

Strategic Plan for Web Archiving at Yale

Web Archiving Working Group

February 2016

Executive summary

The World Wide Web serves as a primary communication medium of the twenty-first century. Yet despite its ubiquity, web content is regularly updated, replaced, deleted or lost, which threatens unique and valuable information. If no action is taken, important primary resources will be lost to current and future scholars and students. Yale University needs to act now in order to protect these digital intellectual and cultural heritage assets and comply with policy and contractual obligations for data retention.¹ Preservation is fundamental to the University's tripartite mission to create, preserve, and disseminate knowledge. In keeping with these priorities, the Web Archiving Working Group (WAWG) recommends the following principles and strategy to advance web archiving at Yale:

- Preservation of certain web content is fundamental to the mission of the University across all of its functional domains
- Collaboration across Yale units, with colleagues at other institutions, with content creators, and with content users is essential to the success of Yale's web archiving strategy
- Web archiving initiatives require a commitment of staff and financial resources in order to succeed
- The rights and intentions of content creators matter and will be taken into account prior to harvesting web content
- Yale will develop shared policies and procedures for web archiving and make them easily accessible.

Within two years, under the leadership of the WAWG, Yale faculty, administrators, staff, and students will recognize the value of web archiving throughout the University. The WAWG will develop University-wide web archiving funding models, begin to address policy and privacy issues associated with web archiving, and cultivate both internal and external partnerships enabling Yale to improve existing practices. Implementing this strategy will allow the University to make informed decisions about its next steps as the theory and practice of web archiving continue to evolve. See Appendix A for a summary of Yale's web archiving efforts to date.

¹ Yale News, "Salovey talks at Davos on preserving cultural heritage," January 2016. Online at: <http://news.yale.edu/videos/salovey-talks-davos-preserving-cultural-heritage?page=13> (accessed 2 February 2016).

Objectives

The objectives outlined below formulate a two-year vision that will establish a solid foundation for web archiving at Yale (see Appendix B for a timeline of objectives). They provide a framework to articulate stakeholder value, foster community and professional collaboration, develop funding models, streamline and expand harvesting, and provide essential guidance on rights issues.

1. Stakeholder value

In FY16, Yale will identify and engage internal and external stakeholders invested in the creation, administration, and use of archived web content. In FY17, key stakeholders recognize and articulate the value of web archiving in relation to the mission of the University.

2. Collaboration

In FY16, Yale units recognize collaboration as a principle that we adhere to as we foster web archiving initiatives across campus. In FY17, Yale's web archiving community looks to expand its collaborative initiatives with external partners.

3. Funding

Yale's web archiving community works to outline sustainable funding models for the future. Yale units currently engaged in web archiving continue to make a financial investment in FY17 and FY18. In addition, Yale stakeholders recognize the importance of collaboration to a University web archiving strategy and jointly fund a two-year term position, starting in FY18, to support community engagement and collaborative efforts toward this strategy. A summary job description is included in Appendix C.

4. Harvesting

By June 2017, Yale's web archiving community will develop and make available a framework for any unit at Yale to make informed website collecting decisions and act on them in an easy, integrated, and well-managed way.

5. Rights

By June 2017, web archiving rights resources will be shared with the general community and the web archiving practice of all key stakeholders at Yale will be informed by them.

Strategy

This Web Archiving Strategy is the culmination of an intensive, strategy-building activity undertaken by the WAWG, a group that consists of representatives from units across the University (see Appendix D).

Web archiving at Yale is a relatively new initiative that will require the continued commitment and expertise of many stakeholders to sustain. The WAWG will initially focus on fostering stakeholder value and further developing the web archiving

community of practice at Yale. Developing a University-wide vision and a supportive web archiving community will empower units to achieve web archiving goals collectively that would be difficult, if not impossible, to accomplish individually. For this reason, collaboration across different Yale departments and offices, as well as with allied partners at other institutions, is a guiding principle.

The WAWG will leverage existing capabilities within multiple units across the University that will contribute staff resources, share technological expertise, commit to open and frequent communication, and develop a distributed funding model to support web archiving initiatives. These early efforts will also result in guidelines, policies, and procedures for web archiving at Yale that will be informed by best practices adopted at other universities and worldwide. By FY18 a dedicated staff member, jointly funded by key stakeholders, will be needed to centrally coordinate web archiving at Yale.

With Yale's web archiving program underway, the WAWG and other stakeholders will work to ensure continuity of practice and keep University policies and procedures up to date. Since the scale of the Internet exceeds the collecting resources of any single department, unit, or organization, Yale's ongoing external partnerships will enable the University's libraries, archives, and museums to continue to build on this solid web archiving foundation and expand the future scope of the Yale's collecting repositories.

Tactics

This section lists the key high-level actions to be taken over the next two years for Yale to accomplish the objectives necessary to realize this University-wide web archiving strategy (see Appendix B for timeline).

Stakeholder and Community Development

1. Develop a community of practice for web archiving and identify potential collaborative projects to build an internal community of practice.
2. Identify potential external project collaborations and engage with the broader web archiving community.
3. Perform a stakeholder analysis and determine a communication plan for all stakeholder groups.
4. Effectively communicate to potential partners the shared resources available to them on campus.
5. Emphasize the ongoing efforts of the Web Archiving Working Group to advance a University-wide program and support existing efforts.
6. Promote web archiving as a priority.

Web Archiving Practice

7. Identify web archiving expertise and tools at Yale and communicate to stakeholders.
8. Learn more about other harvesting tools and/or options for outsourcing.
9. Demonstrate the value of web archiving by developing and gathering use cases, shaping them into case studies, and sharing them with stakeholders.
10. Develop guidelines for the appraisal of web archive collections.

11. Develop tools, educational outreach programs, and training materials to assist stakeholders in understanding the value of web archiving.
12. Gather reference material on web archiving rights and make them available to stakeholders.

Resources

13. Determine what resources are needed and make the case for these resources.
14. Develop a funding model that addresses both University priorities and unit-level priorities.
15. Work with stakeholders to adopt a University-wide funding model.
16. Create a web archiving staff position to facilitate community collaboration and shared objectives (see draft job description in Appendix C).

Policy

17. Develop a shared practice for web archiving at Yale and develop or modify University policies and procedures as needed.
18. Develop a comprehensive framework for all web archiving rights issues, including templates for statements of intent to archive and procedures for addressing ownership and rights.
19. Evaluate the relevance of ADA (Americans with Disabilities Act) compliance issues to web archiving.

Appendix A. Background and Web Archiving Working Group

Background

Although a few Yale staff members have been involved in web archiving initiatives with colleagues at other institutions, until recently there had been no concerted Yale effort to capture and preserve web content. In spring 2014, the Yale Center for British Art initiated a test of Archive-It, a web harvesting service provided by the Internet Archive, as a tool for capturing content from Yale websites and social media. After a successful test period, in 2015 the Beinecke Library, Yale ITS, Manuscripts and Archives, and the Yale Center for British Art entered into a one-year contract with [Archive-It](#). The Archive-It service enables account holders to collect portions of the web and provide full text searching of captured URLs and metadata. Among the sites to be crawled this year are Yale University websites, select social media sites, and websites related to Beinecke's and Manuscripts and Archives' collecting areas.

This pilot year of web archiving has provided Yale with a chance to experiment with a new method of collection development, explore possibilities for describing and providing access to the captured content, and collaborate across campus units. The University-wide [Web Archiving Working Group](#) (WAWG) was established in 2015 to support web archiving activities at Yale and develop a web archiving strategy for the University.

Charge

Develop a web archiving strategy for Yale University, including website harvesting, description of the archived web content, development of access methods, and investigation and management of rights issues. The group will consult with other staff members as needed about digital preservation, description, and discovery; it will also maintain necessary contracts.

Members

Rachel Chatalbash (Yale Center for British Art, co-chair)
Melissa Fournier (Yale Center for British Art)
Gabriela Redwine (Beinecke Library, co-chair)
Michael Lotstein (Manuscripts and Archives)
Maureen Callahan (Manuscripts and Archives)
Daniel Dollar (Yale University Library, sponsor)
Haruko Nakamura (International Collections and Research Support & Humanities Collections and Research Education)
Tang Li (International Collections and Research Support & Humanities Collections and Research Education)
Pam Patterson (Center for Teaching and Learning)
Louis King (Information Technology Services)
Suzanne Lovejoy (Music Library)
Steve Wieda (Yale University Library IT)
Andrea Belair (President's Office)

Appendix B. Objectives Timeline

Objectives	FY 16	FY 17	FY 18
Stakeholder Value	Yale will identify and engage internal and external stakeholders invested in the creation, administration, and use of archived web content.	Key stakeholders recognize and articulate the value of web archiving in relation to the mission of the University.	
Collaboration	Yale units recognize collaboration as a principle that we adhere to as we foster web archiving initiatives across campus.	Yale's web archiving community looks to expand its collaborative initiatives with external partners.	
Funding	Yale's web archiving community works to outline sustainable funding models for the future.	Yale units currently engaged in web archiving continue to make a financial investment.	Yale units currently engaged in web archiving continue to make a financial investment.
		Yale stakeholders recognize the importance of collaboration to a University web archiving strategy and jointly fund a two-year term position, starting in FY18, to support community engagement and collaborative efforts toward this strategy.	Two-year term position begins.
Harvesting		By June 2017, Yale's web archiving community will develop and make available a framework for any unit at Yale to make informed website collecting decisions and act on them in an easy, integrated, and well-managed way.	
Rights		By June 2017, web archiving rights resources will be shared with the general community and the web archiving practice of all key stakeholders at Yale will be informed by them.	

Appendix C: Web Archivist - Sample Job Description

Position Focus:

Under the supervision of [fill in blank], and in collaboration with the Web Archiving Working Group, the position will coordinate web archiving activities throughout Yale University.

The Web Archivist will serve as the main point of contact for all units interested in web archiving, and has responsibility for interacting with units, assisting them as they start web archiving activities, and supporting their ongoing web archiving efforts. The position is responsible for maintaining contracts with vendors, answering technical questions pertaining to web archiving, and working with the Web Archiving Working Group to resolve issues and create workflows related to harvesting, description, rights, access, and long-term digital preservation of captured content. The position will work directly with users in individual units and the Web Archiving Working Group to develop best practices for web archiving at Yale.

This position will promote Yale's web archiving activities internally to a diverse set of stakeholders, while also leading external communications and contributing to inter-institutional collaborative web archiving projects.

The Web Archivist will have the unique opportunity to help develop Yale's web archiving community and program at their beginning stages.

Required Education and Experience:

1. Master's degree from an ALA-accredited library school or equivalent accredited degree, with formal training in archival theory and practice. In selected instances, a post-graduate degree in museum studies or a related discipline in the humanities or social sciences may be substituted for a master's degree in library science.
2. Experience arranging and describing or providing public services for manuscript and/or archival collections.
3. Experience working collaboratively and independently with varied groups within a complex organization and rapidly changing, team environment.
4. Demonstrated knowledge of archival theory and practice may be substituted for formal training.
5. Demonstrated knowledge of current national data content and structure standards related to the archival control of collection materials.
6. Demonstrated knowledge of archival and library management systems.
7. Demonstrated job or school experience with basic preservation and conservation standards for archival and manuscript collections.
8. Demonstrated excellent oral, written, and interpersonal communications and analytical ability.
9. Demonstrated record of designing projects and bringing them to a conclusion in a timely fashion.

Skills:

Technical knowledge of web archiving and/or web technologies and an aptitude for learning new technologies. Ability to adapt readily to change. Creativity, flexibility, and

ability to exercise initiative and sound judgment. Demonstrated ability to work cooperatively within and across departments in a team-based environment.

Excellent communication and organizational skills, including the ability to take direction, accept constructive criticism, and work well in a team-based environment.

Appendix D. Information about the WAWG’s Strategic Planning Process

Using the Web Archiving Working Group’s (WAWG) charge and its goals as a starting point, group members engaged in a four-session strategic planning process facilitated by Kathleen Reilly, a Change Management Specialist in Yale ITS’s Organizational Change Management Department.

Meeting Dates	TOPIC	OUTCOME
November 18	1: Trends and Forces/Key Trends and Implications/Scenario Planning	Identified 5-6 priorities and created scenarios.
November 23	2: Assess Current State/Gap Analysis/Strategic Objectives (Part 1)	Undertook gap analysis and developed strategic objectives for scenarios 1-3.
December 1	3: Assess Current State/Gap Analysis/Strategic Objectives (Part 2)	Undertook gap analysis and developed strategic objectives for scenarios 4-6.
December 2	4: Smartify Objectives and Develop Action Plan	Came away with a draft strategic plan with focused action items for Year 1.

The overall goal for the process was to develop a two-year strategic plan for web archiving at Yale University, with a focused action plan for Year 1. The goals and objectives identified in the main body of this report emerged from the strategic planning process outlined above.