

Communicating Organizational Commitment to Long-Term Sustainability through a Trusted Digital Repository Self-Audit

Ana Krahrmer

Digital Newspaper Program, University of North Texas Libraries, Denton, Texas, USA
E-mail address: ana.krahrmer@unt.edu

Mark Edward Phillips

Digital Libraries, University of North Texas Libraries, Denton, Texas, USA
E-mail address: mark.phillips@unt.edu



Copyright © 2016 by Ana Krahrmer and Mark Phillips. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:
<http://creativecommons.org/licenses/by/4.0>

Abstract:

The digital repository activities of University of North Texas Libraries encompass The Portal to Texas History, the UNT Digital Library, and The Gateway to Oklahoma History, preserving over 1.5 million unique digital items, which comprise over 21 million pages of digital content. The news content, including both digital newspaper and news video content, represents the largest single collection in the UNT Libraries' digital collection. In November of 2015, the UNT Libraries released the findings of a year-long process to complete a self-audit and to renew its goal of overseeing a trusted digital repository according to the criteria outlined by Trusted Repositories Audit and Certification (TRAC). This work involved stakeholders throughout the libraries and served as a vehicle for discussion of organizational commitment, sustainability, and contingency planning for the digital repository activities carried out by the UNT Libraries. The resulting document, "UNT Libraries TRAC Conformance Document," and its related appendices were made publically available on the UNT Libraries website as a demonstration of an institutional commitment to digital preservation in an open and transparent way. This paper and presentation will show the structure of this self-audit, explore the documents generated and audited, and will examine challenges and their solutions as revealed by this process. Additionally this paper will present the findings of this process that specifically relate to the digital newspaper and news video content.

Keywords: preservation, sustainability, news, infrastructure planning

Introduction

University of North Texas (UNT) is a public research university in Texas with a growing student population of over 36,000 FTE, offering a wide curriculum range of subjects and disciplines. UNT Libraries supplies research resources supportive of all students, faculty, and community around the university. Over the past decade, UNT Libraries has built capacity in the area of digital libraries to accommodate a range of research activities for its body of users. In furtherance of this support, UNT Libraries has developed an array of digital library interfaces, each with a unique scope, audience, and purpose but supported by a unified digital preservation framework.

UNT developed its first interface, The Portal to Texas History, as a statewide repository of digitized and born-digital resources, intended for partner organizations around the state who sought digital preservation for and access to their Texas history materials. The audiences of The Portal to Texas History include researchers of all kinds, particularly genealogists, librarians, students, historians, and History educators. Since its initiation in 2004, The Portal to Texas History has expanded to include 325 partners, who have contributed over 725,000 digital items to that repository.

The UNT Digital Library is the second digital repository interface developed by UNT. Its scope is university-generated materials, including research and publication collections, like the UNT Scholarly Works Repository, which functions as UNT's institutional repository, and the UNT Thesis and Dissertation Collection, which houses both electronic theses and dissertations and print theses. The UNT Digital Library also creates digital access to a range of archival collections from the UNT Libraries Special Libraries Division, which entail the Special Collections, Music Library Special Collections, and the Media Library collections. At present, the UNT Digital Library maintains a little over 260,000 unique items. The primary users of the UNT Digital Library are the UNT community of faculty and students.

The third digital library interface is the Gateway to Oklahoma History. This system is operated on behalf of the Oklahoma Historical Society (OHS), and it consists of a growing collection of newspaper and news materials digitized and held by OHS, entailing nearly 1.1 million pages of newspapers and just over 405,000 photographs donated to OHS for digitization by the Oklahoma Publishing Company (OPUBCO). In total, the Gateway to Oklahoma History has over 547,550 unique resources. As with The Portal to Texas History, the primary audience of the Gateway to Oklahoma History is comprised of genealogists, historians, and educators.

These three systems combine to create the UNT Libraries Digital Collections, which serves access and preservation activities for the more than 1.5 million unique resources hosted across the three digital library systems. Because the bulk of the materials across these repositories is freely accessible, a worldwide body of researchers may utilize any of these materials, on any schedule. The "Trusted Repositories Audit and Certification" self-audit process began in 2014 at UNT Libraries with the goal of supporting "repository planning guidance, planning and development of a certified repository, periodic internal assessment of a repository, analysis of services which hold critical digital content on which institutions rely" (CRL, 2007, p. 1). The decision to initiate a TRAC self-audit was particularly significant in light of how much digital content is hosted out of the UNT digital repositories

and of how many users work depend on the content, and this decision paralleled the TRAC goal of supporting periodic internal assessment.

Major News Collections

The UNT Libraries digital collections host and preserve several major digital collections of news content. The largest of these collections is the Texas Digital Newspaper Program (TDNP). TDNP started in 2007, with a goal of preserving all newspaper content in Texas. TDNP staff collect and digitally preserve newspapers from three major content streams: paper, microfilm, and born-digital format, to digitally preserve all of this content and provide free public access to as much of the collection as possible through The Portal to Texas History. The TDNP collection in The Portal to Texas History currently has 4,126,183 pages of newspapers, consisting of 383,772 issues from 1002 titles.

The Gateway to Oklahoma History is a parallel project for the State of Oklahoma. Through this OHS is working to digitize five million pages of pre-1923 content it holds in microfilm. Currently all five million pages have been digitized and are being processed for ingest into the Gateway to Oklahoma History. The Gateway currently has 1,066,536 pages of content consisting of 144,168 issues from 533 titles around Oklahoma. In addition, the Oklahoma Publishing Company Photograph Collection represents the digitized photo morgue from the company that ran the Daily Oklahoman for many years.

Distinct from the newspaper and news photograph collections, the third major collection of new content is a video archive from the Fort Worth-based NBC 5/KXAS television station. In 2014, UNT Libraries acquired the archive of NBC 5/KXAS, which was the first television station in Texas and the Southwest when it signed on as WBAP-TV on September 27, 1948. It is an NBC-owned station in Fort Worth and serves the Dallas-Fort Worth Metroplex. This collection contains 62 years of film and video broadcasts news segments, scripts, and log books along with a wide range of archival materials. UNT Libraries Special Collections, in partnership with other units across the library, are currently digitizing large portions of this collection for access via the UNT Digital Library and The Portal to Texas History. Following a small pilot project to digitize one month of 16mm film footage from 1956, archivists have continued to digitize footage "on demand" to fill requests from researchers across the county. A large-scale digitization effort to scan original broadcast news scripts will be completed in 2016 and will provide researchers a keyword-searchable index to the film archive. Future plans for the collection include continued fundraising to enable the eventual digitization of the entire archive. Further information about the NBC 5/KXAS project is available from the UNT Libraries Special Collections page (<https://www.library.unt.edu/collections/special-collections/kxasNBC-5>).

Together these collections represent the large investment UNT Libraries has dedicated to preserving both analog and digital news content. In support of this dedication, the TRAC process affords UNT Libraries the necessary commitment to sustainability, entailing well-documented workflows and infrastructure governed by appropriate policies and rights agreements.

Other Work with TRAC

To initiate the process, UNT had to determine answers to questions directly related to TRAC and its place in academic libraries, as well as to its role in supporting digital repositories. Because the “TRAC Criteria and Checklist” (CRL, 2007) has an established legacy, early decisions in the TRAC process were informed by other institutions who have worked on the TRAC certification process. These questions were the same ones posed during the original development of the “TRAC Criteria and Checklist.” Kaczmarek and her team in 2006, when the checklist was still being arranged, discuss the library’s role in digital curation and digital collection management: “Are the librarians and archivists custodians, curators or stewards of digital resources and collections? How do they work together or independently to provide the best possible oversight of, and access to, hybrid collections of materials?” (Kaczmarek, Hswe, Eke, & Habing, 2006). As early as 2006, Kaczmarek proceeds to define roles, needs, and specific elements according to what her team hoped to achieve, which was to use the 2005 draft of the *Audit Checklist for the Certification of a Trusted Digital Repository* to evaluate an institution’s ability to be a Trusted Digital Repository. The questions Kaczmarek’s team poses were still relevant to the UNT team’s early decision-making process, especially in terms of asking what and who must be defined for a successful audit.

Repositories that have undergone the audit process in collaboration with the Center for Research Libraries (<https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>) include Chronopolis, on whose advisory panel which Mark Phillips served, and Scholars Portal in Canada. Phillips’ experience with the Chronopolis advisory panel helped UNT with answers to the initial process definitions defined by Kaczmarek, and these definitions served the TRAC team throughout the entire process. In addition, Phillips’ work on the Chronopolis advisory panel provided the team with a firm grasp on how to build a peer review process.

For their own CRL audit process, Scholars Portal developed the “Risk Analysis and Management” assessment, which included a risk assessment typology that served as a template on which UNT could base its own self-audit Risk Analysis and Management document

(<https://spotdocs.scholarsportal.info/display/OAIS/Risk+Analysis+and+Management+Strategies>). As UNT planned its self-audit team and documentation, the goal was always to develop a replicable, cost-effective process to make Trusted Repository certification more accessible as a concrete, actionable concept. Dearborn, Barton, and Harmeyer (2014) embrace the full goal of ISO 16363 (2013) certification for their digital repository, PURR, and UNT’s process in many ways paralleled what they did, though UNT’s goal differed, in part due to demands of cost in certification, and in part because UNT’s TRAC self-audit team sought to make the TRAC process more transparent with the obvious benefit relating to more robust repository management by institutions. The self-audit afforded UNT the opportunity to develop a process revealed through transparent documentation with the hopes that other institutions could replicate it in the future.

“Peer Review” TRAC Self-Audit

In the Summer of 2014, UNT Libraries partnered with George A. Smathers Library at University of Florida to prepare a TRAC peer-review process. To begin, each institution would conduct a self-audit using the Trusted Repositories Audit and Certification checklist. The project initiation provided UNT and UF the opportunity to collaboratively interpret

The project planned with both UNT and UF working together to create a dialog about the interpretation of the various items in the TRAC documentation and how they might apply to institutions like UF and UNT. The original plan called for each institution to complete a self-audit, openly publish this self-audit and then have the other institution perform a critique of the self-audit and documentation. Due to unexpected issues at UF, UNT Libraries was the only institution that completed the self-audit portion of the project. The UF team was of tremendous help during the process and lead to the successful completion of the UNT self-audit. Plans are underway at UF to complete their self-audit, though no specific dates have been established.

UNT and UF selected TRAC for this review process because it was a firmly-entrenched instrument of review in 2014, and as such multiple other institutions had completed the TRAC process and made their documentation publically-accessible creating a body of reference material. Although ISO16363 will likely be considered the “gold standard” in the future, since much of it was adopted from the earlier “TRAC Criteria and Checklist”, UNT will likely move the legacy TRAC documentation into the ISO16363 format in future review iterations.

1 UNT TRAC PROCESS

The UNT TRAC process began in Fall 2014 with the formation of a local project team. This project team consisted of people within the Digital Libraries Division who work on a range activities of the UNT Libraries Digital Collections from selection and acquisition to the final long-term preservation and access activities. The team consisted of five individuals. The project team expected to interface with others in the UNT Libraries for various sections in the TRAC documentation as needed and major interactions were required with the UNT Libraries Administrative Office that handles financial planning and human resources. Additionally individuals in the UNT Libraries Facilities and Systems Department made important contributions to the UNT TRAC process.

To initiate the project, the UNT TRAC team coordinated with the UF team to make basic decisions, such as meeting schedules between both organizations, documentation practices between organizations, and potential visit schedules.

- Information Sharing: UNT and UF decided to use BaseCamp to share information between institutions.
- Research and the Process: Both institutions’ teams scoured different TRAC audit projects, articles, software, and information about the emergence ISO16363 to initiate the process. This helped give a shape to a lot of UNT’s final documentation.
- Internal Meetings: The UNT TRAC team internally decided to try to meet once per week for approximately two hours. Since the UNT team consisted of a total of five people, everyone could not always attend each meeting.

- Collaborative Editing: the UNT team also decided to store and develop its documentation in Google Docs because it provided good collaborative editing and sharing options.

The UNT TRAC team throughout the Fall of 2014 and struggled with formulating an outline for tackling the full TRAC documentation. By February 2015, the team decided on formatting its audit into a TRAC Conformance Document that would explicitly address each of the criteria outlined in the “TRAC Criteria and Checklist”, thus providing a logical place to reference other documents as appendices within this main document. After this decision, UNT divided the three sections from the checklist between the team members.

2 POLICIES

As the TRAC self-audit progressed, the team identified multiple policies that needed to be prepared and approved by the libraries’ administration. During the process, the team arrived at three core policies that drove development of much of the other documentation associated with the overall process: a digital preservation policy, a collection development policy for the UNT Libraries Digital Collections, and a final policy document comprised of an assortment of brief policy statements relating to usage, privacy, redaction, and access to the digital collections. The TRAC team initially drafted these policies and then passed them to the UNT Libraries Policy Workgroup for comments and changes, after which the UNT Libraries’ Leadership Team and Dean of Libraries approved the final version. The TRAC team developed two additional policies supportive of the three core policies: a policy related to licensing for metadata and catalog records created by the UNT Libraries and a policy about the releasing and contributing open-source software.

In addition to serving as exciting reading, these policies provide an overarching commitment to long-term sustainability of the digital collections at UNT, and they shape how and by whom the digital collections can be used.

2.1 Core Policies

2.1.1 Digital Preservation Policy Framework

The purpose of the UNT Libraries’ Digital Preservation Policy Framework (<http://www.library.unt.edu/policies/other/unt-libraries-digital-preservation-policy-framework>) is to formalize UNT Libraries’ “continued commitment to the long-term stewardship for, preservation of, and sustainable access to its diverse and extensive range of digital assets. In alignment with the UNTL mission to create, acquire, organize, disseminate, and preserve digital content, this policy makes explicit UNTL’s long-term commitment to the University of North Texas (UNT) community as its trusted digital repository” (Digital Preservation Policy Framework, 2015). Perhaps the most complex of the three core policies, the digital preservation policy identifies all of the elements of digital preservation that UNT Libraries will examine, implement, and/or maintain over time to achieve a specific result with the goal of ensuring institutional adherence to a Trusted Digital Repository Infrastructure (see Table 1).

Table 1: Digital Preservation Policy with individual sections explained.

| Elements | Action | Intended Result |
|----------------------------|---|---|
| Mandate | Identifies five commitments and the stakeholders involved with these. | Supports long-term goals supportive of all parties affected by the digital collections. |
| Challenges | Identifies eight barriers to digital preservation, though this list is not exhaustive. | Makes concrete those obstacles that will need possible solutions through proper planning. |
| Principles | Identifies the libraries' position in maintaining operating principles and preservation standards. | Positions libraries as an agent in developing and maintaining organizational principles and digital preservation standards. |
| Categories of Commitment | Identifies types of materials for which libraries will ensure digital preservation and how the libraries will achieve this. | Provides a scope for the types of materials that will be digitally preserved and made accessible. |
| Levels of Preservation | Identifies the NDSA levels of preservation to which the libraries adhere. | Supports long-term adherence to an established digital preservation conceptual model. |
| Roles and Responsibilities | Identifies all parties involved in supporting the digital preservation policy framework. | Represents who, in what capacity, and how they should be involved. |
| Collaboration/Cooperation | Identifies the significant role partnership makes in digital preservation. | Supports specific actions UNT will take in furthering preservation activities through collaboration. |
| Access and Use Criteria | Identifies those resources the libraries seek to preserve and build access to and specifies another core policy, "UNT Libraries' Digital Collections Usage and Feedback Policy." | Concretizes the intention of the libraries to provide access to as large a body of content as is possible, but also refers to special circumstances as identified in the Usage and Feedback Policy. |
| Implementation | Identifies what elements affect proper implementation of the policy: infrastructure, feasibility of solutions, standards adoption, and evolution of standards and best practices. | Offers possible actions the libraries can take to navigate the long-term needs associated with digital preservation. |

One more element that appears in the Digital Preservation Policy Framework is the review cycle. In support of the TRAC Criteria & Checklist, a review cycle accompanies each policy developed for TRAC.

2.1.2 Collection Development Policy for the UNT Libraries' Digital Collections

As a library group with materials, patrons, and staff, the UNT Libraries' Digital Collections adheres to a Collection Development Policy, prepared in conjunction with the TRAC Criteria & Checklist. This policy parallels other collection development policies, and it "describes the policies governing the acquisition, curation, and management of materials in the UNT Libraries' (UNT) Digital Collections" (<http://www.library.unt.edu/policies/collection-development/collection-development-policy-unt-libraries-digital-collections>). Two elements in this policy are of unique significance to UNT Libraries, while the remaining elements are variations on other library collection development policies.

The first element of significance deals with consumers of the UNT Libraries' Digital Collections. The policy identifies these as:

- The UNT Extended Community (comprising students, faculty, staff, alumni, and administrators)
- Contributing partners and their constituents
- The larger academic community and researchers
- The general public throughout the world

Because so much of the content within the UNT Libraries' Digital Collections has been added through collaboration, the communities involved in using and maintaining interest in the collections are deliberately defined broadly in this policy.

The second element of note is the collection scope. While the concept of scope as it is used within the policy aligns with how it is used across library collections, the scope of the UNT Libraries' Digital Collections is laid out in this policy based on the three digital library interfaces identified above, with further elaboration provided within the policy (see Table 2).

Table 2: The Digital Library Interfaces at UNT and collections of note in each.

| Digital Library Interface | Scope | Collections of Note |
|-----------------------------|---|--|
| The Portal to Texas History | Materials collected and curated within Texas | <ul style="list-style-type: none"> • Texas Digital Newspaper Program • The Texas Register • Texas Laws and Resolutions Archive • Texas State Publications |
| UNT Digital Library | Materials held by groups at the university, also serves as a digital repository for campus research materials | <ul style="list-style-type: none"> • UNT Scholarly Works • UNT Data Repository • UNT Theses and Dissertations • CyberCemetery • Technical Reports and Image Library |
| Gateway to Oklahoma History | Portal to resources housed at the Oklahoma Historical Society | <ul style="list-style-type: none"> • Oklahoma Digital Newspaper Program • Oklahoma Publishing Company Photograph Collection |

2.1.3 UNT Libraries’ Digital Collections Usage and Feedback Policy

The Usage and Feedback Policy addresses an assortment of issues relating to external users of the digital collections. This policy includes:

- **Four categories of access**
 - Public Access: defined as open access via the Web.
 - UNT Community: access is available to anyone affiliated with the UNT Community.
 - UNT Community Strict: access provided only to current UNT Community members who are required to authenticate using their EUID before each use, no matter their physical location.
 - Physical Premises: access does not require authentication but resources are only available in the physical buildings of the UNT Libraries, on computers connected to the Local Area Network (note WiFi cannot connect to these resources).
- **Embargoes:** resources are digitally preserved in a “pre-release state” but the resources are not available to any users.
- **Rights/Terms of Use:** provides copyright protection and terms-of-use statement for users.
- **User Privacy:** states what type of user information the access systems collect for site usage and improvement purposes, and identifies privacy protection measures.
- **Redaction and Removal Requests:** Perhaps the most controversial requests any collection--digital or otherwise--receives are redaction and removal requests. This statement represents what types of removal requests UNT Libraries will and will not uphold.

- **Intellectual Property and Copyright Challenges:** outlines the process UNT Libraries will take under intellectual property/copyright challenges.
- **User Feedback:** dissects the process UNT Libraries takes in responding to user feedback and explains what types of services UNT Libraries does not offer.

The Usage and Feedback Policy document is, in actuality, comprised of multiple smaller policies that all fit under the umbrella of public interface with the UNT Libraries' Digital Collections. As such, it is intended to address an array of patron needs that have either already arisen over the past few years or that UNT Libraries' TRAC team or the Libraries' administration predict to occur in the future. This document directly ties into requirements on the "TRAC Criteria and Checklist," sections A3, A5, and B6

([http://www.library.unt.edu/sites/default/files/documents/digital-libraries-uploads/UNT Libraries TRAC Conformance Document.pdf](http://www.library.unt.edu/sites/default/files/documents/digital-libraries-uploads/UNT_Libraries_TRAC_Conformance_Document.pdf))

These three policies serve as the hub around which the TRAC team developed multiple other documents required by the "TRAC Criteria and Checklist," including

2.2 Additional Documentation Addressed in Audit

2.2.1 Memoranda of Understanding, Partnership Agreements, Letters of Commitment

Since the creation of The Portal to Texas History in 2004, UNT Libraries has worked with potential partners to proactively secure necessary rights related to the digital content added to the digital collections. This work has produced a document called the Partnership Agreement, which is now used for both Portal and UNT Digital Library projects. The Partnership Agreement has been approved by the UNT System Office of General Council and requires signatures by both contributing partners and UNT before projects are placed into the digital collections. The agreement defines the rights and expectations of both UNT and the partner institution. This includes the full rights for the partners to reuse all content digitized in any way and also identifies descriptive metadata created for a digital resource as being in the public domain. This standard partnership agreement also outlines what would happen in the unexpected event that the UNT Libraries could no longer provide the same level of support or service through its digital repositories. Both digital and physical copies of all partnership agreements are retained by the UNT Libraries.

Another document, the "Memorandum of Agreement for Digital Rights," serves as the signature document for rights holders to allow their content to be made digitally available via the Web. The most common groups to sign this document are newspaper publishers, but UNT Libraries strives to obtain permission from extant rights holders as necessary prior to making content digitally available. As with the partnership agreement, the Memorandum of Agreement for Digital Rights is employed for both Portal and UNT Digital Library projects.

In addition to these signed partnership and rights agreements, University of North Texas has entered into partnerships with other state and federal agencies to preserve important content in the UNT Libraries Digital Collections. Agreements exist between UNT and the Texas Secretary of State, the Texas State Library and Archives Commission, the Government Publishing Office, the National Archives and Records Administration, and the Oklahoma Historical Society. These agreements identify various responsibilities of UNT in maintaining collections in partnership with these organizations, as well as what will happen in the event that UNT is no longer able to maintain its commitment.

These partnership agreements and memoranda of understanding form a set of contingency plans for future access to content in the event that the UNT Libraries Digital Collections are not able to provide continued access and preservation of resources. The TRAC team recognized that in many cases there would be a severe change in the way that resources might be accessed, but these contingencies support long-term sustainability required by the TRAC Criteria & Checklist.

During this process the TRAC team identified two areas where additional memoranda of understanding are needed: the Texas Digital Newspaper Program content contributed to the Portal to Texas History by the UNT Library and collections digitized or housed from the UNT Libraries Special Collections. At this point, UNT Libraries is developing an agreement in support of a contingency plan for these digital collections in the event that UNT is no longer able to maintain them. This planned agreement would transfer preservation copies and associated metadata to another large institution in Texas for access and preservation.

2.2.2 Workflow Documents

In support of the core policies, the TRAC team developed a set of workflow documents that would help to execute the policies as necessary. These supplemental documents directly ties into both the three core policies and the TRAC Criteria & Checklist (see Table 3).

Table 3: Supplemental documents, explanations, and respective policy section each applies to.

| Document Title | Description | Overarching Policy & Section |
|---|---|---|
| “Formal Statement of Conformance to ISO 14721:2012” | Statement of UNT Libraries’ Digital Collections conformance to the OAIS Reference Model | Digital Preservation Policy Framework <ul style="list-style-type: none"> • Levels of Preservation, Implementation Collection Development Policy <ul style="list-style-type: none"> • Guiding Principles |
| “UNT Libraries’ OAIS Information Package Specification” | Documents the Submission, Archival, and Dissemination Information Packages used within the UNT Libraries Digital Collections. | Digital Preservation Policy Framework <ul style="list-style-type: none"> • Levels of Preservation Collection Development Policy <ul style="list-style-type: none"> • Guiding Principles |
| “UNT Libraries SIP-to-AIP Conversion Workflow” | Documents UNT Libraries’ process for converting Submission Information Packages (SIPs) to Archival Information Packages (AIPs). | Digital Preservation Policy Framework <ul style="list-style-type: none"> • Levels of Preservation • Implementation |

| | | |
|--|---|--|
| <p>“UNT Libraries AIP-to-DIP/ACP Conversion Workflow”</p> | <p>Documents UNT Libraries’ process for conversion from AIPs to Dissemination Information Packages (DIPs)/Access Content Packages (ACPs).</p> | <p>Digital Preservation Policy Framework</p> <ul style="list-style-type: none"> • Levels of Preservation • Implementation |
| <p>“UNT Digital Libraries: Risk Analysis and Management Strategy Plan”</p> | <p>Outlines some known risks to UNT Libraries’ Digital Collections in relation to the technical infrastructure, environmental concerns, and risks related to staffing and personnel.</p> | <p>Digital Preservation Policy Framework</p> <ul style="list-style-type: none"> • Levels of Preservation • Roles and Responsibilities <p>Collection Development Policy for the UNT Libraries’ Digital Collections</p> <ul style="list-style-type: none"> • Audience • Retention and Evaluation |
| <p>“Data Loss Escalation Procedures”</p> | <p>Explains the escalation procedures related to potential data loss; created to clarify what steps should be followed and what sort of information should be relayed during a potential or actual data loss event.</p> | <p>Digital Preservation Policy Framework</p> <ul style="list-style-type: none"> • Levels of Preservation • Roles and Responsibilities <p>Collection Development Policy for the UNT Libraries’ Digital Collections</p> <ul style="list-style-type: none"> • Audience • Retention and Evaluation |

In addition to these supplemental documents that tie directly to preservation and sustainability, the UNT TRAC team also prepared formal documents about locally-developed digitization standards and metadata element set based on documentation that the UNT Libraries’ Digital Libraries have maintained over the past decade. This legacy documentation has provided invaluable information, allowing the TRAC team to canonize local practices into the two formal documents, one titled, “Preferred File Formats” and the other, “Digital Projects: Minimally-Viable Records.”

Supportive of this process, the team developed two additional pieces of technical documentation for the TRAC self-audit. The first document is a formal XML Schema for the UNTL metadata format. In past practice, UNTL XML records were generated consistently by an external metadata library called pyuntl, but no formal XML Schema existed that defined this format. The UNT TRAC team authored this schema document and verified all of the UNTL records in production against this schema. The UNT Libraries utilizes the Metadata Encoding and Transmission Standard (METS) to serialize digital objects and their associated technical and structural metadata. The second document was this METS format for AIPs, canonized as a METS Profile with the Library of Congress during this self-audit project.

3 PROCESS RETROSPECTIVE

The UNT TRAC team released the finalized set of documentation, centered around the “UNT Libraries: TRAC Conformance Document” in November 2015, after roughly one year of work. Over the course of that year, the TRAC team developed, deleted, and re-developed workflow and process documentation and policy document drafts. In short, the “TRAC Conformance Document” did not emerge, fully-formed, like Athena from the head of Zeus. The documentation evolved through a drafting process, with some items thrown away as unnecessary or not relevant, while other items, particularly the core policies, developed as the most motivating documents for the large conformance document. This evolutionary process informed the final output of this project, and it will inform the future review cycles for all of the documentation. The process also pointed in possible directions for other institutions who seek to undertake this endeavor.

3.1 What are we trying to certify?

One of the challenges that the UNT TRAC Team faced at the onset was to determine exactly what to certify as a trusted digital repository. This challenge directly relates to the situation of UNT Libraries within a larger university system, whereby the digital repository is managed by an entity beneath the umbrella of multiple other departments and university administrative units. Throughout the project, the team dealt with the three access repositories and the Coda preservation repository that provides preservation service for all of the access systems. Above this infrastructure is the Digital Projects Unit, a unit within the Digital Libraries Division, situated within the UNT Libraries as a whole, which is located on the UNT Denton Campus, and which is organized within the UNT System. Some of the requirements that are in place for financial auditing exist because University of North Texas and UNT Libraries operate as a state entity of the State of Texas. In the end, the TRAC Team defined the UNT Libraries Digital Collections as the unit being certified with support from a wide range of actors across the organization. Once this was decided, it became easier to move forward with the overall process.

3.2 Policies take time

At UNT Libraries, the process for approving official library policies requires first submitting drafts to the libraries’ Policy Workgroup for review from stakeholders across the library. After that a new policy is presented to the Library Dean’s Council (LDC) for additional review and final approval, after which the approved library policies are recognized as official. The TRAC team used the three core policies as a means of formalizing extant practices and conceptual frameworks around digital libraries access and preservation at UNT, and these core policies required the greatest amount of effort on the part of the team because they simply covered so much ground. One minor challenge in policy development arose as the team tried to decide when a given policy was complete enough to move through the process of review and formal acceptance by the libraries’ leadership council. A few revisions to the policies were necessary after they were approved and before the final TRAC Conformance Document was published. The full policy development cycle took approximately six months, from initial drafting to final approval and publishing.

3.3 Examples Help to Move Forward

The UNT TRAC Team was fortunate to be able to pull from existing documentation made publically available by other institutions who have undergone the TRAC process, such as the

Scholars Portal in Canada and the HathiTrust in the United States. Both of these institutions have completed an external certification by CRL for their TRAC work and have made available a wide range of documentation online. The UNT TRAC team requested permission to use the Scholars Portal's Risk Analysis and Management Strategy Plan as a template for the risk analysis documentation. The "TRAC Conformance Document" concept was based on the section-by-section documentation that the HathiTrust created for their certification process.

Some sections of the TRAC documentation required the work of people not on the TRAC team, particularly the TRAC sections dealing with budgetary initiatives and technological infrastructure. The UNT Libraries Administrative Offices helped the TRAC team extensively on the TRAC Criteria & Checklist sections related to financial planning, auditing, and personnel. The team coordinated with the Libraries' Facilities and Systems division in answering questions related to underlying physical and network infrastructure, and network and physical security practices. In the case of both groups, example language from previous TRAC audits from other institutions helped clarify the scope of statements the team needed assistance with on the TRAC self-audit.

3.4 Many Eyes

During Summer 2015, after the first draft of the "TRAC Conformance Document" and its associated appendices were completed, the UNT TRAC Team contracted three external readers to perform a thorough analysis of the document. This included two readers external to the university and one UNT librarian who had a fresh view of the Digital Libraries operation at UNT and could read the document with objective, fresh eyes. All three external reviewers had deep knowledge in the field of digital libraries and a strong understanding of the TRAC process. These external readers were invaluable during this process, particularly in terms of providing new insight into what the TRAC Criteria & Checklist means in certain difficult-to-understand sections, by pointing out when the TRAC team had used UNT (local) jargon to describe something that needed to be better illustrated for external audiences, and by giving three completely different perspectives about how fully or well the TRAC team had interpreted the Criteria & Checklist sections. A number of situations arose in which the UNT TRAC team had interpreted a section of TRAC in a way that was different from the external readers, and the readers' advice proved helpful in coming to a consensus. The three external readers were given approximately one month to complete their assessments, though they finished in slightly less than that time, and they were compensated for their work. Each reader was given a separate copy of the "UNT Libraries' TRAC Conformance Document" in Google Docs to provide comments and critique.

In addition to the external reader analysis, a public draft was distributed to all of UNT Libraries for a three-week comment period. All comments aggregated into one document, combining those of the three external readers with the public draft comments, after which the TRAC team reviewed and incorporated revisions into the final UNT TRAC Conformance Document. The overall process took approximately three months, from initially sending of the documents to the external readers and to the UNT Libraries employees until the team completed the revisions.

3.5 Preserving News Content

Having undergone a full TRAC self-audit, we have identified a number of issues that relate directly to the preservation of news within the context of the UNT Libraries.

First and foremost of these issues is the cost associated with managing a collection the size and scale of the news content within the repository. The newspaper content in the Texas and Oklahoma Newspaper projects alone comprises over half a million items in the UNT Libraries Digital Collections, representing 33% of the total 1,568,316 items currently in the repository. Not only does this require a significant amount of storage to maintain, other challenges exist in handling this size of content. Maintaining high-quality metadata for a collection of over half a million records is a challenge in most institutions, and while UNT Libraries has worked to minimize some of the issues other institutions address as problems that arise in the creation and processing of these large collections, errors can still occur and when they occur, a large number of individual records have to be altered. UNT Libraries, fortunately, incorporates some automated methods to process record corrections. Quality in the metadata record set stems from a built-in lack of variation in records to describe multiple issues within a title--basically batch records for title runs, resulting in only a small amount of field deviation between records: main title, date, citation and physical description fields. Ideally, the metadata creators for the newspapers also add subjects and keywords to the newspaper records to identify unique characteristics represented by each title. Adding these subjects and keywords is often more useful when records are aggregated where full-text searching is not a possibility, such as the metadata aggregation of records by DPLA.

Another issue the TRAC self-audit recognized in relation to the preservation of news is the sheer bulk of research needed to identify, transfer, describe, and preserve the moving-image content contained in the NBC 5/KXAS collection. With an estimated storage capacity of over two petabytes required to store the transfers of the film-based items and another two petabytes estimated for the video this collection stands to be the largest in the UNT Libraries Digital Collections. While this estimated 4+ petabytes of data reaches some years into the future, UNT Libraries has mechanism in place to plan for access to this material. For each of the news clips on film or video, there is also an accompanying script that news anchors read during the broadcast, and these scripts are also being digitized, with optical character recognition (OCR) and metadata records. The scanning and OCR process is expected to be completed by December 2016, with metadata description to be completed in the following year. The resultant project will provide keyword access to over 30,000 scripts, entailing nearly 300,000 pages of content. This collection of scripts will provide a keyword-searchable retrieval mechanism to aid in the identification of film and video news clips that should be prioritized for digitization. In addition to the script digitization, in 2016 and 2017 the UNT Libraries Media Library will be working with the Special Collections Department to digitize the U-MATIC tape collection from NBC 5/KXAS. This collection is expected to result in the ingest of approximately 2 TB of video content per-month over a twenty-four month period.

A third issue related to news preservation arises relating to privacy, particularly in the case of more recent news content like the PDFs in TDNP and the NBC 5/KXAS collection. While genealogists love to locate family members from historic issues of newspapers in the digital repository, a recently-arrested defendant is often much less pleased to locate his/her name in newspapers or in the news transcripts and will write to request information removal. Other situations not related to crime also arise that cause users to find their names and with remove it, though the most common issues derive from the legal notices in the more recent, born-digital PDF newspapers. The TRAC process required documentation about how and when information removal occurs, as and after completing the audit process, UNT Libraries addressed handling removal requests in their "Digital Collections' Usage and Feedback Policy," which the libraries staff have been able to reference multiple times when patrons have requested information removal. This policy, discussed in further detail above, provides a

mechanism for uniform enforcement of the removal policy and ensures that requests for removal will be handled correctly and consistently across library departments.

4 RECOMMENDATIONS & CONCLUSION

As the TRAC audit process has evolved, it has moved from a strict focus on software and technology, as especially indicated by the literature from 2005 onwards, to an expanded focus on the repository within the context of its respective institution. A piece of equipment or software can serve countless users, but when its host institution has no infrastructure for sustaining it, trustworthiness is gone. Implementing the “TRAC Criteria and Checklist” as a self-audit has afforded UNT the opportunity to develop internal documentation that represents the libraries’ commitment to preservation and access. As the initial audit and documentation phase concludes, the TRAC team has plans for future directions and recommendations for groups interested in pursuing something a similar self-audit.

1. Carefully read for understanding the “TRAC Criteria and Checklist” (2007) or ISO16363 before starting: The TRAC review process is daunting, whether it is by external reviewers or a self-audit. Understanding how the process will place under a microscope both your repository infrastructure and your institution before you begin will prepare you for the amount of work needed to complete the process.
2. The shoe doesn’t always fit perfectly: The UNT TRAC team and its external reviewers discovered over time that some of the elements in the “TRAC Criteria and Checklist” did not necessarily apply to UNT’s situation because UNT is a state university. As a result, the team strove to interpret the criteria according to what we believed was its original intent, with the goal of maintaining both fidelity to and accuracy in the process.
3. Be prepared to find weaknesses: It’s better to find and addresses weaknesses in your repository during a self-audit before a disaster or before an external entity finds your weaknesses. The whole purpose of a self-audit is to find and repair holes for long-term stability.
4. Don’t be afraid to coordinate: Multiple sections of the “UNT Libraries’ TRAC Conformance Document” required coordination between the TRAC team and other entities in the library beyond the Digital Libraries. UNT Libraries is a part of an organic whole, and as such, some pieces of work for this project fell into other departments or areas. The Digital Libraries at UNT has a long track record of collaboration with other groups, and this made gathering information a fairly simple process.
5. Processes take time: The UNT TRAC team began with a long-range plan that accounted for approximately eighteen months, and this plan included the cooperative piece of working with University of Florida on the peer review stage. This timeline was sufficient for the team to complete its own work, but a little more time would have been necessary for the cross-institutional, peer review piece. As a result, we recommend that other institutions considering this process should develop both short- and long-range plans to keep progress moving at a steady pace.

Acknowledgments

We wish to thank our colleagues at the George A. Smathers Library, particularly Laurie Taylor and her team. We also thank Morgan Gieringer, Head of Special Collections at UNT Libraries, for her contributions about the NBC 5/KXAS collection. Finally, we would like to thank the remaining members of the UNT TRAC team, Hannah Tarver, Daniel Alemneh, and Laura Waugh, for their dedication to this project.

References

- Center for Research Libraries. (2007). Trustworthy repositories audit & certification: Criteria and checklist. Retrieved from https://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf.
- Dearborn, C. C., Barton, A. J., & Harmeyer, N. A. (2014). The Purdue University research repository. *OCLC Systems and Services*, 30(1), 15-17.
doi:http://dx.doi.org/10.1108/OCLC-07-2013-0022
- ISO (2013). ISO 16363 Trusted digital repository (TDR) – Audit and certification of trusted digital repositories.
- Kaczmarek, J., Hswe, P., Eke, J., & Habing, TG. (2006). Using the audit checklist for the certification of a trusted digital repository as a framework for evaluating repository software applications: A progress report. *D-lib Magazine*, 12(12), 1082-9873.