensitive data.

TDL and UCSD will engage technical staff in our respective data center partners, TACC and SDSC, as well as consultants from the cultural heritage, information technology, and legal sectors, to propose a service model to later implement and share among digital preservation practitioners in the U.S. over the 12-month performance period. Project activities include convening for discussions at the beginning and end of the project, researching legal agreements and drafting legal templates, determining technical requirements in consultation with HIPAA And FERPA compliance experts, cost modeling, and producing a final report outlining the grant findings and activities.

The grant deliverables include a report modeling the establishment of a DDP service in the United States for sensitive data, templates for legal agreements, technical requirements for data transfer, and cost modeling scenarios. These deliverables will assist TDL and UCSD in enhancing their DDP offerings to include services for sensitive data and also help pave the way for other DDP services to do so as well. Additionally, they will serve as resources for the cultural heritage community during preservation planning processes and especially in evaluating service providers. The deliverables and final report will be made publicly available and shared widely with the greater cultural heritage community and health science centers through various existing working and advisory groups, webinars and conference presentations.

Preserving Sensitive Data in Distributed Digital Storage Networks

Statement of National Need

The Texas Digital Library (TDL), in collaboration with the University of California, San Diego Library, seeks a one year planning grant to gather requirements and create supporting documents for a service model for the first 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, this planning grant will support the research and data gathering needed to model a nationwide distributed digital preservation service[1] for private and sensitive content.

Although distributed digital preservation (DDP) services have been offered in the United States for over a decade, there is no distributed service offering for sensitive data. For this reason, Personally Identifiable Information (PII) or Personal Health Information (PHI), as well as other sensitive data in the custody of libraries, academic health science centers, and archives is at an escalated risk of loss. Academic health science libraries, especially, face a growing backlog of digital PHI governed by HIPAA which requires preservation. Additionally, university-held cultural heritage collections are likely to have materials governed by FERPA requirements as well as valuable cultural heritage materials that contain PII such as social security numbers and other data deemed sensitive or private based on local and jurisdictional policies. Data specialists in academic health science centers and their colleagues in university special collections, archives and libraries share a common need for sensitive data preservation. The project team posits that the requirements for DDP of content containing sensitive and confidential information will be similar across all types of content; indeed, the digital preservation services represented by the project partners are content agnostic, excluding any content which is considered sensitive for any reason. Further, the team assumes that the bar set by HIPAA and FERPA is sufficiently high to protect many other kinds of nonregulated sensitive data. The project seeks to propose a nationwide model for a DDP service that would close gaps in current preservation offerings for sensitive data for various types of institutions.

The technology, infrastructure, and expertise needed to build a DDP service for sensitive data exist, but the connections, agreements and processes to put it all together to form a viable service are lacking. Since such an offering would be the first of its kind, it would be unwise to proceed to its creation before identifying the legal, technical, and financial requirements to build it effectively. While building this service is outside the scope of this one year planning grant, the two lead institutions are interested in using the grant deliverables to assess their capacity to meet the outlined requirements and to initiate discussions with possible network partners. These future activities would ideally lead to the establishment of a national network for sensitive data protection.

Both grant partners have well-established business models and extensive experience in building and providing DDP services. The Texas Digital Library (TDL), administratively based at the University of Texas at Austin (UT), is a consortium of Texas higher education institutions that builds capacity for preserving, managing, and providing access to unique digital

[1] MetaArchive Cooperative. *A Guide to Distributed Digital Preservation.* Atlanta: University of North Texas Libraries: 2010. digital.library.unt.edu/ark:/67531/metadc12850/. Accessed March 14, 2019.

collections of enduring value. The mission of the TDL is to advance and advocate the role of digital libraries and digital scholarly communication technologies that support the research and teaching missions of institutions of higher education in Texas and to promote cooperation, communication, and resource sharing among its members.[2]

Since 2015, the TDL has also offered access to DDP storage systems.[3] Early iterations of its digital preservation services allowed members to store and manage multiple copies of data in Amazon Web Services storage locations and/or at the Texas Advanced Computing Center (TACC). In 2012, the TDL joined the Digital Preservation Network (DPN) and worked in partnership with UT Austin and TACC to build and launch in 2016 one of four production nodes in that network, which ceased operations in 2019.

In 2017, the TDL joined the Chronopolis DDP network headquartered at the University of California San Diego Library, providing access to Chronopolis services to its members (via TDL's DuraCloud implementation) and serving as a replicating node for the network using storage at TACC.[4]

The University of California, San Diego Library manages the internationally-recognized DDP service, Chronopolis. The Chronopolis network spans four sites across the United States and is one of the earliest established DDP services in the world, having been in operation for over 11 years. The UCSD Library partners with the University of Maryland Institute for Advanced Computing Studies (UMIACS), the National Center for Atmospheric Research (NCAR), and the TDL to maintain geographically distinct data centers, which are referred to as "nodes." Chronopolis offers preservation storage through the DuraCloud and Texas Digital Library services. It was certified as a Trusted Digital Repository by the Center for Research Libraries in 2012 and plans to undergo ISO 16363 certification.[5]

Both project partners maintain close working relationships with organizations affiliated with their home institutions that could provide key resources for a DDP service for PII and PHI. Both TACC, located at UT Austin, and the San Diego Supercomputer Center (SDSC), affiliated with UCSD, offer protected data storage; while these individual storage locations do not constitute a geographically distributed digital preservation service, they could serve as essential components of as the project team works to model a best-practice DDP network for PII.

In carrying out their missions and supporting their current members, both the TDL and Chronopolis have observed that data containing Personally Identifiable Information (PII) or Personal Health Information (PHI), as well as other sensitive data managed by libraries, academic health science centers, and archives, are at an escalated risk of loss. Consultations with TDL member university libraries and archives reveal that more than half of all 22 member institutions have sensitive data content which requires digital preservation actions. While some digital preservation actions can be performed successfully onsite by digital archivists and

[2] "Texas Digital Library Bylaws." https://www.tdl.org/wp-content/uploads/2018/05/TDLBylaws_201805.pdf. TDL.org. Accessed March 16, 2019.

[3] "Announcing DuraCloud @TDL for digital preservation." TDL.org. November 12, 2014. https://www.tdl.org/2014/11/announcing-duracloud-tdl-digital-preservation/

[4] "Texas Digital Library Joins Chronopolis Digital Preservation Network." May 11, 2017. https://www.tdl.org/2017/05/texas-digital-library-joins-chronopolis-digital-preservation-network/

[5] International Organization for Standardization. *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories.* ISO 16363:2012 (CCSDS 652.0-R-1). Accessed March 14, 2019. https://www.iso.org/standard/56510.html. ISO 16363 is the highest level of digital preservation certification available.

librarians, the lack of a sensitive data DDP service was identified as a significant gap for these institutions. As a result, data are at a high risk of loss because they are usually only stored locally and rarely replicated elsewhere; thus, these data are excluded from services which provide the essential and standards-based components of digital preservation such as geographical distribution. Sensitive data can be found in almost all archives and is prevalent in many cultural heritage organizations but because of the legal and technical complexities involved in preserving such data over a network of providers no existing non-profit DDP network currently provides a HIPAA/FERPA compliant preservation service.[6]

DDP storage for health data, student education records, and other sensitive data will benefit the national population at large by protecting from the loss, degradation, and exposure of records essential to public health advancement, human welfare, and sociopolitical stability. This project seeks to propose a nationwide model for a DDP service that would tackle the complexities of private and sensitive data preservation, clarify the legal and technical requirements needed, and close these gaps in current preservation storage offerings for sensitive data.

The collaborative and mutually beneficial work proposed in the project aligns with the Digital Preservation Declaration of Shared Values, which both Chronopolis and TDL representatives helped to draft and signed. This planning grant to develop a service model for the first nationally DDP service for sensitive data is especially relevant to the category *National Digital Infrastructures and Initiatives* as it brings together various experts and advisors to develop resources which will help establish new digital preservation services for library collections. Further, in alignment with the IMLS Transforming Communities Strategic Goal of building capacity, IMLS planning grant funds will support the research needed to model a nationwide DDP service for private and sensitive content. The TDL and UCSD are uniquely qualified to lead this research given their extensive experience in DDP service design and expertise in reviewing and assessing the needs of data depositors.

Project Design

Project goals

The goal of this proposed grant is to establish a service model which will lay the groundwork for establishing a DDP for sensitive data and will include: templates for the legal agreements that will need to be in place between the participating nodes, their respective data centers, and the institutions depositing data; a list of the technical requirements to ensure data security and integrity as it transfers between secure locations; and finally cost models to outline the initial and ongoing resources participating nodes will have to allocate to provide this service.

[6] The Academic Preservation Trust (APTrust) does not have HIPAA/FERPA certification. It uses Amazon commercial services exclusively and does not accept sensitive data unless it is encrypted to standards that meet its depositors' own individual institutional requirements for handling of such data. APTrust will allow ingestion if the depositors encrypt it themselves, and APTrust does not hold the keys to decrypt the content.

Personnel

Project Director (Lead): Kristi Park

Kristi Park has been the Director of the TDL since 2015 and has worked with TDL since 2009. She will serve at 1% time as co-PI. As project director, Kristi will provide overall direction of the project, supervising the project manager and grant-funded student employee. She will also serve as the convener of project meetings and as a key participant in dissemination of project outcomes.

Project Manager (Lead): Courtney Mumma

Courtney Mumma is the Deputy Director of the TDL. She has served as a services manager and deputy director since 2017 and will contribute 5% of her time as co-PI. Courtney will serve as the project manager, leading and assigning all of the major grant activities. She will work directly with the major partners, including TDL health science members, consultants, and student researcher, as well as TDL member libraries with an interest in preserving sensitive and restricted data. Courtney will assist in collecting and synthesizing legal, technical, cost, and other resource requirements for the transfer and preservation of such data, incorporating direct consultation with the Texas Advanced Computing Center (TACC).

Project Lead: Sibyl Schaefer

Sibyl Schaefer is the Chronopolis Program Manager at the UCSD. She will dedicate 5% of her time to the project including assisting with all of the major grant activities. She will work with UCSD Health Sciences researchers to determine their requirements regarding the storage and preservation of sensitive and restricted data and consult with SDSC to outline the technical requirements and related costs for providing preservation services for sensitive and restricted data.

Graduate Student Research Assistant

The Graduate Student Research Assistant must be in good academic standing, and demonstrate initiative, effective verbal and written communication skills, and the ability to work independently. The GRA must have a keen eye for detail and accuracy. Preferred qualifications include: Familiarity with digital preservation, technical requirements gathering, legal concerns of managing sensitive data or personally identifiable information, and knowledge of higher education or cultural heritage organizations. The GRA will work at 50% time (20-30 hours per week) on investigating legal agreements required between all parties involved in order to create distributed PII nodes in a DDP network, gather technical requirements for data security as it moves between secure locations, as well as assist with other grant administration duties.

Invited participants

Representatives from the Academic Preservation Trust (APTrust), the Smithsonian Institute, Northeastern University, the University of North Texas Health Sciences Center, the University of Texas Southwestern Medical Center, the Dell Medical School at the University of Texas at Austin, and the Maryland Advanced Research Computing Center (MARCC) at 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. The project team will contract with SecurityMetrics[7], experts in HIPAA compliance, to provide templates for policy creation and review final deliverables. These representatives, in addition to project leads and contracted legal counsel, may attend the in-person and virtual project meetings to contribute their technical requirements and organizational expectations for a sensitive data DDP service. Their contributions will inform the service model and be captured as part of the technical requirements. Draft versions of the legal agreements, technical requirements, cost model, and report will be shared with this group prior to a virtual meeting at the end of the project. Feedback will be incorporated into the final versions of these documents.

The TDL and UCSD will work closely with staff in their respective partner data centers, the Texas Advanced Computing Center (TACC) and San Diego Supercomputer Center (SDSC). As part of the University of Texas at Austin, TACC[8] designs and operates some of the world's most powerful computing resources. TACC has partnered with the TDL on several projects over the years and currently provides storage resources for the TDL's Chronopolis node. TACC independently offers secure HIPAA/FERPA compliant storage[9] to local partners. Similarly, SDSC provides HIPAA compliant storage to its faculty and researchers at the UCSD.[10] Both computing centers will share their expertise in providing sensitive data solutions, including service and cost modeling, with the project team.

Timeline

ACTIVITY	DETAILS	START	END
Project Lead Meetings	Project leads meet online every two weeks for planning, problem solving, and keeping on task and on time.	9/1/19	8/1/20
Hire GRA	Project leads hire GRA and orient them to the project.	9/1/19	10/1/19
Gather Data	GRA and project leads: ● Gather data about legal issues impacting distributed PII nodes in a DDP network. Includes, among other activities, interviewing relevant personnel at TACC and SDSC about legal issues encountered in developing non-distributed storage for PII. ● Collect use cases about private and sensitive content from invited participants and current TDL and Chronopolis members. ● Collect existing contracts for non-distributed HIPAA and FERPA storage at UCSD and TACC. ● Document costs of non-distributed HIPAA and FERPA storage at UCSD, TACC, and others as identified. ● Collect existing contracts for DDP storage provided	9/1/19	11/1/19

[7] SecurityMetrics. https://www.securitymetrics.com/. SecurityMetrics website. (Accessed March 14, 2019). A quote for legal services from Security Metrics is included with this proposal.
[8] The Texas Advanced Computing Center. https://www.tacc.utexas.edu/ TACC website. (Accessed March 14, 2019).
[9] The Texas Advanced Computing Center. "TACC User Portal." https://portal.tacc.utexas.edu/user-guides/corral#access-policies-category1 TACC website. (Accessed March 14, 2019.
[10] San Diego Supercomputer Center. https://www.sdsc.edu/. SDSC website. (Accessed March 14, 2019.)

ACTIVITY	DETAILS	START	END
	by Chronopolis and TDL. Includes agreements between participating Chronopolis nodes and between depositing institutions and their service provider (TDL or Chronopolis). ● Identify and document areas of concern in existing Chronopolis and TDL contracts as they relate to HIPAA and FERPA content. ● Document or gather existing documentation about personnel and workflows for ingest, storage, and replication for DDP services provided by TDL, Chronopolis, and others as identified.		
Determine Technical Requirements	Project leads, GRA, SecurityMetrics (HIPAA/FERPA consultants), and TACC & SDCC: ● Via phone, web conferencing, and local meetings amongst Austin-area partners: ○ Identify technical (e.g. ingest, transfer, encryption, etc.) and other requirements (eg. physical security, access protocols, etc.) for HIPAA/FERPA compliant data security as it moves between institutional users, services, and storage providers.	10/1/19	5/1/20
Meet In Person	Project leads, GRA, and invited participants: ● Convene in Austin, TX. ● Project leads present information about existing storage services, including non-PII distributed digital preservation services provided by TDL and Chronopolis, as well as PII-compliant (but non-distributed) storage at TACC and SDSC. ● Discuss information compiled to-date in “Gather Data” and “Determine Technical Requirements” stages above, including institutional use cases. Discuss and expand on use cases. ● Identify the gaps between existing services and a distributed digital preservation service for PII. ● Outline components of a service model, including a service proposition; key actors in design and governance; user interaction; technology and human resources needed for service delivery; associated costs; and metrics for evaluating performance.[11]	11/1/19	12/15/19
Analyze Data	Project leads and GRA:	12/15/19	3/1/20

[11] Turner, Neil. “Introducing the service model canvas.” UX for the Masses. http://www.uxforthemasses.com/service-model-canvas/ (accessed March 13, 2019). While this resource is intended for UX-centered design processes, it may be adapted for use in outlining service model components for the purposes of this project.

ACTIVITY	DETAILS	START	END
	● Via phone, email, and regular project meetings ○ Analyze all data gathered and information from the in-person meeting. ■ Is the information comprehensive, accurate, and understandable? ○ Research and outreach as-needed to fill in gaps in data if identified.		
Draft Final Report	Project leads and GRA: ● Compile requirements based on the data gathering up to this point for the templates for legal agreements needed between all parties who might participate in a DDP network for private and sensitive data. ● Compile requirements for the service and cost models. ● Compile technical and other requirements. ● Outline and draft essential components of the final project report.	3/1/20	7/1/2020
Draft Contract Templates	Project leads working with SecurityMetrics HIPAA/FERPA consultants: ● Create draft template legal agreements between: ○ participating data center nodes, TDL/TACC and UCSD/SDSC; ○ depositing institution and service provider (*e.g. between UT Southwestern Medical Center and TDL);* ○ service providers and the data center's storage facilities providing HIPAA/FERPA compliant storage (*e.g. between TDL & TACC, UCSD & SDSC*).	5/1/20	7/1/20
Draft Service and Costs Models	Project leads will draft service and cost models based on information gathered and feedback from the in-person meeting.	5/1/20	7/1/20
TCDL	TDL staff will share project progress at Texas Conference on Digital Libraries (TCDL).	5/1/20	6/1/20
Disseminate Drafts	Project leads will share drafts of the final report and templates with invited participants for their initial review. ● Share as available and collect feedback about the	6/1/20	8/1/20

ACTIVITY	DETAILS	START	END
	project report, legal templates, service and costs models via Google Docs, email, phone, and the virtual meeting (see below). ● Discover whether the legal templates are viable in institutions (including at least one academic health science center and one cultural heritage collection). ● Submit legal templates to the UT and UCSD contracts/legal offices for review.		
Meet Virtually	Project leads, TACC representative, and invited participants: ● Convene a virtual meeting. ● Discuss feedback on all drafts (final report, service and cost models, and templates for legal agreements). ● Review and refine the documents. ● Discuss the dissemination and implementation plan.	8/1/20	9/1/20
Revise and complete final report and all templates	Project leads will revise and complete the final report (with service and costs models) and all templates.	8/1/20	10/1/20
Disseminate	Project leads: ● Share project details, final report and outcomes at conferences and via TDL webinars. ● Publish report and templates on tdl.org and in the TDL Institutional Repository, and distribute via social media, email, and listservs.	5/1/20	11/1/21

Assumptions and risks

As with any project, there are certain assumptions and risks inherent in the proposed plan. The most potentially impactful risk is that the graduate research assistant is not hired in a timely fashion. The project has been shifted back from an original August, 2019 start date to a September, 2019 start date to allow additional time for this placement. The project leads are also willing to undertake some of the necessary research and administrative tasks to ensure the project is not significantly delayed should this issue arise.

Another risk is that the legal counsel required for this project will cost more than outlined in the budget. A quote for legal services is attached to this proposal to justify the budget line. Should this quote balloon unexpectedly, the project leads will request additional legal assistance from legal resources on campus.

Two assumptions were posited at the beginning of this proposal. One is that because both the TDL and Chronopolis offer services that don’t differentiate based on content type, (i.e., health science data vs. cultural heritage materials), separate sensitive data services for different content types will not be necessary. While this assumption could prove false, the actual preservation service offered is bit-level preservation. Any migration processes or additional metadata collection — two preservation services where content type would matter — are the responsibility of the data provider.

The second assumption is that meeting the requirements for HIPAA and FERPA storage is sufficiently thorough to protect all sensitive data. These regulations provide guidance on what is needed to secure sensitive information and have also resulted in an industry designed around their compliance. This project will benefit from using the structure provided by the regulations as well as industry expertise.

Project evaluation

During their meetings every two weeks, the project team will assess the project's status, whether objectives are being met in a timely manner, and whether any disruptive issues have arisen. Should there be a problem requiring adjustment, the team will identify the steps needed to course correct and communicate them to anyone essential to accomplishing those goals.

Drafts of all project components -- including final project report, legal agreements, technical requirements, and service model -- will be disseminated to invited participants for review and feedback on perceived utility and feasibility, as well as on compliance with HIPAA/FERPA guidelines. As noted elsewhere, this group of participants includes HIPAA/FERPA compliance experts (SecurityMetrics), storage providers with experience providing PII-compliant storage, and potential end users of the planned service. The project team will also submit any legal contracts to the relevant offices of UT Austin and UCSD for review. Feedback will be incorporated into final versions of the project report and outputs.

Dissemination

The TDL and UCSD intend to make the report and all findings publicly available, and share it widely with information technology professionals, librarians, archivists, and digital preservation practitioners working in higher education organizations of varied sizes, as well as directly with academic health science centers and the greater cultural heritage community through various existing working and advisory groups, webinars, and conference presentations. The TDL's health science members UT Medical Branch and UT Southwestern Medical Center, as well as other academic health science centers identified during the course of the grant, will be provided detailed information about the grant findings and consultation from the TDL project team about next steps. Grant leads will share the report and findings with the National Digital Stewardship Alliance (NDSA) Infrastructure and Standards & Practice Interest Groups. The TDL will engage its own Digital Preservation Services User Group, comprised of representatives from six member institutions, which meets once per month, in discussions about the grant as it progresses.

The TDL offers free, public webinar series, and intends to use their webinar platform to disseminate the findings of the planning grant. These webinars will be recorded, closed captioned, and made available in the TDL DSpace repository[12] and on TDL's YouTube channel[13]. Additionally, TDL and UCSD project leads intend to propose presentations about the grant's findings to iPRES, Open Repositories, the NDSA's Digital Preservation, the Texas Conference on Digital Libraries, CNI, and engage opportunities for presenting to other digital

[12] Texas Digital Library. TDL DSpace Repository: https://tdl-ir.tdl.org/. (accessed March 14, 2019).

[13] Texas Digital Library YouTube Channel. https://www.youtube.com/user/texasdigitallibrary. (accessed March 14, 2019).

preservation, library, and archives conferences as they come to light. The TDL and UCSD project leads will also support and encourage project participants to disseminate broadly in their communities.

National Impact

Libraries and archives have built robust community-driven networks for preservation of all types of content except sensitive data. This project will lay the groundwork for transforming this practice to include this valuable and at-risk data at the national level. The final grant deliverables include a report modeling the establishment of a DDP service in the United States for sensitive data, templates for legal agreements, technical requirements for data transfer, and cost modeling scenarios. The service model produced will estimate resources needed collaboratively and across time. These deliverables will ideally assist the TDL and UCSD in enhancing their current DDP offerings using this service model.

With one of the highest priority deliverables being a service model, the project teams intend to combine lessons learned from successful models like DuraSpace, the TDL, and Chronopolis as well as less successful ones like the recently defunct Digital Preservation Network to ensure the service model is realistic and aids in sustainability planning for future DDP sensitive data storage services. Cost modelling will allow the team to offer transparency in pricing and clarity about the resources needed to support institutions who need to preserve this specific content niche. Additionally, the project deliverables, shared broadly, will serve as resources for the cultural heritage organizations during their preservation planning processes and especially in evaluating services providers. In most cases, institutions will not have previously had the experience of working with service providers to preserve their sensitive content. The documentation of the legal, technical, and resource allocation requirements will serve to help institutions make informed choices. By bringing in a variety of participants and incorporating their feedback and perspectives into the final documents, the deliverables should be readily adaptable by other organizations. Although many of the participating organizations are of the cultural heritage realm, assimilating the requirements of local campus health sciences schools partnered with UCSD and the TDL should extend the applicability of the deliverables to other communities as well.

It is the intent of the project team that the project deliverables support the design and implementation of DDP storage networks for sensitive data, within the TDL and UCSD as well as with other networks and institutions outside of the project team and its partners.

ACTIVITES	SEP 2019	OCT 2019	NOV 2019	DEC 2019	JAN 2020	FEB 2020	MAR 2020	APR 2020	MAY 2020	JUNE 2020	JULY 2020	AUG 2020
1 | Project Lead meetings online every two weeks throughout project period.	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
2 | Hire GRA; train and orient to project.	🟩	🟩										
3 | Gather data about legal issues; collect use cases and existing contracts; document areas of concern, costs, and current workflows.		🟩	🟩									
4 | Determine technical requirements.		🟩	🟩	🟩	🟩	🟩	🟩	🟩				
5 | Convene grant personnel and advisors (one-day, in-person meeting in Texas; will take place between 11/1/19 and 12/15/19).			🟩	🟩								
6 | Analyze all data gathered and information from the in-person meeting. Research and outreach as-needed to fill in gaps in data if identified.				🟩	🟩	🟩						
7 | Draft final report.							🟩	🟩	🟩	🟩		
8 | Draft template legal agreements.									🟩	🟩		
9 | Present project progress at Texas Conference on Digital Libraries, May 2020.									🟩			
10 | Share drafts of the final report and templates with invited participants for their initial review.										🟩	🟩	
11 | Convene project leads, consultants, and invited participants to discuss feedback on all drafts; review and refine the documents; and discuss the dissemination and implementation plan. (Virtual meeting will take place between 8/1/20 and 9/1/20.)												🟩
12 | Complete final report (with service and cost models) and templates.												🟩
13 | Dissemination: Share project details, final report and outcomes at conferences and via TDL webinars; publish report and templates on tdl.org and in the TDL Institutional Repository; and distribute via social media, email, and listservs.									🟩	🟩	🟩	🟩

Schedule of Completion

DIGITAL PRODUCT FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (e.g., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

All applications must include a Digital Product Form.

□ Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

Part I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

The report, including any templates for legal agreements between the participating nodes, their respective data centers, and the institutions depositing data, a list of the technical requirements to ensure data security and integrity as it transfers between secure locations, and details about the service and cost modelling outline of the initial and ongoing resources participating nodes will have to allocate to provide this service will all be licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0/).

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

Under the Creative Commons Attribution-ShareAlike 4.0 International License, the project team offers all materials in hopes that they will be used as a basis for service-building for private and sensitive data in the digital preservation community. As such, the project team only requires citation of their work in any derivative products.

A. 3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

This grant will not produce any products with privacy concerns, or cultural sensitivities, or requiring obtaining permissions or rights. The grant products will help establish the foundation for a network to handle sensitive data, but the grant itself is not collecting, disseminating, or preserving any such data.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and the format(s) you will use.

The project team will produce a PDF report, including any templates for legal agreements between the participating nodes, their respective data centers, and the institutions depositing data, a list of the technical requirements to ensure data security and integrity as it transfers between secure locations, and details about the service and cost modelling outline of the initial and ongoing resources participating nodes will have to allocate to provide this service.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

Drafts of the final report and its attachments will be shared across the project team using Google Docs. The final report will be transformed into a PDF. Any conference presentations related to the grant will also be accessible as PDFs in Texas Digital Library’s DSpace Repository (https://tdl-ir.tdl.org/). TDL also offers free, public webinar series, and intends to use their webinar platform to disseminate the findings of the planning grant. These webinars will be recorded as MP4 and made available in the TDL DSpace repository and on TDL’s YouTube channel (https://www.youtube.com/user/texasdigitallibrary).

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

PDF, MP4 (youtube videos), .SRT (caption files for videos), DOC, ODT, PPT

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

The project team will use Google Docs to manage documentation sharing and drafts across a distributed team. Google Docs will allow flexible versioning and access controls. The final report documents and any presentations created in Google Docs will be converted to PDFs and ultimately stored in a collection named for the planning grant in the TDL DSpace Repository.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

The final report, addenda, and all presentations and videos related to the grant will be made available in TDL's DSpace repository and preserved in Chronopolis.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

DSpace uses the Dublin Core metadata structure. Metadata content in DSpace includes: Authors (including ORCID capture), Contributors, Title, Date issued, Type (presentation, article, image), Language, Description, Abstract, Publisher, and Keywords. DSpace metadata is available in all of the formats listed here: https://tdl-ir.tdl.org/oai/request?verb=ListMetadataFormats. All of the formats listed are available via crosswalks in the DSpace configuration: https://github.com/DSpace/DSpace/tree/dspace-6_x/dspace/config/crosswalks/oai/metadataFormats. TDL has regularly used OAI_DC and METS/MODS.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

Metadata related to the final report, its addenda, and all presentations and videos related to the grant will be made available in TDL's DSpace repository and preserved in Chronopolis.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

TDL will use DSpace's built-in OAI-PMH server (https://tdl-ir.tdl.org/oai/) to facilitate programmatic access to the metadata of all digital content created during this project. TDL will also provide LOD access using Fuseki on TDL's DSpace repository, populating the triplestore and making the metadata available in RDF format.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The final report, addenda, and all presentations and videos related to the grant will be made available in TDL’s DSpace repository (https://tdl-ir.tdl.org/) and preserved in Chronopolis. Videos of any webinars will be closed captioned and openly accessible online on the TDL YouTube Channe (https://www.youtube.com/user/texasdigitallibrary/).

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

January 2019 TDL Member Forum: https://youtu.be/q-TDmjApL3Q
TDL Digital Library Archives: https://tdl-ir.tdl.org/handle/2249.1/67069

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

NA

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

NA

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

NA

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

NA

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

NA

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

NA

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

NA

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

NA

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

NA

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

NA

URL:

NA

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

NA

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

NA

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

NA

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

NA

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

NA

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

NA

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

NA

A.8 Identify where you will deposit the dataset(s):

Name of repository:

NA

URL:

NA

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

NA