

The Impact of the Shift to Cloud Computing on Digital Recordkeeping Practices at the U-M Bentley Historical Library

Max Eckard

Bentley Historical Library
University of Michigan

Dallas Pillen

Walter P. Reuther Library
Wayne State University

ICHORA 2020 | Wednesday, October 28, 2020 | Virtual



IT Guiding Principles

The following principles guide the pursuit of the strategy:

- Technology will support and encourage learning, research, and interdisciplinary collaboration.
- Systems will make their data available for other processes whenever possible, while respecting limits required for personal privacy, regulatory compliance, and IT security.
- Technology choices will favor solutions offered as external cloud services.
- Services will consist of all elements required for the adoption and best use of the service, including training, support, appropriate funding, and accessibility for all.
- IT providers will work together to limit redundant or outdated services so that investments can be redirected toward new technology service needs.

The Shift to Cloud Computing

Adoption of Cloud Services

- Cloud services facilitate the use of “**shared documents, files, software, knowledge, and applications**” through the internet and remote servers
 - Google’s G Suite, Microsoft 365, Box, Dropbox
- Significant adoption throughout industry, government, and academia
- Adoption spurred by **increased collaboration** afforded by cloud-based tools and **assumed cost-savings** associated with moving from local administration to externally-supported services
- Remote infrastructure administered by third-party entities presents **unique recordkeeping challenges**

Recordkeeping Challenges

- Challenges include...
 - Maintaining links between **records and metadata**
 - **Transferring records** out of remote/closed systems
 - Loss of access to records
 - Record **destruction or loss**
 - Ensuring **authenticity, integrity, and evidential value** of records is maintained
- No local ownership and limited control of records in externally-hosted services
 - Remote infrastructure not defined or made known to users
 - Relevant metadata may not be maintained or transferred

The Cloud at U-M

- 2010 IT assessment → IT Rationalization project
- Effort to replace “more than 40 e-mail and calendar services” and to address budgetary shortfalls
- NextGen Michigan Collaboration Project
 - 2011 Cloud Strategy group recommended “externally provided services”
- 2011 selection of Google’s G Suite for Education collaborative tools
- 2012 implementation of M+Box for storing and sharing files
- 2016 IT Strategic Plan: “Technology choices will favor solutions offered as external cloud services”
- Late 2010s: cloud-based tools in widespread use throughout all levels of the university

Archivist Involvement and Response

- Archivists at U-M served on various NextGen project committees
- Drafted feedback documents to “take a stronger role in the definition, identification and preservation of records”
- Many implementation details made without archivist involvement
- Cloud tools proliferated rapidly and have had a lasting impact on digital recordkeeping practices at the U-M Bentley Historical Library



About the Bentley Historical Library

About the University Archives

- Mission: Among other things, to “collect the materials for... the University of Michigan



University Archives and Records* 601.08

I. Policy

For the purposes of this policy, “university records” are defined as all records regardless of their form, prepared, owned, used, in the possession of, or retained by administrators, faculty acting in administrative capacities, and staff of university units in the performance of an official function.

A History of Decentralization

- U-M is decentralized
- By state constitution independent in its record keeping practices
- **1935:** Even from its establishment, the Bentley recognized that not all was going to be centralized
- **1936:** Committee on University Archives
 - Already recognized that it was “manifestly impossible and undesirable to assemble all such records in one place”
 - “As one of its first acts, a useable catalogue of existing archives“ and their location, should be compiled”
 - A revealing statement of a precustodial mentality

University Records, *Regardless of Their Form*

- **1979:** Hosted “Conference on Archival Management of Machine-Readable Records” conference
- Foundational period...
 - **1997:** Received and developed a preservation strategy for its first significant collection of born-digital archives
 - **2000:** Initiated a web archiving program
- Gaining more experience and understanding...
 - **2006:** In partnership with U-M Libraries, began providing online access to born-digital archives via DeepBlue, U-M’s digital preservation and access repository, based on DSpace
 - **2007:** Hosted “Development of Case Studies for the Effective Management of University Digital Records,” which led to the development of the “Campus Case Studies” portal hosted by the Society of American Archivists

Transforming in Response to Digital Environment

- **2009-2010:** “E-Mail Archiving at the University of Michigan” (or “MeMail”) project yielded a significant improvement, a dedicated “Technical Lead” position
- **2011:**
 - Digital Curation Division established with the goal of developing and implementing solutions for the long-term preservation and management of digital materials
 - Later it became the norm for processing archivists to process collections holistically, regardless of physical, analog, or digital format
- **2014-2016:** ArchivesSpace-Archivematica-DSpace Workflow Integration project united three Open Source Software platforms to allow for the more efficient creation and reuse of metadata and to streamline the ingest of digital archives

Adapting Digital Recordkeeping Practices to This Shift

Bentley Involvement in U-M RM and the Cloud

- **2010:** Bentley archivists, including the Director, were engaged in campus-wide conversations about records management, including records management in the cloud
- **2011:** Took part in a Records Management Task Force. Identified issues of...
 - “Custody, control and access”
 - Ownership (“Which office is responsible for maintaining the official copy?”)
 - Need for clarity around “self-provisioned cloud storage”; “widespread adoption of cloud-based services [was] driving a loss of control over university content”

Bentley's RM Program and the Cloud

- **2016:** Bentley itself began to develop a formal records management program
- **2016:** New Assistant Records Manager would...
 - Assist units in the identification of “inactive records with long-term, archival value”
 - Provide clarity and guidance for units using cloud storage including “how to create and store records, and how to destroy unneeded information or transfer records to the archives”

CREATION | RECEIPT

This includes the creation of original content within a unit, the capture and integration of content from various sources, and the receipt of information from outside the unit.

DISTRIBUTE

This includes both internal and external distribution, as information that leaves an organization becomes a record of a transaction with others.

(RE)USE

Use and reuse takes place after information is distributed internally, and can generate business decisions, document further actions, or serve other purposes.

MANAGE

Management includes processes such as filing, retrieval and transfers. Information management increases its searchability, sortability, browsability, and shareability.

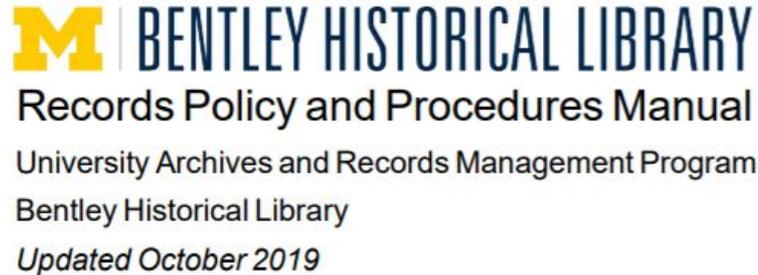
DISPOSE

Disposition is the practice of handling information that is less frequently accessed. Decisions regarding its ongoing retention are made here, often resulting in destruction or transfer to the University Archives.

Overall Approach

- Continues to take an active role in appraisal and in the identification of records with long-term, archival value regardless of where they're stored (e.g., on local University storage, email, removable media, the cloud, etc.)
- Attempts to provide guidance for units creating and managing their own active, semi-active, and inactive records, including those on the cloud
- Develops practical strategies for transferring those records from the cloud to the Bentley

Guidance on Cloud Use from RM Perspective



- Use the move towards a cloud-based collaboration service as a kind of “trigger” for the identification of inactive records and a determination of their final disposition
- Cloud storage services are not appropriate for inactive records with long-term, archival value

Cloud as Medium for Transfer of Records to Bentley

- Bentley archivists have developed strategies for...
 - Transferring data out of these platforms (e.g., using Google Takeout for G Suite)
 - Using these platforms as a convenient means through which to transfer data stored locally
- Not without their drawbacks...
 - “Cloud services, especially personal accounts, may or may not comply with laws, regulations or other policies for sensitive or legally protected information”
 - Generally slower than other methods for larger transfers
 - “Some significant properties of files... may be overwritten when they are downloaded”
- In general, digital information in the cloud is not always captured in ways that are familiar to archivists
 - “Working with cloud transfers” can thus require “an additional layer of management”

None of these obstacles is unsurmountable, of course!

Cloud-Based Services in Bentley Internal Workflows

- Box (with some added steps to verify the integrity and authenticity of what was transferred) is used in a transactional way for some of the most critical and sensitive records the Bentley collects
- Also used to provide individual researchers streaming access to digitized audiovisual material, some of which is copyrighted
- ...
- That said, Bentley archivists have also been intentional about *not* using cloud solutions when they are not appropriate
 - Have recently ensured that material they curate in long-term preservation storage no longer be moved to Amazon Web Services (AWS) cloud storage

Issues With Which Archivists are Grappling

The “Snapshot” Approach and Significant Properties

- “Snapshot” approach to transferring inactive university records of enduring value to the Bentley
 - Tools like Google Takeout export a snapshot of *some data as it existed* on the day it was exported in a *similar but not identical* format
- However, important significant properties or characteristics of those records may be lost
 - In this case, collaborative aspects of, for example, Google Docs, like version history, suggestions, and comments are lost
- Even if they were able to capture those aspects, the Bentley is ill-equipped to provide access to them in any meaningful way

Evidence or Process (or Something Else)

- There is a juridical (and memory, identity, and community) legacy in archives (Cook)
- A key assumption of the “evidence” paradigm is that archives are made up *finite, static* entities that support an assertion
- Due to this shift to the cloud and the dynamic nature of tools optimized for real time collaboration, the “by-products of such interaction are no longer finite entities, but processes that are always changing” (Duranti)
- This calls the evidential value of records into question

The “Custodial” Approach and the Cloud

- BHL has adopted more distributed, postcustodial approach to provide guidance for units creating and managing their records
- Cloud computing considerations are not just technological
 - Organization management, human behavior, regulation, records management
- U-M not subject to state retention schedules
- BHL does not have authority to set retention policies
- U-M ITS provides documentation on platforms but not on how to manage cloud-based business processes, functions, or the data that support them
- BHL has had most success providing guidance for records that will eventually be transferred to the archives

Cloud Computing and Transparency

- “Walled gardens” - closed, corporate-controlled systems
- Prioritization of collaboration/broad participation over transparency
- U-M SPG mandates that IT resources are used in a way that respects the integrity of systems
- Not clear that the systems themselves are designed with integrity and authenticity as a primary goal
- Potential negative consequence of “snapshot” approach

Where Things Stand Now

Box at U-M

[U-M Box](#) is Michigan's implementation of the Box.com cloud storage and collaboration service. Upload, download, share, and edit files via the web and mobile devices.

Notice: Starting in December 2021, the University of Michigan will no longer be offering U-M Box as a storage service. Please visit the [Box Retirement Project](#) site for the latest updates and information.

**“JEEZ LOUISE! This is
incredibly annoying!”**

Conclusion

- U-M IT strategy will continue to prioritize external, cloud-based services
- Archivists at the Bentley will continue to adapt to the shift
- Practical approaches grounded in archival theory
 - “Snapshots” of records using available data-export capabilities
 - Adopting a more distributed, postcustodial approach to guiding units on records management, particularly for those records destined for the archives
 - Maintaining an active role in the appraisal of records to be transferred to the Bentley