

2017

Collaboration Made It Happen! The Kansas Archive-It Consortium

Cliff Hight

Kansas State University, chight@ksu.edu

Ashley Todd-Diaz

Towson University, atoddiaz@towson.edu

Rebecca Schulte

University of Kansas, bschulte@ku.edu

Michael Church

Kansas Historical Society, mchurch@kshs.org

Follow this and additional works at: <http://digitalcommons.usu.edu/westernarchives>

Recommended Citation

Hight, Cliff; Todd-Diaz, Ashley; Schulte, Rebecca; and Church, Michael (2017) "Collaboration Made It Happen! The Kansas Archive-It Consortium," *Journal of Western Archives*: Vol. 8 : Iss. 2 , Article 4.

Available at: <http://digitalcommons.usu.edu/westernarchives/vol8/iss2/4>

This Case Study is brought to you for free and open access by the Journals at DigitalCommons@USU. It has been accepted for inclusion in Journal of Western Archives by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.



Collaboration Made It Happen! The Kansas Archive-It Consortium

Cliff Hight
Ashley Todd-Diaz
Rebecca Schulte
Michael Church

ABSTRACT

This case study explores the formation, current membership, and future goals of the Kansas Archive-It Consortium (KAIC), one of the larger consortia contracting with the Web archiving service Archive-It. KAIC, which is composed of the state historical society and five public universities, has its foundation in a statewide culture of collaboration, and participants have agreed on an informal governance structure with a strong commitment to broadening accessible Web resources for researchers. After establishing consortial consistency during its first two years, members have shared documentation with partners and are beginning to do collaborative collecting. In the future, the consortium will seek additional members and work with Archive-It to develop a consortial search tool. This Web archiving collaborative has helped member institutions overcome challenges by having group discussions, sharing documentation and guidelines, and jointly serving a primary user group, Kansas residents.

Introduction

The Kansas Archive-It Consortium (KAIC) is a Web archiving partnership among six institutions in the state of Kansas—the state historical society and five public universities—that has contracted with the Web archiving service Archive-It. Current partners include Kansas Historical Society, Emporia State University, Fort Hays State University, Kansas State University, University of Kansas, and Washburn University. As one of the larger consortia in Archive-It's portfolio, KAIC is another manifestation of the culture of collaboration that has developed among Kansas libraries and archives during the past two decades. Further, it benefits consortial members by allowing them to learn from one another and helps a primary user group, Kansas residents, by seeking to create a single source for Kansas-related Web archives. This article is a case study of the Web archiving consortium's establishment, the benefits

and drawbacks of collaboration, and how establishing Web archiving partnerships improves the resources available to researchers.

Literature Review

Professional literature touches on each of these topics—the value of Web archiving, establishing Web archives, and collaborative Web archiving—and reviewing some of these writings adds context for this case study.

The Internet has revolutionized how people create, publish, and share information. It has changed how communication occurs and important events are documented and remembered by opening an easy dialog between information creators that enriches and deepens discussions and intellectual discovery. However, since so much human communication and interaction now occur in the online environment instantaneously around the world, archivists have a responsibility to develop reliable and sustainable methods for preserving this vital source of history and culture.

According to the International Internet Preservation Consortium (IIPC), Web archiving is “the process of collecting portions of the World Wide Web, preserving the collections in an archival format, and then serving the archives for access and use.”¹ This task can become daunting considering Web content is not only abundant, but also unusually ephemeral.² Creators can modify or remove information published on the Web at any time, adding an element of urgency to the value of Web archiving.

With regard to establishing Web archives, the Internet Archive charted the way in 1996, followed by the Library of Congress in 2000, the IIPC in 2003, and the British Library in 2004. It is only more recently that smaller institutions have had the means and training to begin thinking about establishing Web archives.³ Although Web archives follow the same general lifecycle as traditional archives, which includes selection, acquisition, arrangement, description, preservation, and access, Web archives are unique enough to require their own standards, best practices, and graduate courses.⁴ One international study revealed how Web archiving initiatives aligned in terms of size, scope, workflows, and access models. Of note, they

1. “Why Archive the Web?,” International Internet Preservation Consortium, <http://netpreserve.org/web-archiving/overview> (accessed November 29, 2016).
2. Hugo C. Huurdeman et al., “Lost But Not Forgotten: Finding Pages on the Unarchived Web,” *International Journal on Digital Libraries* 16, no. 3 (September 2015): 247-265, <http://dx.doi.org/10.1007/s00799-015-0153-3> (accessed November 29, 2016).
3. Miguel Costa, Daniel Gomes, and Mário J. Silva, “The Evolution of Web Archiving,” *International Journal of Digital Libraries* (May 2016): 1-15, <http://dx.doi.org/10.1007/s00799-016-0171-9> (accessed November 29, 2016).
4. Jinfang Niu, “An Overview of Web Archiving,” *D-Lib Magazine* 18, no. 3/4 (March/April 2012), <http://dx.doi.org/10.1045/march2012-niu1> (accessed November 29, 2016).

discovered that Web archives tend to be managed by small staffs, driven by the concern for losing Web content, and focused more on the processes of acquisition and curation of content.⁵

The first step in establishing a Web archives is determining selection guidelines, which often are based on domain (e.g., .edu, .gov, .ca), topic or event, or format. Using automated software and Web crawlers, such as Nutch and Heritrix, libraries and archives can easily establish objective criteria and a collecting scope that can help streamline the acquisition process. Similar to traditional archives, it is important to develop a collecting policy for Web content to maintain the scope of what is collected and to plan storage space requirements by considering the size of certain media and formats, the extent of certain domains, and the complexity of certain topics or events. Recent studies have approximated that even after the selection process, the amount of Web content preserved internationally is calculated in petabytes.⁶

Another step in creating Web archives is to decide whether to seek creators' permissions and how to maintain the authenticity and integrity of the content.⁷ Additionally, archivists must describe the content to support its access and use. Various metadata schemas exist to aid description, including Dublin Core, which the Internet Archive uses, and the Metadata Object Description Schema (MODS), which the Library of Congress employs. Also, OCLC has established a Web Archiving Metadata Working Group made up of 26 representatives from Research Library Partnership institutions to develop a flexible set of best practices for metadata. The results of this effort are expected in 2017.⁸

An additional institutional decision is whether to provide access or keep Web archives dark. Initiatives like the UK Web Archive and the Internet Archive's Wayback Machine provide public access to archived Web content; however, Jinfang Niu noted that some of these initiatives institute an embargo ranging from a few months to a year on recently captured Web content so as not to compete with active websites.⁹

Although it has been two decades since initial forays into Web archiving, it is still an emerging field in archivy. As such, there are many cases in which collaborative approaches have proven more useful than individual efforts. In 2003, twelve

5. Costa, Gomes, and Silva, "The Evolution of Web Archiving," 7-13.
6. Huurdeman et al., "Lost But Not Forgotten," 247-248; Costa, Gomes, and Silva, "The Evolution of Web Archiving," 2.
7. Niu, "An Overview of Web Archiving," 4, 6.
8. "The Problem," OCLC Web Archiving Metadata Working Group, last modified March 10, 2016, <http://www.oclc.org/research/themes/research-collections/wam.html> (accessed December 1, 2016).
9. Niu, "An Overview of Web Archiving," 6.

organizations united to form the IIPC with the common goal of working together to jointly fund projects and develop Web archiving solutions.¹⁰ The British Library, a founding member of IIPC, has broadened its collaborative initiatives by joining forces with five other legal deposit libraries in the United Kingdom to develop the UK Web Archive in 2013.¹¹ Not only do these collaborations assist partners with appraising a broader range of Web content, but they also provide joint problem solving, manpower, and the potential for pooled funding and shared costs.

The professional literature on the value of Web archiving, establishing Web archives, and collaborative Web archiving has influenced KAIC partners during the creation of this consortium, and the literature continues to affect members while working through present challenges and developments.

Beginning the Kansas Archive-It Consortium

Individuals and organizations in Kansas have been creating Web content for over two decades, but until recently archival repositories in the state have made limited efforts to preserve these materials. Many of these cultural heritage institutions are under Kansas state statutory obligation to preserve and provide access to appropriate public records in all formats—including those that now are available only on the Web—but limited resources at these institutions have slowed efforts to meet this requirement.¹² Archivists began strategizing on their own in the past ten to fifteen years to understand and use existing technological tools and services to meet their Web archiving needs.

In addition to these individual efforts to preserve Web content within the past decade or more, there is a broader culture of collaboration among Kansas libraries and archives that has existed for fifty years. This *esprit de corps* has influenced how Kansas organizations successfully partnered in a Web archiving consortium; a brief review of these earlier projects among some KAIC members—Kansas Regional Library System, Territorial Kansas Online, and Kansas State Publications Archival Collection—will provide a clear understanding of the current partnership's foundation.

10. "About IIPC," International Internet Preservation Consortium, <http://netpreserve.org/about-us> (accessed November 29, 2016).

11. "About," UK Web Archive, <https://www.webarchive.org.uk/ukwa/info/about> (accessed November 29, 2016).

12. The relevant statutes and regulations from the State of Kansas include: Agency 53, State Records Board, http://web.archive.org/web/20160215165049/www.sos.ks.gov/pubs/kar/2009/4%20053_53-State%20Records%20Board,%202009%20KAR%20Vol%204.pdf (accessed August 13, 2016); 2016 Kansas Statutes Chapter 75.—State Department; Public Officers and Employees, Article 35, 75-3501 et seq., accessed August 13, 2016, http://www.ksrevisor.org/statutes/ksa_ch75.html; and 2015 Kansas Statutes Chapter 45.—Public Records, Documents and Information, Article 4, 45-401 et seq, http://www.ksrevisor.org/statutes/ksa_ch45.html (accessed August 13, 2016).

In 1965 the State of Kansas created statutory requirements for improving library services through a state-wide regional library system using federal funds from what is known today as the Library Services and Technology Act (LSTA). Because LSTA funds primarily support programs at many smaller libraries in the state, it does not allow the budgetary resources necessary to underwrite collaborative, statewide digital programs.

Nonetheless, the Kansas Historical Society (KHS) and the Kansas Collection at the University of Kansas (KU) Libraries used a grant from the Institute of Museum and Library Services to develop the Territorial Kansas Online project from 1999 to 2001.¹³ This virtual repository uses more than 2,000 items from KHS and KU to describe the early years of Kansas Territory. Staff at both institutions collaborated to develop subject content, to establish curriculum objectives, and to select and digitize items for inclusion in the virtual repository.

In 2007, KHS and the State Library of Kansas partnered on a DSpace-based electronic archive, the Kansas State Publications Archival Collection (KSPACE), for online publications of state government agencies.¹⁴ It grew as KHS and State Library staff used manual processes to add content. By 2013, staff had increasing difficulties managing and updating the site and, after periodic outages limited staff and public access, KHS took the site down entirely in 2015. Since then the State Library has migrated content into a hosted CONTENTdm site called the Kansas Government Information (KGI) Online Library.

All of these joint ventures by Kansas libraries and archives have established a culture of collaboration in the state, and KAIC is a manifestation of this mindset. For more than 50 years the state's information and cultural heritage organizations have been working together to overcome challenges and achieve progress by pooling resources, talents, and insight. Despite this history, KAIC is the first initiative that archivists have created among state institutions. Not only does this collaboration benefit the consortial members by allowing them to work together, but it also benefits a primary user group, Kansas residents, by creating a single source for Kansas-related digital archival materials and primary sources.

Although a collaborative culture has existed among Kansas libraries and archives for decades, initial efforts in Web archiving occurred individually. It took time for staff at these institutions to recognize the power of collaboration in the Web archiving realm. In 2000, KU began to explore the preservation of Web content. The project was undertaken by the bibliographer of the Wilcox Collection of Contemporary Political Movements, one of the largest collections of American left-

13. The Territorial Kansas project website is <http://www.territorialkansasonline.org> (accessed August 13, 2016). Archived versions of the site are available at http://web.archive.org/web/*/www.territorialkansasonline.org.

14. Archived versions of the KSPACE site are at http://web.archive.org/web/*/kspace.org/.

wing and right-wing organizations in the United States. Because political groups had shifted their information sharing to the Web by the mid-1990s, KU staff began testing and reviewing tools for Web archiving. They determined that the offline browsing webspider Teleport Pro was superior to other software products they tested.¹⁵ For a year, the project successfully captured more than 150 organizations' websites, which staff then arranged by broad topics. There was much uncertainty at the time about how to treat websites within a library or archive context, and there were no clear guidelines for description or access. KU archival and metadata staff met to start developing descriptive guidelines, but staff turnover and organizational change led to the end of the project in 2001.

Another project occurred in 2006 and 2007 when KHS participated in testing the Web Archives Workbench (WAW) beta tools developed under the ECHO DEpository, a project that was part of the National Digital Information Infrastructure and Preservation Program (NDIIPP). According to the 2007 WAW user guide, "OCLC's Web Archives Workbench is a suite of Web archiving tools for identifying, selecting, describing and harvesting Web-based content for ingest into an external digital repository. It is based on an archival model for selecting digital materials for preservation developed by the Arizona State Library."¹⁶ KHS developed a seed or "entity" list of many Kansas state agencies to use for running test crawls. Ultimately KHS staff pursued a detailed analysis of select agency websites including the Kansas Department of Transportation (KDOT) website and held discussions with KDOT on the Web archiving project within the context of a larger records management strategy. Although KHS staff was initially optimistic about the WAW tool, KHS stopped using it soon after testing. Contributors to this decision were necessary improvements from the application developers, a lack of broad support for WAW, shifting institutional priorities, and the growing success of and interest in vendor-based Web archiving tools, particularly the Internet Archive's Archive-It service.

KU embarked on another individual Web archiving effort in late 2010 with some inaugural testing of Archive-It. Project personnel initially included KU Libraries' Assistant Dean for Collections and Scholar Services, Web Administrator, and University Archivist. The latter two worked with staff at Archive-It before and after the first trial crawl of the University Chancellor's website. There were challenges with capturing the right file directories because of the robots.txt file on that site, but scoping the crawl differently served as a workaround. After successfully capturing a more complete representation of the site and reviewing websites of other universities with Archive-It agreements, KU decided in 2012 to contract with Archive-It.

15. Tennyson Maxwell Information Systems, Inc., created Teleport Pro; additional information can be found at <http://tenmax.com/>. Useful features included the ability to limit the type and size of files captured, the creation of an exact duplicate or mirror of the website, and an organizational structure for the data file of the site captured. Also, it did not place its own advertising banner on the sites captured as did the other software tested.
16. OCLC, Web Archives Workbench BETA User Guide, March 9, 2007, https://www.ideals.illinois.edu/bitstream/handle/2142/9458/Final_WAW_User_Guide.pdf?sequence=2 (accessed August 13, 2016).

At Kansas State University (KSU), the University Archivist began exploring Web archiving options in 2011, including participation in an Archive-It informational webinar and trial, as well as looking at tools from other vendors. Other repositories in KAIC had not done any Web archiving tests previous to the development of the consortium.

Amidst the individual Web archiving tests and explorations, archivists began discussing possible collaborations. The University Archivist at KSU corresponded with the State Archivist of Kansas at KHS about Web archiving in 2011. Both repositories are public agencies with responsibilities to preserve public records, and these archivists wondered if it would be viable to have KHS partner with Kansas Board of Regents' universities and enter into a single contract with a Web archiving vendor to capture Web content, reduce duplication, and maximize resources. The idea percolated in the following years and at times included discussions with other university archivists in Kansas during meetings of the state's Electronic Recordkeeping Committee and the Kansas State Historical Records Advisory Board.

In December 2013, archivists from eight Kansas institutions (KHS and seven public universities) met to gauge interest in devoting resources to Web archiving and doing it collaboratively. Initially, the group discussed many topics related to Web archiving: the migration of organizational and personal records to the Web, how to effectively search and display Web archives, what vendors were available, and how to collaboratively develop policies and procedures. Further, participants considered governance options, possible funding opportunities, and potential concerns with pursuing the collaborative. The group revisited many of these topics in meetings that occurred during 2014, resulting in the formation of an official consortium by the end of that year. This partnership allows each participant to have its own account that is managed individually and also have access to collaborators who can share lessons learned.

In developing the consortium, participants recognized that the considerable gains in efficiency and economy could offset limitations of state and institutional demographics. Collaborative collection of Web-based materials could reduce duplication of preserved content, enhance workflow and documentation sharing, and stretch limited budgets. The relative geographic isolation of most member institutions, however, would present a significant hurdle to face-to-face collaboration and technology support. Fort Hays State University, for example, is located more than 150 miles from its nearest fellow participant. This distance, coupled with one of the lowest population densities in the USA, would limit physical access to and availability of other professionals pursuing Web archiving.¹⁷ Despite their geographic dispersion, participants decided to capitalize on their collective bargaining strength when a comparative cost analysis of a consortial agreement indicated a significant savings

17. "Kansas QuickFacts," U.S. Census Bureau, 2015, <http://web.archive.org/web/20150905072233/quickfacts.census.gov/qfd/states/20000.html> (accessed August 13, 2016). In 2010, the population density in Kansas was 34.9 people per square mile, ranking 42nd in the nation.

over individual contracts. Partners coping with budgetary constraints began to view the consortium not only as a matter of convenience, but a smart fiscal decision. In the latter half of 2014, negotiations ended and an agreement was finalized.

Consortium Member Profiles

After six of the eight initial institutions committed to Archive-It, participants settled on naming the group the Kansas Archive-It Consortium, or KAIC. The two institutions that did not continue with the group either did not have the resources to join or chose to go a different direction with Web archiving. The inaugural members—profiled individually in Table 1 and below—included Emporia State University (ESU), Fort Hays State University (FHSU), Kansas Historical Society (KHS), Kansas State University (KSU), University of Kansas (KU), and Washburn University (WU).

Name	Founded	2015 Enrollment	Repository FTE	Web Archiving FTE
Emporia State University	1863	6,094	2.0	0.05
Fort Hays State University	1902	14,210	1.0	0.05
Kansas Historical Society	1875	n/a	17.0	0.8
Kansas State University	1863	24,146	9.0	0.07
University of Kansas	1865	28,091	11.0	0.2
Washburn University	1865	7,945	1.0	0.01

Table 1. Snapshot of KAIC Institutions

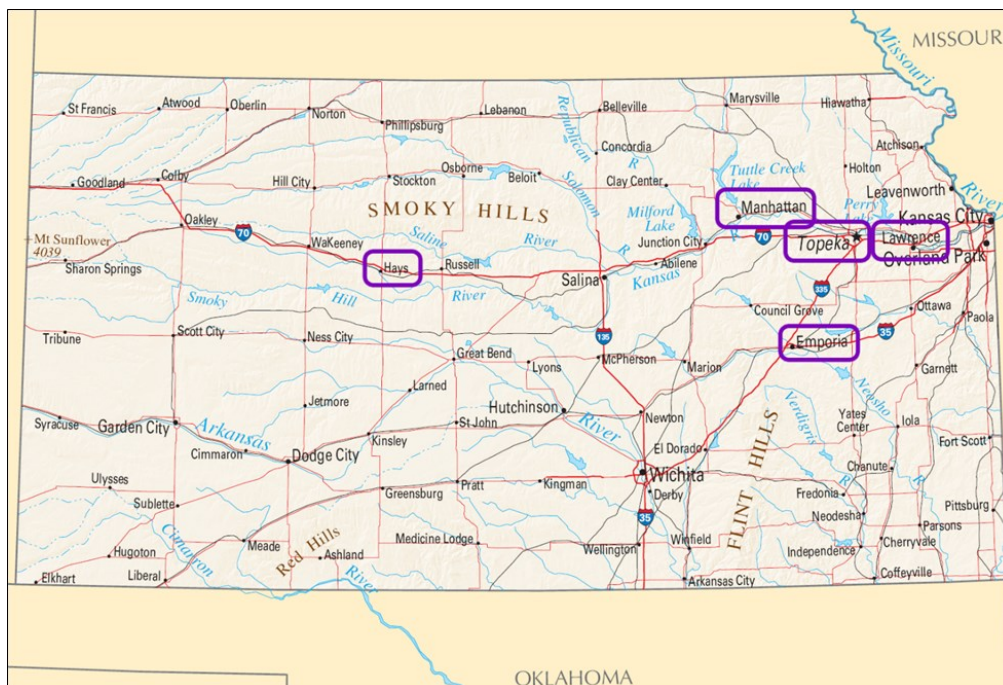


Figure 1. Map of Kansas with KAIC members' locations circled in purple (Credit: The National Map, U.S. Geological Survey)

Emporia State University (ESU) is a small regional university located approximately 100 miles southwest of Kansas City. The University's Special Collections and Archives acts as the institutional memory, maintaining a primary collecting focus of ESU-related materials. In 2013 ESU celebrated its sesquicentennial year, which presented numerous opportunities for its Special Collections and Archives to identify where gaps existed in their collections. As of 2013 digital materials were not actively being collected, though a campus-wide records survey revealed that the majority of materials included in the institution's records retention schedule were only created digitally. The department needed to establish a digital archive to address this shortcoming, in part by joining KAIC as an early step towards creating a digital initiative to collect both Web content and digital records.¹⁸

Fort Hays State University (FHSU) is the largest university in western Kansas—about 270 miles west of Kansas City—and joined KAIC to increase its capacity to preserve news and events that were being promoted predominately on the University's website. The main goal for Web archiving at FHSU is to provide the

18. The public Web archive for Emporia State University (ESU) is found at <https://archive-it.org/organizations/892>.

university community with preserved content for future generations of researchers.¹⁹ Additional benefits of participating in KAIC include greater access to professional expertise and improved communication with partners.

The Kansas Historical Society (KHS) in Topeka seeks to harvest and preserve select Web content created by Kansas state government, businesses, and organizations in order to further its mission to facilitate government accountability, economic development, and the education of Kansans. Through its participation in KAIC, KHS has harvested Web content from more than 300 websites in five collecting areas including websites pertaining to KHS collections, political organizations of Kansas, Kansas state government agencies, community organizations of Kansas, and historical and genealogical societies of Kansas.²⁰ KHS became a member of KAIC in order to realize the financial and communal benefits from working collaboratively and cooperatively with like institutions on the acquisition, preservation, and distribution of Web content unique to Kansas.

Kansas State University (KSU) is a land-grant university based in Manhattan with its University Archives housing historically-significant institutional records. Recently, it has expanded to become a repository for appropriate born-digital materials, including Web content. KSU participated in KAIC because personnel recognized benefits from collaboration would include improved financial value, access to expertise beyond the institution, and the opportunity to improve the research experience for users. KSU's goals for Web archiving include crawling and preserving Web-based documentation that aligns with subject strengths specified in its collection development guidelines.²¹

The University of Kansas (KU) is a major public research and teaching institution based in Lawrence with five regional campuses. KU Libraries constitute the largest library in the state, and has a growing collection of digital assets that includes the websites of University offices and departments harvested since 2013.²² Other research collections include the Kansas Collection and the aforementioned Wilcox Collection of Contemporary Political Movements. The archivists and curators responsible for managing these collections plan to identify, harvest, and make available websites that document these collecting areas. KU joined KAIC in part because of its tradition of cooperating with other Kansas institutions, as well as to help pool limited resources to create a successful statewide project that benefits participants and user groups.

19. The public Web archive for Fort Hayes State University (FHSU) is found at <https://archive-it.org/organizations/891>.

20. The public Web archive for Kansas Historical Society (KHS) is found at <https://archive-it.org/organizations/606>.

21. The public Web archive for Kansas State University (KSU) is found at <https://archive-it.org/organizations/890>.

22. The public Web archive for University of Kansas (KU) is found at <https://archive-it.org/collections/3577>.

Washburn University (WU) is a publicly-funded, locally-governed university in Topeka. The University Archives and Special Collections focuses on preserving institutional memory. Because records creation has shifted predominately to digital formats, WU needed a more comprehensive digital solution, especially for websites. Because of insufficient personnel and technical expertise to handle these emerging records formats effectively, WU joined KAIC to utilize the knowledge and cost-effectiveness of a statewide partnership. By using Archive-It to preserve important websites and pages, WU is beginning to preserve digital records and make them more easily accessible.²³

Organizational Structure and Administration

During the initial KAIC meetings, members discussed a number of topics related to structuring and administering the organization. Conversations touched on goals, governance, meeting logistics, financial administration, and elements that became part of the group's memorandum of understanding (MOU). As participants considered institutional and consortial goals, they agreed to pursue cost savings, create and share Web archiving efficiencies, and develop collaborative collecting.

KAIC members discussed a governance model during early meetings and initially agreed on a steering committee with a rotating chairmanship. Later, the group altered its approach and settled on an informal organizational structure. Currently, the University Archivist at KSU serves as the project manager with duties that include developing meeting agendas, following up on assignments, serving as a primary point of contact with Archive-It, and maintaining effective communication among partners. The main reason for the adaptation in governance was that members recognized formality at such an early stage could limit flexibility and that organic development would occur as the group matured.

Another concern with moving to a rotating chair model was the group needing to establish some consistent practices—such as regular meetings and consortial guidelines—to increase the likelihood of sustained participation from the members. Additionally, a project manager approach presently provides stability over a rotating chair model during fluid budgetary times at member institutions and acts as a stabilizer when such budget volatility increases the possibility of layoffs or furloughs, which could further reduce Web archiving staffing. As the consortium develops and partners maintain consistent involvement, the time may come when the group is prepared to move to such a governance model.

There was a stretch of 14 months when KAIC did not officially meet—mainly during trial and training periods with Archive-It—which hindered collaborative opportunities while learning to use the tool. In early 2015, the group decided to start

23. The public Web archive for Washburn University (WU) is found at <https://archive-it.org/organizations/975>.

meeting quarterly to review progress, address challenges, and share information. In a state that is roughly 400 miles east to west and over 200 miles north to south, distance is an issue to consider when scheduling any statewide meetings. Thankfully, technological advancements have minimized this challenge as the group has chosen to conduct most of its meetings via conference calls. This allows each of the partners to be equally involved in discussions and decisions. The only exception is an annual face-to-face meeting to reinforce consortial relationships and to discuss items that might be best handled in person.

Consortia such as KAIC allow Archive-It to streamline new subscriptions by negotiating an agreement with and receiving payment from one institution on behalf of other consortial members. Archive-It staff helped encourage KAIC members to subscribe by identifying possible account sizes for each institution and noting the cost savings associated by participating in the consortium. Because all participants are using budgeted allocations with no soft funding, these cost reductions represent an important savings for essential services. Additionally, the continual tightening of institutional budgets spotlights the cost benefits of shared contracts.

Archive-It requires a single billing entity when making consortial agreements, and KHS fulfills that obligation during the annual renewal period that is based on a fiscal calendar. Because all KAIC members are state agencies or Board of Regents institutions, KHS is able to request reimbursement of each member's subscription rate through an interfund transfer. Terms governing the reimbursement of KHS funds are included in an MOU among all KAIC members and KHS. The latter terms are required to comply with the Kansas Prompt Payment Act.²⁴

As the contracting party, KHS negotiated on behalf of KAIC to establish terms of the service agreement with Archive-It. This process was much easier for KHS than expected, mainly because KHS's previous negotiations with commercial vendors included rights issues over digitized content. Because Archive-It serves a different purpose than previous vendors, the agreement process took less time and concluded a main contract outlining general provisions and an exhibit with each member's account level and subscription cost. Each year an additional exhibit is incorporated into the agreement including the specific terms of service and allotments for the new term. This flexibility allows KAIC members to change institutional account levels based on the changing needs of each organization. For example, KSU recognized its data and document allotments during the first two years were much larger than currently necessary, so it reduced its account level to align more closely with previous yearly data and document amounts.

Archive-It has worked with KAIC members when unique contractual situations have arisen. For example, since KU already had an individual agreement, Archive-It

24. 2016 Kansas Statutes Chapter 75.—State Departments; Public Officers and Employees, Article 64, 75-6401 et seq., http://www.ksrevisor.org/statutes/ksa_ch75.html (accessed August 13, 2016).

transferred the new terms to the consortial agreement and pro-rated the service fee difference for the time between the individual contract and the consortial agreement.

Name	Account Size	Seeds	Crawls	Crawl Frequency
Emporia State University	128 GB 3M (million) URLs	Total: 53 Public: 0	Total: 740 Test: 22	One-time: 4 Weekly: 3 Bi-monthly: 14 Monthly: 17 Quarterly: 10 Annual: 5
Fort Hays State University	128 GB 3M URLs	Total: 17 Public: 17	Total: 100 Test: 3	One-time: 1 Monthly: 4 Quarterly: 3 Semi-annual: 6 Annual: 4
Kansas Historical Society	768 GB 8M URLs	Total: 400 Public: 0	Total: 333 Test: 116	One-time: 9 Daily: 1 Weekly: 15 Monthly: 27 Semi-annual: 169 Annual: 167
Kansas State University	256 GB 4M URLs	Total: 305 Public: 11	Total: 132 Test: 78	One-time: 281 Bi-monthly: 2 Quarterly: 7 Semi-annual: 2 Annual: 4
University of Kansas	512 GB 6M URLs	Total: 572 Public: 100	Total: 224 Test: 142	One-time: 357 Weekly: 1 Annual: 2
Washburn University	256 GB 4M URLs	Total: 20 Public: 0	Total: 4 Test: 3	One-time: 1

Table 2. KAIC Web Archiving Metrics (as of 11/30/2016)

As an informal organization, KAIC currently has no founding documents or bylaws. In the absence of such documentation, the aforementioned MOU governs member subscriptions. It establishes agreement among the individual members that KHS shall act as the contracting party and billing authority, documents KAIC

members' approval and acceptance of the Archive-It service agreement (attached as an exhibit to the MOU), and formalizes procedures governing payment of the subscription costs and reimbursement of each member's cost to KHS. Each participant receives a copy of the MOU via email from KHS and signs and returns an approval sheet indicating their acceptance of the terms and conditions. Every renewal term requires an updated MOU that documents KAIC members' acceptance of the terms of service and allotments outlined in the new exhibit to the Archive-It service agreement for that term.

Beyond this formal agreement among members there is great flexibility that each partner has in establishing an appropriate account level with Archive-It. KAIC participants have four different account levels based on their individual institutional needs, which is very helpful for organizations of varying sizes and missions. Archive-It's adaptability only strengthens its position as an effective Web archiving service. Such tractability was another reason KAIC members decided to enter an agreement with Archive-It.

In addition to the varying account levels, KAIC partners have developed independent seeds and crawl frequencies. Table 2 provides summary metrics that describe account sizes, numbers of seeds and crawls, and crawl frequencies.

Web Archiving Efficiencies

An important goal for the group was to create and share Web archiving efficiencies with partners. These efforts have included sharing institutional Web archiving documentation, such as metadata guidelines, to help members develop efficient practices. An additional purpose of this goal is to build capacity in smaller organizations since the larger repositories usually have more resources to devote to Web archiving.

The consortium is not designed to force participants to do Web archiving in one specific way. For example, each institution can establish its own metadata standards, crawl parameters, and account display information. KAIC has not established any requirements for these and other facets, instead it has only provided sources for participants to develop their own Web archiving work.

Regarding the metadata guidelines, these are contributions from individual participants; currently there is limited interest in combining them into one set of required guidelines across the complete group. Because resources are different at each institution, KAIC seeks to provide options for members that are malleable to each participant's needs. At this time KHS, KSU, and KU have shared guidelines that outline the level at which Web archives will be described at their respective locations.²⁵ These guidelines, as well as those that will be shared by other

25. See Appendix A for sample metadata guidelines from KHS, KSU, and KU.

participants, can help members adapt existing practices rather than invest additional resources in developing them from scratch.

Collaborative Collection Development

Related to discussions of efficiencies is KAIC's goal to develop ways to collaboratively collect Web content. An early step has been the creation of a shared spreadsheet listing all seeds crawled by KAIC members, which currently numbers over 1,300 URLs. Each institution has its own sheet and contributes the pages it is preserving. This resource provides partners with a single location to check to prevent duplicative crawling. Further, it allows members to suggest sites that other institutions may wish to preserve.

Although most partners currently focus primarily on collecting the Web presence of their own institutions, many of their collecting policies also include specialized areas, such as the consumer movement or information about famous Kansans. There is limited concern for one institution to inadvertently collect materials created by another member, yet the chances are higher of two institutions potentially collecting the same website because its content fits both collecting scopes. While the practice of closely following a collection development policy is something familiar to archivists, the decision to capture a version of a website may be much easier to make than the decision to pursue a one-of-a-kind manuscript collection. Because multiple people can archive a website simultaneously, the decision to crawl becomes one of convenience, while also looking to avoid the duplication of effort, rather than one of ethics. Due to this change brought on by the digital age, members have begun discussing collaborative collecting. To provide context, Table 3 provides current Web archive collecting foci of KAIC partners.

In a certain sense, KAIC applies elements from the acquisition concept of the documentation strategy to Web archiving.²⁶ Typically, collaborating institutions that acquire materials documenting a specific topic, time period, or geographic area carry

26. Professional literature on documentation strategy includes: Helen Willa Samuels, "Who Controls the Past?," *American Archivist* 49, no. 2 (Spring 1986), 109–124, <http://americanarchivist.org/doi/pdf/10.17723/aarc.49.2.t76m2130txw40746> (accessed August 15, 2016); Mark Greene and Todd Daniels-Howell, "Documentation with an Attitude: A Pragmatist's Guide to the Selection and Acquisition of Modern Business Records," in *The Records of American Business* ed. James O'Toole (Chicago: Society of American Archivists, 1997), 161–229, <https://babel.hathitrust.org/cgi/pt?id=mdp.39015071452240> (accessed August 15, 2016); Doris J. Malkmus, "Documentation Strategy: Mastodon or Retro-Success?," *American Archivist* 71, no. 2 (Fall/Winter 2008), 384–409, <http://americanarchivist.org/doi/pdf/10.17723/aarc.71.2.v63t471576057107> (accessed August 15, 2016); and Larry Hackman, "The Origins of Documentation Strategies in Context: Recollections and Reflections," *American Archivist* 72, no. 2 (Fall/Winter 2009), 436–459, <http://americanarchivist.org/doi/pdf/10.17723/aarc.72.2.g401052h82h12pm3> (accessed August 15, 2016).

Name	Collecting Area(s)
Emporia State University	<ul style="list-style-type: none"> • University web presence
Fort Hays State University	<ul style="list-style-type: none"> • University web presence
Kansas Historical Society	<ul style="list-style-type: none"> • State government agencies • State political organizations • Community organizations of Kansas • Historical and genealogical societies of Kansas • Kansas life and culture (specifically, websites relevant to KHS collections) • Kansas newspapers
Kansas State University	<ul style="list-style-type: none"> • Consumer Movement • Cookery • Kansas life and culture (focus on agriculture and rural life, grain science and milling, and Kansans in the military) • University history (including institutional records, student organizations, and affiliated organizations)
University of Kansas	<ul style="list-style-type: none"> • University history (including institutional records, student life, and culture) • Kansas life and culture (focus on the African American experience and websites relevant to KU collections) • American left and right wing political movements for the Wilcox Collection of Contemporary Political Movements
Washburn University	<ul style="list-style-type: none"> • University web presence

Table 3. Current Web Archive Collecting Areas of KAIC Partners

out this approach.²⁷ With Web archiving, repositories in a partnership like KAIC share their collecting policies and determine how to collectively acquire content on a particular topic without duplicating each other. This method requires all consortium members to define their collecting scope and share it with the group. Communication is paramount in developing trusting relationships that result in building a joint collecting scope that is inclusive and representative of what the individual institutions are already undertaking.

For example, KHS, KSU, and KU have an overlapping collecting area related to documenting Kansas life and culture. An October 2016 face-to-face consortial meeting focused on collaborative collecting and steps the group can take to reduce duplicative Web archiving. One action was to share each institution's collecting policy so partners could identify similar topical areas. An example discussed at the meeting was the records of the League of Women Voters of Kansas, which has identified the Kansas Collection at KU as the official repository for its records. Yet, there is also a small amount of physical records at KHS, which led KHS to include the League's website on its list of potential sites to crawl. The repositories communicated and decided KU should handle the preservation of that site. As the consortium continues meeting, partners will formalize at broad levels each repositories' collecting boundaries as they seek to preserve the Web content documenting specific realms of Kansas life and culture.

Collaborative collection development will also include some partners providing additional support to other members. While Archive-It provides most of the training resources and technical assistance for individual members, KAIC encourages members to help each other overcome common problems or improve results through cross training. Agendas for every consortial meeting include time for each partner to discuss difficulties with any facet of Web archiving. One example where this openness proved beneficial occurred when KHS provided supplementary support so that Washburn University (WU) could continue as a consortial member by training staff to initiate and scope crawls.

WU is a founding member of KAIC, but staffing levels and institutional priorities have limited their Web archiving efforts to the point that no crawls had occurred during the first two years of the consortium. WU mentioned during a membership renewal discussion with KHS that they might have to leave the consortium due to inactivity. They recognized that much of the Web archiving work is the initial test crawls and scoping, and suggested they would be more likely to continue if they could receive help with those efforts. KHS agreed to have one staff member provide

27. Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: The Society of American Archivists, 2005), 131, <http://web.archive.org/web/2015127181745/files.archivists.org/pubs/free/SAA-Glossary-2005.pdf> (accessed August 15, 2016); and "Documentation Strategy," Society of American Archivists, 2015, <http://web.archive.org/web/2015128214205/www2.archivists.org/glossary/terms/d/documentation-strategy> (accessed August 15, 2016).

limited assistance in training WU personnel to create test crawls and revise scoping rules to preserve appropriate Web content. WU had already created a prioritized seed list, but had not added it to Archive-It. When KHS and WU staff met, they added the tier one seeds to Archive-It and initiated test crawls on those seeds. During a second session, KHS staff explained how to review test crawls and re-scope them before initiating the actual crawl. Although KHS provided only about three hours of assistance, it was enough for WU to remain in the consortium.

Lessons Learned

In just over three years, KAIC has moved from an idea to a collaborative Web archiving group with members assisting each other in many ways. From the logistical and technical—like the KHS and WU project—to the development of Web archives—such as shared collection development—partners have learned some of the benefits of collaboration. Other positive lessons learned by working together include recognizing there is strength in numbers to negotiate cost savings, having access to other professional expertise, sharing best practices and guidelines with fellow members, and strengthening professional relationships in the state.

Emporia State University (ESU) had a unique benefit from participating in KAIC by providing a learning opportunity for graduate students in ESU's School of Library and Information Management. As the only library school in the state, ESU was able to leverage its consorcial membership to expose future archives professionals with practical experience in Web archiving. They incorporated Archive-It into class projects and curricula, which provided enhanced interaction with an innovative resource they likely will use in the field.²⁸

In addition to the benefits of collaboration, KAIC members recognize there also are challenges. First, it takes time to build trusting relationships, which is key to successfully working together. Existing responsibilities stretch each of the partners, and adding Web archiving in a shared environment requires members to decide how to fit it into their priorities. At times progress comes more slowly as participants juggle all of their respective institutional duties and assignments beyond KAIC tasks.

Another challenge of collaboration is pursuing shared goals among partners with sometimes significant variations in resources, skillset levels, and technological capacities. Members have differing budgets, so each must prove to allocators they are showing fiduciary skill in using public funds. Similarly, KAIC partners have varying proficiency levels related to Web archiving, which can hinder group progress if not addressed with care and tact. Plus, each participant works within a separate

28. For a list of educational partnerships with Archive-It, see <https://webarchive.jira.com/wiki/display/ARIH/Archive-It+Educational+Partnerships> (account required). For more information about Archive-It Educational Partnerships, see <https://archive-it.org/blog/learn-more/educational-partnerships/>. Examples of Emporia State University (ESU) educational usage are at <https://archive-it.org/organizations/857> and <https://archive-it.org/organizations/858>.

technological environment that affects the capacity to preserve Web content. Amidst these disparities, members have to navigate through the challenges toward the shared goals.

A more tangible lesson from participating in KAIC was the need to improve documentation regarding group decision making. Members kept minutes of meetings, however there were occasions when partners could not find when particular decisions were made. This challenge in part is rooted in the group's early decision to keep KAIC informal. As it matures, though, members will need to reassess how formal it should be.

Finally, members realized in KAIC's third year that they should have clarified in the beginning the number of Archive-It users per institution. This simple step would have helped institutions to deploy Web archiving resources more effectively from the start. The lesson learned was to more thoroughly consider facets of user agreements when multiple partners are involved.

Future Directions and Possibilities

In addition to the lessons learned so far, KAIC partners are considering future directions as a Web archiving consortium. While group membership currently numbers six institutions, it is not closed. Participants decided to keep involvement limited while building a sustainable organization and developing a relationship with Archive-It, after which expansion among other Kansas organizations would be possible. To give a sense of KAIC's growth potential beyond the current members, there are 28 public two- and four-year colleges and universities, 18 private colleges and universities, and numerous historical societies, museums, and historic sites in Kansas. The state's rich educational and cultural heritage is represented by these diverse institutions. Pursuing an expanded membership would help KAIC further develop a rich, well-rounded digital resource with statewide representation. Furthermore, existing consortial members could use their software experience and policy and procedure developments to mentor new members. These guidance efforts could also become case studies to share with Archive-It personnel and partners, as well as the broader archival profession.

One of the early influences on creating KAIC was the potential to create a large body of archived Web resources that users could search in a single location. As the group expands its collaborative collecting efforts and investigates increasing its membership, this consortial search tool will become a significant asset. Although Archive-It currently offers an option to search across all collections created by Archive-It partners, as well as some search scoping options, at this time it cannot limit a search to only consortial members. This capability could be a very useful tool for researchers with specific interests, such as the history of Kansas or Kansans. With a consortial search box, users could easily search within KAIC-generated Web archives and find relevant results. Being united by such a search tool would strengthen the consortium and provide an obvious place to direct researchers as they

navigate their collections. The addition of such a feature might also benefit Archive-It's recruiting efforts as institutions or consortia would see how easily their collections may be unified and navigated by researchers. Elements also could be deployed in the current search tool to enhance users' abilities to search effectively. Archive-It has been receptive to this idea during early discussions with KAIC, which should help subscribers and researchers have cautious optimism that such a tool will become a reality.

Conclusion

Web archiving is an emerging frontier that still holds many questions and challenges for archivists. Tapping into the culture of collaboration that exists among Kansas libraries and archives, KAIC began forming roughly three years ago to help each other navigate this new territory. As members began collaboratively archiving Web content, they discovered multiple benefits including shared documentation for metadata entry, consortial acquisition efforts based on documentation strategy concepts, and cost savings that have helped to ease strained budgets. KAIC currently is one of the larger consortia in Archive-It's portfolio, and has plans to expand membership to other Kansas institutions. Broadening the consortium not only adds new voices to the conversation, but it also enhances the resources available to a primary user group, Kansas residents. With many of the larger repositories in the state participating in KAIC, Kansans can more conveniently locate relevant Web archives than if members were capturing content independently. As KAIC maintains its current trajectory, it will prove the fiscal and collaborative value of consortial arrangements and demonstrate to researchers the benefits of aggregating Web archives.

Appendix A: Sample Metadata Guidelines

Kansas Historical Society

Archive-It Metadata Procedures (Draft)

Seed level metadata guidelines

Elements in **bold** are required.

- Title:
 - Name of agency plus “website”.
 - Example: Kansas State Board of Pharmacy website
- Creator:
 - Name of agency using LC authority heading, if available.
 - Example: Kansas. State Board of Pharmacy
- **Description:**
 - Include record group scope note from KSHS Archives Catalog, if available. (For Collections-Related records, use notes in following order of precedence: 5203: Abstract. If empty: 5202: Scope and Content. If empty: 5208: Summary. See email from Bob Knecht 4/16/2015.)
 - If not, compose one based on information available on the agency website.
 - This is the only field required by Archive-It.
- Subject:
 - Do we use LC subject headings or a more free-form faceted approach? NC is free form.
 - At this point, it seems advisable to use LC subject headings from KSHS Archives Catalog record group record, if available. If not, consult Atlas for suggested LC subject headings for the agency in question.
- Publisher:
 - Should we repeat the creator in this field? It seems redundant but other Archive-It partners appear to follow this practice with some frequency.

- Contributor:
 - Include a parent agency, if appropriate.
 - Example: Governor's Office is a Contributor for the Kansas African-American Affairs Commission because the KAAAC is administratively placed in the Governor's Office.
- Type:
 - It seems more user friendly to use terms other than DCMI's list such as:
 - website.
 - social media (perhaps even as granular as social media—Facebook page or social media—Twitter feed).
 - blog.
- Date:
 - Should we include this at the seed level? The date of capture should be displayed and be fairly obvious.
- Format:
 - Should we include the MIMETYPE?
- Identifier:
 - Unit ID.
- Source:
 - Consider using for subdomains—add the URL for the parent domain in the seed metadata for the subdomain(s) and vice versa?
- Relation:
 - Perhaps this would be a more appropriate field for subdomain information?
 - Should we include a URL to the Archives Catalog record group record in this field?
- Coverage:
 - Do we need temporal or spatial coverage? Seems unnecessary.
- Rights:

- Do we need to develop a rights statement? I don't believe we can assume all Kansas state agency Web content is in the public domain.
- Collector:
 - Use "Kansas State Historical Society. State Archives Division".
- Language:
 - Use user-friendly version: English rather than eng.
- Group:
 - Do we want to use this option? If so, how?

Kansas State University

Archive-It Metadata Guidelines (Feb 2016)

For all collections in our Web archive, we recommend metadata for the following fields:

- Collection level
 - Title
 - Archived Since
 - Description
 - Subject
 - Creator or Publisher
 - Collector
- Seed/Document level (A seed is any URL that is being captured by a crawler, and a document is any file on the Web that has a distinct URL. Some examples of a seed could be an entire website, a specific part of a website, or a specific URL (this is also called a document). A document could be an image, PDF, article, etc.)
 - Title
 - URL
 - Creator or Publisher
 - Coverage

- Date

For descriptions of these elements, please see <http://www.dublincore.org/documents/usageguide/elements.shtml>.

University of Kansas

Archive-It Metadata Recommendations (7/16/14; revised 3/2015)

Overview

Archive-It is currently being employed as an online archive of KU Web content. While there are many potentially valuable sources of content within the website, a particular emphasis has been placed on content pertaining to the Chancellor, Provost, and Commencement as prime, top level, resources for initial development and exposure.

Since there is a large, and growing, amount of available Web content, initial focus will be given to creating collection level metadata for the top-level categories. Individual records for pages of note (e.g. Chancellor Reports) can be created and grouped under each collection as warranted.

Metadata Schema (Collection Level)

As per OAI standards, simple Dublin Core is the required minimum element set to describe resources in Archive-It. Its limitations however can be mediated by the addition of other elements. From the general DCMI Terms, *accrualMethod*, *accrualPeriodicity*, *accrualPolicy*, *relation.isPartOf*, *extent*, and *coverage.spatial* have been included. These additional elements will give additional context to the digital resource and its method of ingest, and will be useful from a preservation standpoint.

Content Standard (Collection Level)

Given the difference between Web and physical archives, it is difficult to effectively map existing content standards to an element set designed for a very different medium. Some consideration has been given to match appropriate DACS guidelines to DC elements, but GCMi guidelines were followed more closely.

Standard Archive-It Metadata Fields Used (indicates required)*

Element	Guideline	Example
Title*	Major Heading of Top Level Page	Office of the Chancellor
Subject*	Institution, Named Entities (use VIAF, LCSH naming conventions), Separate Subjects with semicolon	Gray-Little, Bernadette, 1944-
Description*	Basic outline of resources within that part of the website	The webpages pertaining to the Office of the Chancellor at the University of Kansas.
Publisher*	Institution Responsible for Publishing of Site	University of Kansas
Rights	Institution holding rights to resource, and any licenses granted with it	Copyright by University of Kansas
Collector	Institution that collected the resources	University of Kansas
Language	Primary language of resources	English (US)

Additional Metadata Fields Used by KU (indicates required)*

Element	Guideline	Example
Coverage.Spatial	Use NAF. If not specified, use City, State	Lawrence, KS
accrualMethod*	Method by which resources were aggregated	Web Crawl
accrualPeriodicity*	Frequency with which items are ingested	Monthly
accrualPolicy*	Policy governing addition of items to collection	All Web content available at time of crawl.