

TRENDS IN
ARCHIVES
PRACTICE

MODULE 8

**BECOMING A
TRUSTED DIGITAL
REPOSITORY**

STEVE MARKS

with an Introduction by Bruce Ambacher

Edited by Michael Shallcross



**SOCIETY OF
American
Archivists**

CHICAGO

Appendix A

Case Study: University Archives, University of Illinois at Urbana-Champaign

by Bethany Anderson

The following case study describes an informal TDR self-assessment utilizing Steve Marks's *Becoming a Trusted Digital Repository* as a guide. It is written from the perspective of an academic archives within a large institutional research library.

Established in 1963, the University Archives at the University of Illinois at Urbana-Champaign preserves the administrative records, faculty/alumni papers, and institutional records of several associations and professional organizations. The University Archives (UA) is one of several special collections units located within the University of Illinois Library. Like other special collections units, UA works in collaboration with the Library's digital preservation (DP) service unit to ensure the long-term preservation of and access to its born-digital and digitized materials. Although UA does not directly manage the Library's digital preservation repository (Medusa), members of its staff serve in advisory roles, helping to shape its preservation policies.

Like many archives, UA is good at creating access to its holdings, but when it comes to digital preservation planning, our repository's staff members have realized that we must leverage our skills, creativity, and knowledge to share the responsibility of stewarding digital materials and ensuring their long-term preservation and accessibility. Indeed, creating access is not enough to gain the trust of users; archives must also demonstrate that they are actively preserving the resources of enduring value with which they are entrusted. Conversely, keeping digital resources in a dark archives risks reducing their cultural and intellectual capital and likewise jeopardizes the trust of funders, donors, and stakeholders. The best way to preserve archives is to plan for access in perpetuity. As this module demonstrates, if archives are to be trustworthy, they must safeguard the provenance, authenticity, and integrity of materials while also facilitating access. For UA, this

assessment reinforced the importance of formulating preservation plans informed by access needs, underscoring Brewster Kahle's contention that in digital environments, "Access drives preservation."⁷¹

To determine how well the Library is meeting preservation requirements, we conducted an informal TDR assessment. It proved to be a vital exercise that will help UA refine and attune its digital records appraisal, arrangement, and description practices and workflows to better facilitate preservation. Although we were relatively unfamiliar with the criteria of ISO 16363 and had previously had only limited interaction with digital preservation systems (other than as a depositor or creator of digital resources), Steve Marks's module guided us through the process with ease.

In the main body of this module, Marks notes the importance of organizing one's responses during a self-assessment or formal audit. To more easily track and document the many facets of our multi-unit preservation operations for digital content, we created a wiki for the audit checklist.⁷² Answering the criteria required evaluating both current workflows and documentation, sometimes involving members of the DP unit, who clarified digital preservation policies and procedures. In most cases we found that, even if the University of Illinois Library had not already formulated and articulated many of its critical policies and procedures related to the preservation of digital content, its committees and working groups are currently developing a more integrated and effective preservation environment that reflects compliance.

Ascertaining the Library's compliance with some of the criteria proved to be quite straightforward, and we were able to locate specific policies and governance documentation. While the DP unit has articulated its mission statement and Collection Policy and defined the user groups that compose its Designated Community (Sections 3.1.1, 3.1.3, and 3.3.1), many policies related to its organizational infrastructure and procedural accountability are currently in draft form, including its Preservation Strategic Plan (Section 3.1.2). It has yet to institute

71 Brewster Kahle, "Mass Digitization, Library Lending, Internet Archives," presentation to UNESCO Memory of the World Meeting, Vancouver, September 2012, http://www.unesco.org/archives/multimedia/index.php?s=films_details&pg=33&id=2771#.VG0gcYflo3I.

72 See <https://wiki.cites.illinois.edu/wiki/display/libraryarch/University+Archives+Trusted+Digital+Repository+Self-Assessment>.

a policy and process for self-assessment, including a framework for measuring the integrity of information objects, as well as a review cycle for its digital preservation documentation as best practices and methods in the field evolve and change (Sections 3.3.2.1, 3.3.5, and 3.3.6). According to Marks, demonstrating accountability through self-assessment is as important as making decisions to determine which content information and information properties will be preserved. The Library will need to formalize mechanisms to promote regular self-assessments and greater transparency, especially by documenting its policies and procedures to record institutional knowledge, decisions, and preservation actions, as well as to create open access to internal and operational documentation (Section 3.3.4). It will likewise need to commit to a review cycle for this documentation as part of the constant process of evaluation, improvement, and reevaluation (Section 3.3.6).

The criteria revealed that we have yet to formalize and delineate many steps in both DP's and UA's workflows to manage digital objects and enhance their accessibility, as well as to promote greater engagement with our Designated Community. UA and DP have not formally defined or documented which significant properties of digital content to preserve, nor which properties would be most meaningful to our Designated Community (Section 4.1.1). Archivists have long been concerned with preserving the authenticity of the digital object, but what if that original object cannot be opened by modern software suites? What if the file format cannot be identified or the file is corrupt? How would we create access, and how would we know what kind of access users will require? What does this mean for the preservation of the original object? While these are not easy questions to answer, the assessment reinforced the importance of creating and maintaining dialogues with our Designated Community. Such dialogues will help us begin identifying which aspects of digital content we should actively preserve to better serve this community's needs and gain its trust.

The assessment proved to be an essential element in identifying preservation gaps in the Library's organizational infrastructure and digital object management, and it also helped us evaluate our technical infrastructure. Since Medusa's inception, the DP unit has actively managed its storage locally, replicating bitstreams across two storage

clusters/data centers as it monitored its hardware and object storage architecture, disk space, motherboard temperatures, CPU/Memory utilization, and server performance. However, the Library recently partnered with the National Center for Supercomputing Applications (NCSA) at the University of Illinois to provide storage for the Medusa digital preservation service. Going forward, the NCSA will monitor Medusa's infrastructure and core components and maintain two copies of digital content at two different locations on campus, while a third copy of files are backed up on the Amazon Glacier Web service (Sections 5.1.1.2 and 5.1.2). Because NCSA has the capacity and infrastructure to support a petabyte storage environment, the digital preservation service will be able to scale as its storage/backup model evolves in response to emerging technologies and its digital resources increase.

Conducting an informal self-assessment not only helped us to better understand our own preservation operations, it also provided significant insight into the frameworks, policies, and practices that UA and DP will need to put in place if we wish to be successful and trustworthy. It enabled us to evaluate UA's own methods for acquisition, arrangement, and description, putting particular emphasis on the relationship with our Designated Community—the group of scholars, staff members, and students who constitute our main user group. Ultimately, the assessment showed that the services provided by DP, such as the Medusa repository environment, are necessary but not fully sufficient means by which we can demonstrate trustworthiness. The University Archives must establish and have in place deposit agreements that grant it legal custody and responsibility for preserving digital materials, it must delineate and formalize guidelines and processes for structuring and preparing AIPs for ingest, and it must select and implement preservation and access strategies such as file format migration rules, regardless of the specific preservation means provided in the basic repository hardware and software. Last, we realized that, while the OAIS is an important model to adopt and integrate into our own framework for managing digital materials, we must remain flexible and cognizant of our unique context, developing curation strategies that suit our own and our Designated Community's needs.

Like many archives and special collections, UA seeks to maintain the authenticity of digital objects as it seeks to promote access and preservation. To make this a success, UA must coordinate and collaborate with others in the broader Library, developing a community of practice around digital preservation. As repositories negotiate which aspects of their digital objects will be preserved for posterity, they must not only be aware of the community for whom they take these measures while documenting these decisions but also of the organizational change required to engage with digital resources.