LIBRARY COOPERATION GRANT 2017

Project Title: Development of a Statewide Aggregation Service for Texas Digital Collections
Project Location: Austin, Texas
Project Director: Kristi Park, Executive Director, Texas Digital Library (UT Austin Libraries)
New program? Yes
Project partners: UT Austin Libraries (Texas Digital Library), University of North Texas Libraries, Houston Public Library

Abstract: The purpose of this project is to increase the visibility of Texas digital collections through statewide aggregation and inclusion in the Digital Public Library of America (DPLA). The proposed service, developed and sustained by a partnership of academic and public libraries, will increase access to valuable Internet resources stewarded by Texas libraries and archives, but which may be underutilized currently. Because the development of a fully functioning statewide aggregation hub is a 2-year project, we propose to focus on 3 objectives in the first year: 1) needs assessment, 2) sustainability planning, and 3) technology development, concluding with a demonstration project to aggregate collections from a select group of institutions.

Project Partners: This grant proposal involves three Project Partners: UT Austin Libraries (through the Texas Digital Library), the University of North Texas Libraries, and the Houston Public Library. The TDL is a consortium of 22 academic libraries that builds capacity for preserving and providing access to unique digital collections of enduring value. Its role in the project is to serve as the grant administrator, provide overall coordination through the project PI Kristi Park, and provide the bulk of staff time to developing and operating an aggregation hub. The two additional partners represent important public and academic library constituencies and will serve on a Steering committee to develop a governance and financial model for the hub.

Needs Assessment

While many collections of value are already made available to the Digital Public Library of America through DPLA's relationship with the University of North Texas' Portal to Texas History, a great many more are missing from this widely used and prestigious Internet resource. This project will work to create a *comprehensive* aggregation service of Texas collections to ensure that *all* collections of enduring value from Texas public, school, special, and academic libraries (as well as many cultural heritage organizations and historical societies) have a pathway to the large-scale collection of the DPLA.

All digital collections in the state of Texas should be available through the Digital Public Library of America (DPLA), which provides access to more than 10 million items through its aggregated metadata repository. The DPLA is an all-digital library that aggregates metadata and thumbnails for millions of photographs, manuscripts, books, sounds, moving images, and more from libraries, archives, and museums around the United States. It is a national discovery tool for

public domain and licensed digital content held by the nation's archives, libraries, museums, and other cultural heritage institutions. It is aimed at providing large scale public access to these local depositories.

It has been demonstrated repeatedly that Inclusion in the DPLA leads to more visibility and usage of these valuable materials, offering contributing institutions the opportunity to reach more users, increase access to their content, and collaborate in new ways. For example, since joining DPLA, the Minnesota Digital Library has seen a 55% increase in visits and 62% in unique visitors. The Mountain West Digital Library has seen a bump of 105% in visits and 109% in unique visitors. This traffic is passed on to the contributing institutions that have shared content with these statewide partners (Source: http://dp.la/info/hubs/)

It is important to note that the DPLA works through a "hub" model -- a national network out of more than forty state/regional digital libraries and myriad large digital libraries in the US that aggregate and feed metadata about digital objects into the single database and portal that is the DPLA. Currently the DPLA model allows for only one aggregation "hub" per state.

In Texas, the Portal to Texas History has served as this hub for Texas collections. Hosted by the University of North Texas, this well-respected digital gateway consists of more than 650,000 digital items from over 300 partners. It hosts digital content from multiple contributing partners, and, in many cases, its staff digitize physical items, create metadata, process, upload, and digitally preserve the content.

However, while the Portal serves a very large number of archival collections -- and does so with exemplary quality -- its model does not accommodate every institution wanting to have its collections accessible through the DPLA. Out of 92 four-year academic libraries that are part of the Texas Council of State University Libraries and Texas Independent College and University Libraries, 41 (less than half) are represented in the Portal to Texas History holdings. Several academic libraries with very large digital collections of immense value are not represented at all, including Baylor Libraries, Texas A&M Libraries, and Texas Tech Libraries.
The same gap exists for public libraries: 94 public libraries are represented in the Portal to Texas History collections, compared to 558 public libraries statewide. (Source: Texas Public Libraries Annual Report for 2014) Similarly, many large museums and other collecting institutions (including the Harry Ransom Humanities and Research Center at UT Austin) do not contribute materials.

Furthermore, even in cases where an institution has contributed materials to the Portal, those items may represent a fraction of that institution's overall digital collections. For example, the Houston Public Library's Special Collections Division has contributed to two Portal collections (The John J. Herrera Papers and The Mexican American Family and Photo Collection); however, the Houston Area Digital Archives contains more than 10,000 items not included there. Likewise, the Portal contains a small collection from the Alexander Architectural Archive at UT Austin, but thousands of other items from the UT Digital Repository, the Benson Latin American collections and others are not included.

Institutions choose not to contribute content to the Portal to Texas History for a number of reasons, including that they wish to host and control the content themselves, or that they believe their digital collections are not relevant to a perceived collections focus of the Portal on "Texas history." A comprehensive aggregation service should accommodate these varying use cases, in which institutions wish to host their own content or use an alternate service.

Discussions of the need for a truly statewide solution have been on-going for several years. Informal assessments of TDL members have indicated a need for an alternative to the Portal to Texas History pathway, and the DPLA has received requests from Texas institutions that want to participate in DPLA but cannot under the current model. Some of these discussions culminated in a meeting of the current project partners at the Texas State Library and Archives Commission in the fall of 2015, to discuss a path forward to meet that need. This proposal is, in part, a result of those discussions and the shared realization that a broader solution to DPLA-related content aggregation must be found.

An additional beneift to developing a comprehensive aggregation service for DPLA will be the ability to create a single searchable metadata repository for Texas collections, which will make visible the collective wealth of the many and varied collections of Texas libraries, archives, and other collecting institutions. The creation of a statewide aggregation hub, with a web-based discovery interface, will allow Texas to showcase these collections in aggregate and provide greater visibility to the work that Texas libraries as a whole are doing to preserve and provide access to their collections of enduring value.

Program Design

We see the development of a fully functioning statewide aggregation hub as a two-year project, with the first year devoted to achieving three goals (1) research and needs assessment, (2) sustainability planning and (3) technology development, culminating in a demonstration project that aggregates collections from a select group of institutions and feeds them to the DPLA.

Some preparation work to develop this service will commence prior to the start of the funding period and will be undertaken at the expense of the Texas Digital Library. This will include several public meetings to raise awareness about the proposed program, identify institutions interested in participating in the aggregation service, and gather feedback from these stakeholders on what would make a successful service for them.

Program Objectives:

Objective #1: Develop comprehensive inventory of major digital (archival) collections in Texas and perform gap analysis to determine which collections are not currently included in the DPLA.

Achieving a broad understanding of the nature and scope of Texas digital collections across the state is essential to a full-scale roll-out of an aggregation hub. To achieve this, a graduate research assistant (GRA) from the UT School of Information will conduct a survey of Texas collecting institutions and do additional research to create an inventory and analysis of the digital collections of Texas. The Texas Digital Library will recruit and hire a graduate research assistant in the summer of 2016, in anticipation of beginning grant-funded work on September 1.

This inventory will include, among other information, the following elements:
1) Technology platforms used
2) The harvesting capabilities of those platforms (e.g. whether they are OAI-PMH compliant)
3) Metadata standards used and gaps in metadata implementation across collections
4) The size of collections (i.e. number of items),
5) Presence or absence of collections in the Portal to Texas History

The outcome of this inventory will be a useful dataset for future work beyond the initial demonstration project, and will include the raw data collected as well as analysis. This analysis will identify institutions and collections most ready to participate in an aggregation service, in anticipation of a phased implementation over the following several years.

The project PI (the Director of the Texas Digital Library) will supervise the graduate research assistant engaged in this work.

Objective #2: Develop a sustainable model for operations and governance for a statewide DPLA aggregation hub.

While project participants are committed to continuing the development and on-going operation of the service beyond the project year, a significant objective of the project will be the development of a sustainable operational and financial model for the service. This includes determining where responsibility for elements of the service reside, how new collections will be on-boarded, how the service will be governed, what the short- and long-term staffing needs are for maintaining the project, and how funding, beyond any provided by project partners, will be secured.

The Project Partners will address this need through the development of a sustainability plan, developed over a series of 4 meetings of the Project Steering Committee, which will consist of the leads from each partner institution (Kristi Park from the Texas Digital Library, Dr. Martin Halbert from UNT Libraries, and Judith Hiott from the Houston Area Library Automation Network).

The partners will meet in person at least twice. Two meetings will take place in conjunction with other statewide library meetings (e.g. the TLA Conference in San Antonio in April and the TCAL meeting in Austin in September). Travel costs may be covered for project partners, including

Steering Committee and Technical Advisory Group members, to attend the meetings. Additional meetings (up to 2) will be held via Skype or a similar web-conferencing technology. The resulting plan will address:

1) Operational structure of the hub
2) Long-term staffing needs
3) Governance
4) Funding sources
5) Marketing and Outreach

In addition to its primary task of developing a sustainability plan for the aggregation hub, the Steering Committee will provide overall guidance for the grant project, and will assist in identifying participants for the demonstration project described below.

Objective #3: Implement a demonstration hub that aggregates collections from the Project Partners' digital collections, as well as a select group of other participants.

As an essential step in developing a full-fledged, production service for aggregating content for the DPLA, the Project Partners will undertake a demonstration project with a select group of participants and collections. This project will provide a full prototype that technology staff can iterate upon in the second year of implementation, as well as provide a framework for developing requirements for the aggregation hub and for gathering feedback from stakeholders.

Technology staff at the TDL, led by assistant director Ryan Steans, will evaluate, prototype, and implement the demonstration DPLA Service Hub that will aggregate and standardize metadata from the Portal to Texas History, Houston Public Library's digital collections, and a select group of other digital collections representing a range of collection type, technology platform, and metadata schema. The Harry Ransom Humanities Research Center at UT Austin and Baylor University Libraries have been identified as early willing participants for this demonstration project, but the Steering Committee members will work to identify other aggregation targets as well.

This demonstration hub will be developed based on requirements derived from information from other statewide service hubs, requirements of all DPLA service hubs, and the special needs of content identified by project partners and participants. The demonstration hub, which will reside in the Amazon Web Services cloud with the TDL's other hosted services, will test aggregation of content from the project participants and feed the metadata to the DPLA. It will also provide a centralized web-based discovery interface for searching and accessing items through the aggregated metadata.

A Technology Advisory Group will serve as a resource and guide for the technology team at TDL. Mark Phillips of UNT Libraries as well as a technology designate from the Houston Public Library will serve in this capacity, along with others they identify as helpful individuals during the

course of the project. A technology kick-off meeting will be held in the Fall of 2016 to discuss the demonstration project and technology requirements.

Project Impact

We believe the development of a statewide aggregation service that makes it possible for *all* digital archival collections to become available through the DPLA will have the following positive impacts:

1. It will Increase the visibility and accessibility of valuable Texas digital collections globally.
- The project will make valuable digital materials part of the large-scale DPLA collection, provide added visibility, as well as the opportunity for the materials to be used in novel ways -- for example, through DPLA's programming interface (API), as well as its ever-growing set of exhibit tools and educational programs.
- It has been demonstrated repeatedly that inclusion in the DPLA leads to more visibility and usage of collections made available there, offering contributing institutions the opportunity to reach more users, increase access to their content, and collaborate in new ways. Other statewide hubs have reported increases in visits of 50-100% after inclusion in the DPLA.
- Local libraries and cultural heritage institutions can create exhibits using their own and other materials found in the DPLA, using the DPLA's exhibit tools (http://dp.la/exhibitions) and promote these exhibits to their constituents and communities.
- The increased usage that DPLA makes possible helps libraries demonstrate their value and increases their visibility in their respective communities.

2. It will enrich a valuable resource for educators, students, and learners of all types.
- By enriching DPLA content with a more comprehensive set of Texas digital materials, the project will enhance DPLA's educational programs, including its Primary Source sets, which are designed to help students develop their critical thinking skills by exploring topics in history, literature, and culture through primary sources discoverable in the DPLA. Each Primary Source set includes an overview, ten to fifteen primary sources, links to related resources, and a teaching guide.

3. It will complete the picture of Texas and Texas collections that is currently incomplete in the Digital Public Library of America and elsewhere.
- By filling in the gaps that currently exist in Texas' contributions to DPLA, this project will provide a more complete and rich experience to users searching for Texas materials in the Digital Public Library of America.
- Furthermore, by aggregating collections statewide *and* providing a search interface for those collections, this project will provide a showcase of Texas

collections, creating a point of pride statewide in the riches of the state's collective heritage.

Key Personnel

A Steering Committee consisting of representatives of each Project Partner will provide general oversight of the project and generate the sustainability plan.

Kristi Park, Director, Texas Digital Library (UT Austin Libraries) will serve as Principal Investigator. Kristi Park leads the Texas Digital Library, a consortium of 22 academic libraries providing infrastructure at scale to support digital initiatives at its member institutions and statewide. In addition to coordination and attendance of the four scheduled Steering Committee meetings, she will provide overall coordination of the project, donating roughly 10% of her time.

Dr. Martin Halbert, Dean of Libraries, University of North Texas will serve as a member of the Project Steering Committee and attend four meetings over the course of the project year. Dr. Halbert has extensive expertise in managing collaborative digital library projects involving both academic and non-academic institutions.

Judith Hiott, Chief, Houston Automated Library Network (Houston Public Library) will serve as a member of the Project Steering Committee and attend four meetings over the course of the project year. In her position as Chief of HALAN, she has invaluable experience managing large digital collections and systems for a large public library system.

UT Libraries, through the Texas Digital Library, will provide staff time for the coordination of projects and development of the demonstration project.

Ryan Steans, Assistant Director, Texas Digital Library (UT Austin Libraries) will lead the technology team at the Texas Digital Library in the development and implementation of a demonstration aggregation hub. In his role at the Texas Digital Library, Ryan Steans manages technology infrastructure at scale for 22 member institutions of the Texas Digital Library, including an aggregated statewide repository of theses and dissertations. Approximately 10% of Mr. Steans' time will be donated to the project.

Gad Krumholz, Senior Software Developer, Texas Digital Library (UT Austin Libraries) will devote 15% of his time to software development activities related to the demonstration hub. Mr. Krumholz has broad frontend and backend development experience, including extensive work on digital repository and library workflow applications.

Kristen Lanum, Software Engineer, Texas Digital Library (UT Austin Libraries) will devote 15% of her time to software development activities related to the demonstration hub, with a particular focus on the web front end and discovery service. Ms. Lanum brings experience in user experience and interface design.

Technology Advisory Group

A Technology Advisory Group will serve as a resource and guide for the technology team at TDL. **Mark Phillips,** Associate Dean for Digital Libraries at the University of North Texas as well as a designate from the Houston Public Library will serve in this capacity, along with others they identify as helpful individuals during the course of the project. Members of the Technology Advisory Group will be invited to attend a kick-off meeting to discuss technology requirements and expectations and will be available as resources for feedback and consultation during the development of the demonstration hub.

Graduate Research Assistant (to be appointed)

A Graduate Research Assistant will be recruited and hired from the UT Austin School of Information to undertake activities related to the statewide inventory of digital collections. The position will be a part-time position (20 hours per week) with all time devoted to the project. A job description for this position is included with this proposal.

Timetable

Phase 1: Preparation and Investigation (September 1-January 31)

- September 1-15, 2016. On-boarding of Graduate Research Assistant. This will include orientation to the DPLA hub project and training in use of Qualtrics survey platform. (Note: Because of the academic calendar and university hiring practices, the GRA may have been recruited and hired prior to September 1, but will not be paid with grant funds until the granting period begins.)
- September 15-30, 2016. Develop work plan for inventory and analysis of digital collections.
- September 15-October 31, 2017. Develop survey instrument for inventory of digital collections and compile distribution lists for survey.
- Mid-September 2016. Kickoff Meeting: A joint meeting of the Project Steering Committee and Tech Advisory Group in Austin, Texas (at Perry-Castaneda Library). The goals of the meeting will include: (1) Identify potential participants for the demo aggregation project, (2) Generate consensus on high-level requirements for aggregation hub. Breakout sessions will allow Tech Advisory Group to flesh out requirements for aggregation hub and the Steering Committee to identify existing and missing elements of sustainability plan and reach consensus on a work plan for the year.
- By October 31, 2016. Make initial contact with demonstration hub participants
- October 1-December 31, 2016. Initial software development and design work. This will include the development of 2-3 rough prototypes of an aggregation service, along with development of wireframes and models for a web interface and discovery layer.
- December 1-January 31, 2016. Solicit stakeholder feedback (from the Technical Advisory Group, Steering Committee, and selected demonstration project participants).
- Late January 2017. Hold Steering Committee Meeting #2 (Skype Meeting). The goals of this meeting will be to review development work and sign off on movement to the formal development phase. Work will continue at this meeting on a sustainability plan.

Phase 2: Formal Development
- Late January-May 31, 2017. Development of aggregation service for demonstration project, including aggregator, search/discovery, and web interface.
- By February 28, 2017. Distribute digital collections inventory survey.
- March 1-April 15. Perform follow-on research to fill in gaps left in survey, through phone calls and other research.
- Mid-April 2017. Hold Steering Committee Meeting #3, an in-person meeting in conjunction with the TLA conference in San Antonio.
- April 15-May 31, 2017. Draft analysis of inventory results, including a segmentation of collections based on ease of implementation.
- April 1-May 30, 2017. Identify and coordinate with all project participants in preparation for demonstration aggregation project.

Phase 3: Implementation and Assessment
- June 1-July 15, 2017. Perform aggregations of demonstration project participants' collections.
- July 1-31, 2017. Invite demonstration project participants to test web interface and discovery/search tools.
- Mid July 2017. Hold Steering Committee Meeting #4 (Skype call). The goals of this meeting will be to finalize elements of sustainability plan and review progress on demonstration project.
- July 15-30, 2017. Collect feedback from demonstration project participants on aggregation hub through a Qualtrics survey.
- July 15-August 15, 2017. Draft and revise sustainability plan.
- Mid-August 2017, 2017. Send sustainability plan to DPLA staff for review and feedback.

Evaluation

Because we see the development of a fully functioning aggregation hub as a multi-year project, the first year is devoted to planning and development activities reaching a limited number of users. Program evaluation will be focused on assessment of whether and how well the project met its core objectives to (1) assess the landscape of digital collections statewide (2) develop a sustainable model for hub operations and (3) demonstrate a technologically viable aggregation service.

Objective #1: Inventory and analysis of collections
Successful completion of the digital collections inventory will include:
1. Contact through survey distribution with all (100%) of the academic libraries and public libraries in Texas, as well as a portion of historical societies and museums.
2. Data collection from all (100%) of the academic libraries and most (75%) public libraries through either initial survey or other research
3. Identification of important variables in digital collections statewide, i.e. successfully answer the questions "what platforms and metadata schema must an aggregation hub accommodate?"

Objective #2: Develop a sustainable hub model.

A successful outcome of objective #2 related to sustainability planning will be the the ability of the Steering Committee to fully complete all aspects of the DPLA service hub application, found on the DPLA website at: http://dp.la/info/service-hub-application/. The Steering Committee will provide elements of its sustainability plan to DPLA staff for feedback and review.

<u>Objective #3: Demonstration Hub</u>
A successful demonstration project for a DPLA aggregation hub will have the following features:
1. Successful aggregation of metadata from multiple digital collections, with multiple technology platforms and metadata schema.
2. Development of a website and discovery interface for searching the aggregated metadata.

Assessment of these outcomes will come from two two sources:
1. The Technology Advisory Group will review and provide on-going feedback on the demonstration project from a technology perspective.
2. Stakeholders in the demonstration project will be asked to complete a survey gauging the success of the aggregation and evaluating the demonstration search interface.

Sustainability

The Texas Digital Library Governing Board supports the development of an aggregation hub housed at the TDL in service to Texas libraries. TDL has been in existence since 2005 and in that time has developed a stable funding base and membership, as well as a scalable platform for digital initiatives within academic libraries. It is the TDL's mission to build capacity within the state of Texas for preserving, managing, and providing access to unique digital collections of enduring value. TDL's governing body sees the development of an aggregation hub, in collaboration with public library and other entities in the state, as a service that will fill a direct need to many of its members -- whose collections are not yet findable in the DPLA -- but also to researchers, students, and public users statewide. Such a project will strengthen the research and teaching missions of our academic library members, while also serving a much broader base of public libraries and cultural heritage centers, and their constituencies.

Like the TDL. the other Project Partners on this proposal are committed to developing a sustainable hub model through the course of the grant period that will address long-term sustainability. The sustainability plan developed by the Project Steering Committee will address these long-term needs, identifying staff resources needed and available funding to ensure sustainability of the hub for a minimum of 3 years.

Marketing of the DPLA service hub will be essential to the success of the service over the long-term. The Steering Committee will fully address this component of sustainability in its sustainability plan, to be developed over the course of the grant. During the grant period, outreach will be focused on the select group of participants in the demonstration project. However, the Project Partners will use their networks of contacts and associations to communicate and build support for the work of the service hub. Additionally, the TDL will report

on progress of the project through its formal communications channels (web site, public email lists and newsletters) as well as seek opportunities to present on the project at meetings and conferences throughout the state.