



Midwest Archives Conference 2017 Annual Meeting

Collaborative Web Archiving: Lessons from Kansas

Sherry Williams, chair



Megan Macken



Cliff Hight, substitute
chair



Patty Nicholas



Introduction

- First web pages
 - Thanks, Tim Berners-Lee!
- Responsibility to preserve & make accessible
- Our plan
 - Overview
 - With limited resources
 - With more resources

```
The World Wide Web project

WORLD WIDE WEB

The WorldWideWeb (W3) is a wide-area hypermedia[1] information retrieval
initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this
document, including an executive summary[2] of the project, Mailing lists[3] ,
Policy[4] , November's W3 news[5] , Frequently Asked Questions[6] .

What's out there?[7]Pointers to the world's online information,
subjects[8] , W3 servers[9], etc.

Help[10] on the browser you are using

Software Products[11] A list of W3 project components and their current
state. (e.g. Line Model[12] ,X11 Viola[13] ,
NeXTStep[14] , Servers[15] , Tools[16] , Mail
robot[17] , Library[18] )

Technical[19] Details of protocols, formats, program internals
etc

<ref.number>, Back, <RETURN> for more, or Help: | █
```

Emulated version of the first widely accessible web page, 1992
From <http://line-mode.cern.ch/www/hypertext/WWW/TheProject.html>

Introduction

- For additional info on KAIC, see:
 - “Collaboration Made it Happen! The Kansas Archive-It Consortium,” *Journal of Western Archives*, at:
<http://digitalcommons.usu.edu/westernarchives/vol8/iss2/4>

Current Issue: Volume 8, Issue 2 (2017) Web Archiving Special Issue

Articles



Introduction to the Special Issue on Web Archiving
Nicholas Taylor



Developing Web Archiving Metadata Best Practices to Meet User Needs
Jackie M. Dooley, Karen Stoll Farrell, Tammi Kim, and Jessica Venlet

Case Studies



Case Study: Washington and Lee's First Year Using Archive-It
Alston B. Cobourn

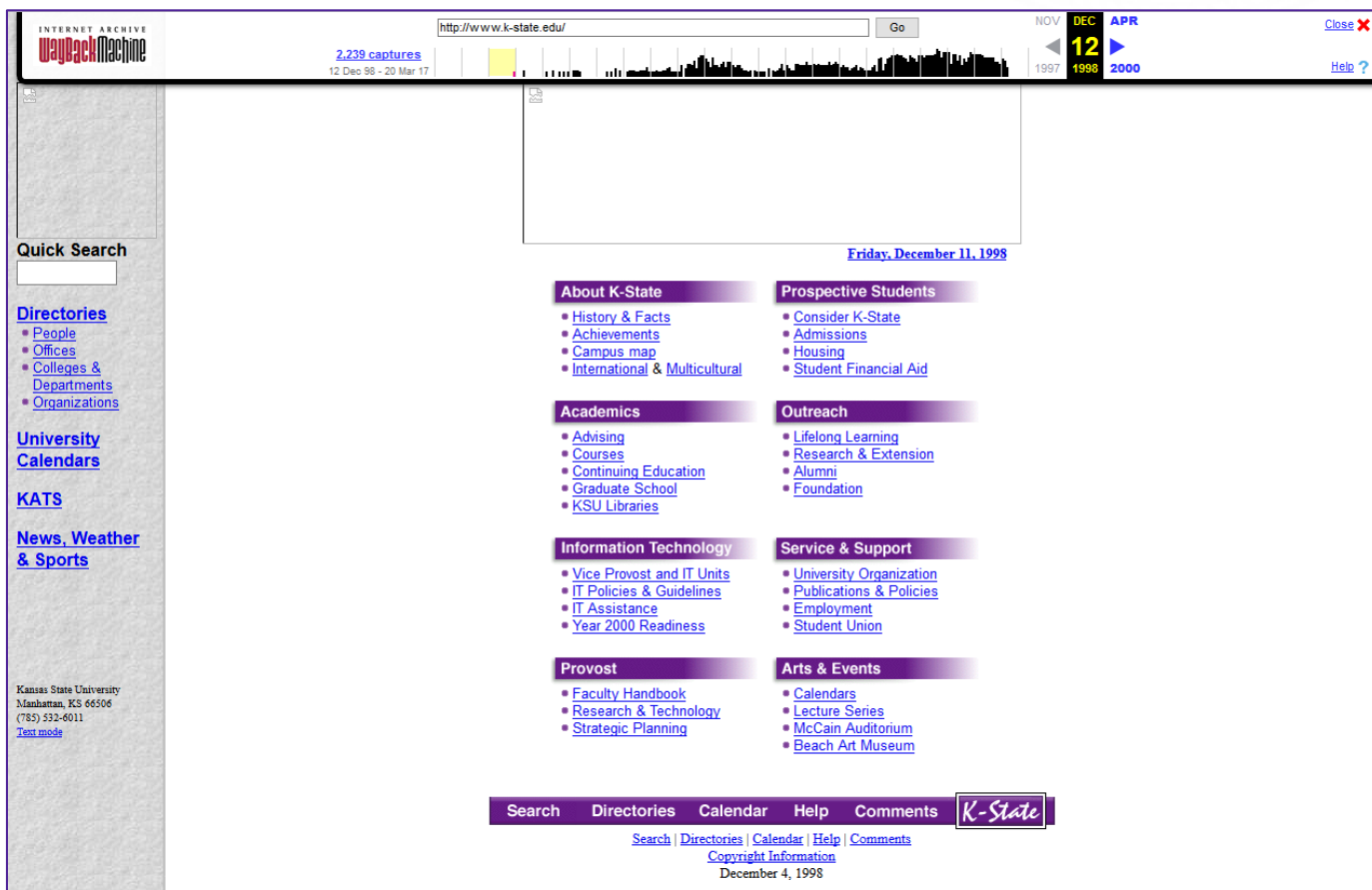


Using RSS to Improve Web Harvest Results for News Web Sites
Gina M. Jones and Michael Neubert



Collaboration Made It Happen! The Kansas Archive-It Consortium
Cliff Hight, Ashley Todd-Diaz, Rebecca Schulte, and Michael Church

Starting the Collaborative



First K-State home page in Wayback Machine, 12/12/1998

Cliff Hight

University Archivist | Morse Department of Special Collections

In the Beginning...

- Found common challenge
- Discussions with others in Kansas
 - 2011–2013: chats with state archivist & others
 - 2013: request of state archives director



Photo: close up of <http://www.flickr.com/photos/empeiria/8657432375/>

In the Beginning...

- Naming the consortium
 - Kansas Web Archiving Collaborative (KWAC)



Photo: <http://www.flickr.com/photos/9422878@N08/>

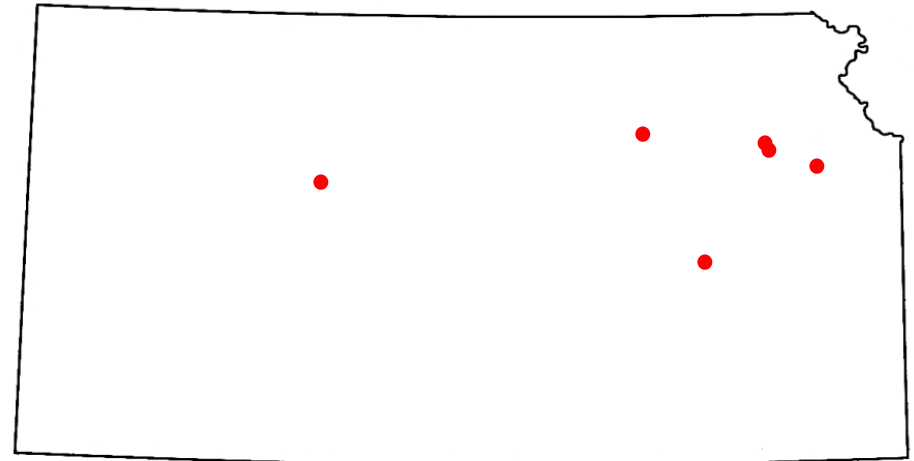
- Kansas Archive-It Consortium (KAIC)



Photo: <http://www.flickr.com/photos/marcyleigh/>

In the Beginning...

- Current member institutions
 - Emporia State University
 - Fort Hays State University
 - Kansas Historical Society
 - Kansas State University
 - University of Kansas
 - Washburn University



Map: http://www.netstate.com/states/maps/images/ks_outline.gif

Organizational Structure

- Administrative
 - Flexible approach
 - Project coordinator
 - Archive-It
- Maintaining relationships
 - Build trust
 - Consistent communication
 - Use technology wisely



Two K-State students using computers in their dorm room, undated

Financial Collaboration

- Creating messages for resource allocators
- Strength in numbers
- Independent flexibility



K-State students purchasing season tickets for football, 1958

Lessons for a Web Collaborative

- Recognize variations among partners
- Use technology
 - A/V conferencing
 - Shared online space
 - Shared seed list
 - Partner/shared metadata guidelines



Computers in the K-State library for database searching, 1988

Benefits and Difficulties

- Benefits
 - Others who can help
 - Stronger relationships within group
 - Improved collection development
 - And others
- Difficulties
 - Requires time to coordinate
 - Differences in resources
 - Keeping up with other duties
 - And others



Sheep shearing at K-State, 1911

Summary

1. Communicate, communicate, communicate
2. Plan, plan, plan
3. Formal documentation
4. Unique partner opportunities
5. Clarify number of users
6. Collaborative collecting is possible



New IBM 650 installed at K-State, 1958

Collaborative Collecting

- Based on concept of “documentation strategy”
 - First from Helen Samuels & others
 - Summary:
 - Many repositories
 - Similar topic
 - Defined collecting scope
 - Formalized institutional involvement
 - Appraisal criteria
 - Acquisition
- Connections to web archiving



Student reads newspaper at K-State, 1969

Kansas Archive-It Consortium (KAIC) at Fort Hays State University

Patty Nicholas
Library Specialist,
Special Collections and Periodicals
Forsyth Library



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

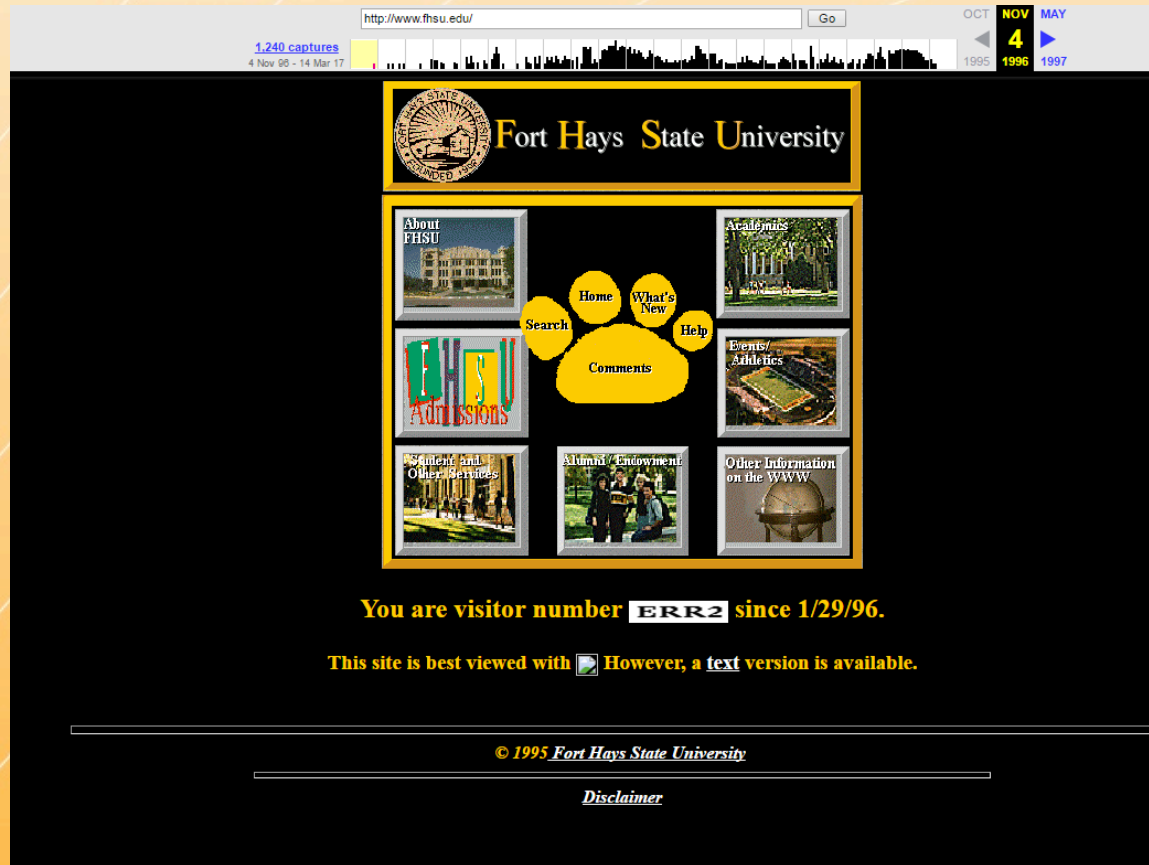
Fort Hays State University Hays, Kansas

- Founded in 1902
- Only 4 year institution of higher learning in the western part of the state
- Spring 2017 enrollment
 - 4257 – Campus
 - 6652 – Virtual College
 - 1744 – International partner institutions



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.



FHSU's first entry on the Wayback Machine

November 4, 1996



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

Consortium Beginnings

- I was the University Archivist at the time in late 2013 when the meetings and conference calls with other consortium members began
- Visited with the man who was our library director at that time about the consortium, and he liked the idea
- Quotes were received by consortium members in May 2014
- After we received the quotes and data budgets, I got the okay to proceed with the membership to Archive-It from our library director



FORT HAYS STATE
UNIVERSITY

Forward thinking. World ready.

1st Challenge

- I had to deal with concerns from some of my colleagues
 - Thought I should have gone to the University Web Site committee for input
 - Not sure if our director had the authority to offer university content to an external entity
 - Whether the university or the library should pay the annual costs
 - Are there other pages within the university's site that should be crawled



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

How we proceeded with joining the consortium

- University Web Site Committee
 - I gave them information regarding Archive-It and the consortium
 - The web content manager said she was not sure it really even needed to be approved by the committee, so we went ahead with the process
- Our IT person went to the University's Computing Center
 - Our university has to go through an office there for computer, tablets and software purchases
 - It was decided this would be a library purchase, not a university purchase, on an annual basis



FORT HAYS STATE
UNIVERSITY

Forward thinking. World ready.

2014-2015

- First MOU for the consortium was available to sign in August 2014
 - FHSU's portion
 - Up to 3 million URLs archived
 - Not to exceed 0.125 terabyte(s) in data
- Became a member on November 12, 2014
 - Decided on 4 sites to crawl
 - 1 daily
 - 3 weekly
- By the end of 2015, we had added two more sites



FORT HAYS STATE
UNIVERSITY

Forward thinking. World ready.

Big Problem

- After our new dean of libraries came in, she made some changes. In November 2015, a colleague, Sherry Severson, was moved into the University Archives and she took over the Archive-It project for Forsyth Library
 - I was asked to return to work in the Periodicals area and also remain in Special Collections
- By the end of January 2016, we realized that we were going over our data budget
 - Remember the daily crawl I mentioned on the last slide?
 - Tiger Media Network – the online news site for the university
 - When I made the decision to crawl daily, I did not realize how much data it would take up
- Sherry and I talked with a representative from Archive-It to get some ideas on how often to crawl big sites
 - It is currently being crawled monthly
- We were invoiced for the data that was over our data budget



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

Numbers

Current Subscription Details: July 2016- June 2017 as of April 5, 2017

Data Budget	189 GB
New Data	53 GB
New Documents	2,067,383

	2015	2016	All Time
Total Data	80.4 GB	269 GB	402.4 GB
Total Documents	2,413,965	8,045,892	12,527,240

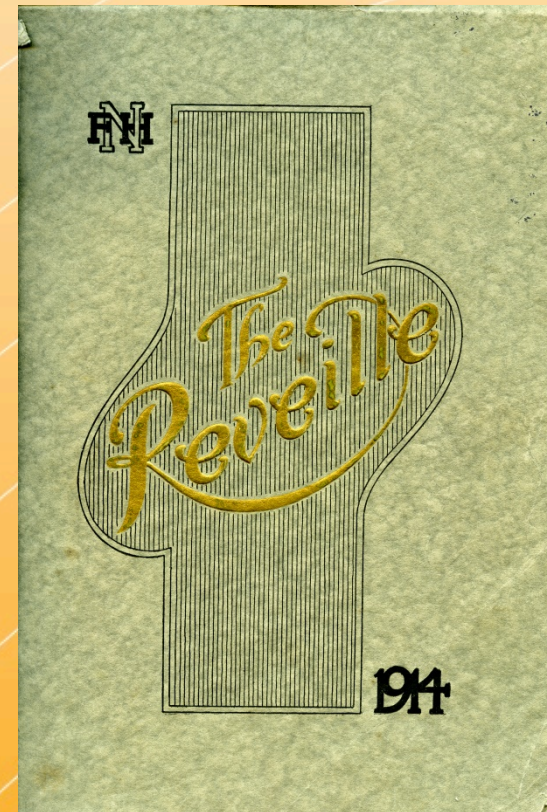


**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

Collections

- We currently have 17 active collections on a scheduled crawl
 - 4 monthly
 - 3 quarterly
 - 6 semi-annual
 - 4 annual
- Two other active collections were 1 time crawls
- 20th active collection is currently not scheduled, but we are keeping our eye on it
 - It is a new publication put out by the Alumni Association and we don't know how often it will be published



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

Collections

- Our collections include:
 - 5 colleges
 - 2 top administrative offices
 - FHSU Athletics
 - They use a commercial web service for their web site
 - Tiger Media Network
 - University Relations
 - Forsyth Library
 - Plymouth Schoolhouse (Omeka site)



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

Issues

- Hard to determine how often we should crawl various sites because of unknown future data numbers
 - Current Data budget – 189 GB
 - Sherry decided she wanted to be more conservative in how often to crawl various sites
- We are the furthest away from the other members of the consortium
 - Getting approval to attend non-teleconference meetings can be denied due to limited travel budget
- With a limited budget, how do we determine which sites are to be crawled? The following is what I went by:
 - Crucial information that is contained in the site
 - The university catalog of courses is no longer printed at FHSU
 - News of the university on a daily or weekly basis
 - Tiger Media Network
 - University Relations News
 - Potential and current student information
 - The colleges within the university
 - Popular
 - Alumni Association
 - Athletics



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.

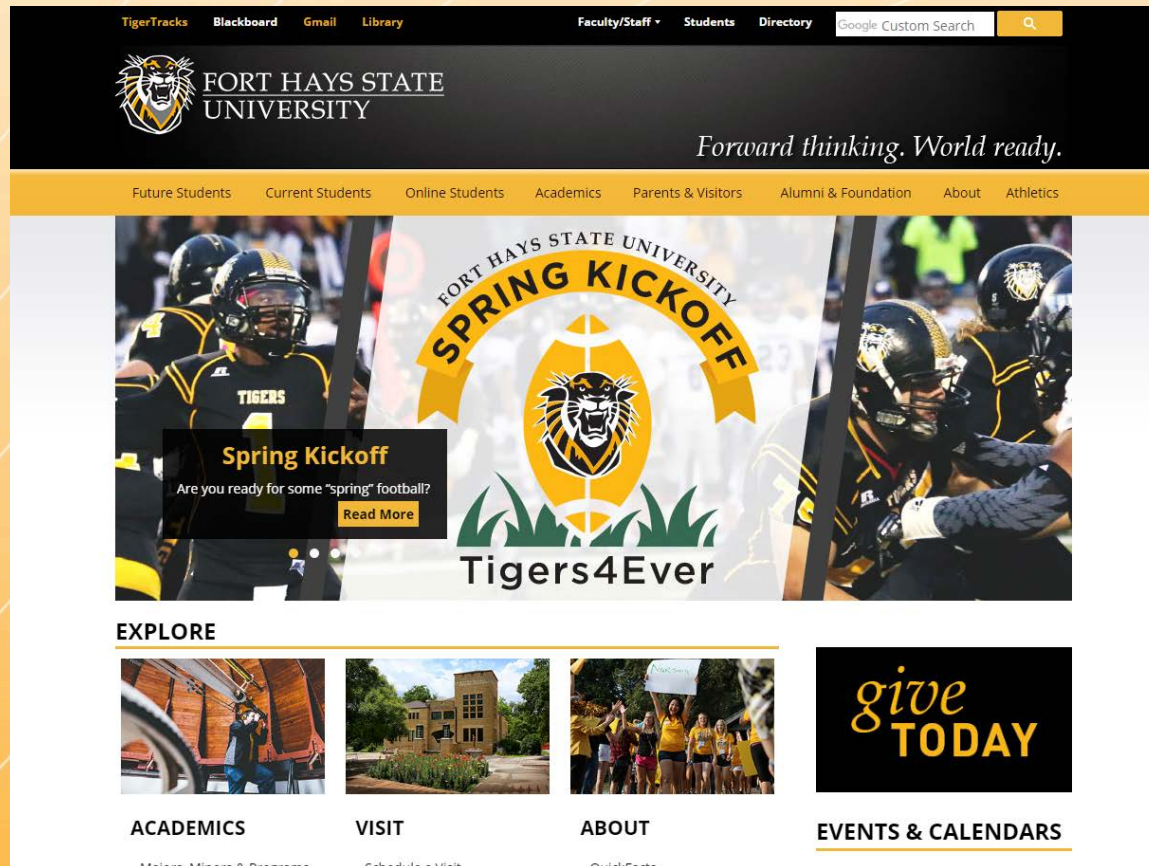
Advantages of being a part of KAIC

- Working together with various institutions from across the state to achieve a common goal
- Sharing of staff expertise and best practices
- The joint purchasing agreement helps to reduce the costs of archiving our web pages
- Advocacy from other institutions can help with getting your own institution on board



FORT HAYS STATE
UNIVERSITY

Forward thinking. World ready.



FHSU Web Page Today

www.fhsu.edu



**FORT HAYS STATE
UNIVERSITY**

Forward thinking. World ready.



Millennium time capsule, 1999. Courtesy KansasMemory.org.

Archiving Websites

at the Kansas Historical Society

Megan Macken, Digital Archivist

YOUR
STORIES
OUR
HISTORY | KANSAS
HISTORICAL
SOCIETY

Basic Vocabulary

- Seeds
- Crawling
- Scoping
- QA
- Crawler Trap
- Robots.txt
- Wayback Machine



Rotate-o-Matic Super Astronaut, Horikawa Company, 1960s.
Courtesy KansasMemory.org.

<https://support.archive-it.org/hc/en-us/articles/208111686-Glossary-of-Archive-It-and-Web-Archiving-Terms>

Collaboration?



Galle Family, Moundridge, Kansas, 1998. Courtesy KansasMemory.org.



KAIC Comparison

Institution	FTE	TB	Seeds	Local
Fort Hays State University	0.05	0.125	21	100%
Emporia State University	0.05	0.125	53	100%
Washburn University	0.01	0.25	21	100%
Kansas State University	0.07	0.5	33	27%
University of Kansas	0.2	0.5	618*	38%
Kansas Historical Society	0.15-0.8	0.75	395	0.001%

*481 of the 618 seeds are one-time, single-page crawls



Ella Bird Lott's 80th Birthday, 1941.
Courtesy KansasMemory.org.

KSHS Staffing

Title	Tasks	Web Collection	Status	Hrs/Mo
Digital Archivist	All			24
Digital Initiatives Coordinator	Contract; Selection; QA; Preservation			1-2
Electronic Records Archivist	QA; Preservation	State Government Agencies	Turnover	1-2
Public Records Archivist	QA	State Government Agencies	Turnover	-
Head of Acquisitions & Collections	Selection; QA	Collections of KSHS; Community, Hist/Genealogical + Political Orgs	Retiring, replaced?	1-2
Director of State Archives			Retired	-
Asst. Director, State Archives	Selection; QA	State Government Agencies	Promoted/ not replaced	-
Archivist/Pres. Coordinator	Selection; QA	Collections of KSHS; Community, Hist/Genealogical + Political Orgs		1-2

Web Archiving Steps

1. *Selection*
2. Running Test Crawls
3. Scoping (pre-QA)
4. Crawling
5. *Quality Assurance*
6. Patching
7. Going Public
8. *Preservation*



Pen and ink drawing, Myron A. Waterman, 1893. Courtesy KansasMemory.org.

KSHS Staffing

Title	Tasks	Web Collection	Status	Hrs/Mo
Digital Archivist	All			24
Digital Initiatives Coordinator	Contract; Selection; QA; Preservation			1-2
Electronic Records Archivist	QA; Preservation	State Government Agencies	Turnover	1-2
Public Records Archivist	QA	State Government Agencies	Turnover	-
Head of Acquisitions & Collections	Selection; QA	Collections of KSHS; Community, Hist/Genealogical + Political Orgs	Retiring	1-2
Director of State Archives	Selection; QA		Retired	-
Asst. Director, State Archives	Selection; QA	State Government Agencies	Promoted/ not replaced	-
Archivist/Pres. Coordinator	Selection; QA	Collections of KSHS; Community, Hist/Genealogical + Political Orgs		1-2

Metadata

Settings

Metadata

Crawling History

Notes

Seed Scope

	url	Publisher	Language	Descriptio	Agency co	Title	Unit ID	Creator	Contribut	Relation	Collector	Subject	Subject	Subject	Subject	Subject	Subject	Subject
Tit	http://admin.ks.gov/																	
	http://admin.ks.gov/offices/chief-counsel/																	
Cn	http	Administrative procedure -- Kansas (1)																
	http	Affordable housing (1)																
	http	Civil rights -- Kansas (1)																
	http	Discrimination -- Kansas (1)																
	http	More ▾																
Su	http																	
De	http																	
	http	Creator						Sort By: Count (A-Z)										
Pu	http																	
	http	Kansas Department for Children and Families (1)																
Lang		Kansas Health Care Stabilization Fund (1)																
Co		Kansas Housing Resources Corporation (1)																
		Kansas Human Rights Commission (1)																
Da	Age	Kansas Legislature Legislative Administrative Services (1)																
Ty		More ▾																
Fo	Unit																	
Idi		Publisher						Sort By: Count (A-Z)										
So	Add	Information Network of Kansas (1)																
		Kansas Housing Resources Corporation (1)																
		Kansas Human Rights Commission (1)																
		Kansas Water Office (1)																

Sites

Search Page Text

Page 1 of 1 (12 Total Results)

Sort By: Title (A-Z) | Title (Z-A) | URL (A-Z) | URL (Z-A)

URL: <http://admin.ks.gov/offices/printing-mailing/>
Captured 3 times between Nov 8, 2015 and Nov 8, 2015

Title: Kansas Health Care Stabilization Fund website
URL: <http://hcsf.kansas.gov/>
Captured once on Nov 8, 2015
Subject: Kansas. Health Care Stabilization Fund -- History, Kansas. Health Care Stabilization Records and correspondence, Malpractice insurance -- Kansas, Medical personnel -- Malpra
Kansas, Kansas. Insurance Dept. -- History
Creator: Kansas Health Care Stabilization Fund
Publisher: Kansas. Health Care Stabilization Fund
Language: English
Unit ID: 216184
Collector: Kansas State Historical Society. State Archives Division
Relation: <http://www.kshs.org/dart/units/view/216184>
Agency code: 270

Meta-Collaborations



Prize Cakes Culinary Department, Kansas Free Fair Album, 1921. Courtesy KansasMemory.org.

Internal Documentation

[Home](#) [My page](#) [Projects](#) [Help](#)

Logged in as **mmacken** [My account](#) [Sign out](#)

Web Archiving

Search: Web Archiving ▼

[Overview](#) [Activity](#) [Roadmap](#) [Issues](#) [New issue](#) [Gantt](#) [Calendar](#) **[Wiki](#)** [Files](#) [Settings](#)

Archive-It

[✎ Edit](#) [★ Watch](#) [🔒 Lock](#) [🔄 Rename](#) [🗑 Delete](#) [⏪ History](#)

[📁 Public Portal](#)
[📁 Login](#)
[📁 View Seeds in Wayback Machine](#)

- [To Do List and Outstanding Questions](#)
- [Kansas Archive-It Consortium](#)
- [Metadata](#)
 - [Archive-It Metadata Procedures](#)
 - [Metadata Import/Export](#)
- [QA](#)
 - [Webcrawl Scoping](#)
- [Selection Criteria](#)
 - [Government Records](#)
 - [Collections](#)
- [Support](#)
- [User Accounts](#)

Additional KSHS documentation is located at
I:\Outreach\Partnerships\web_archiving_collaborative

- [External resources](#)

[New file](#)

Wiki

- [Start page](#)
- [Index by title](#)
- [Index by date](#)

KAIC Comparison

Institution	FTE	TB	Seeds	Local
Fort Hays State University	0.05	0.125	21	100%
Emporia State University	0.05	0.125	53	100%
Washburn University	0.01	0.25	21	100%
Kansas State University	0.07	0.5	33	27%
University of Kansas	0.2	0.5	618*	38%
Kansas Historical Society	0.15-0.8	0.75	395	0.001%

*481 of the 618 seeds are one-time, single-page crawls

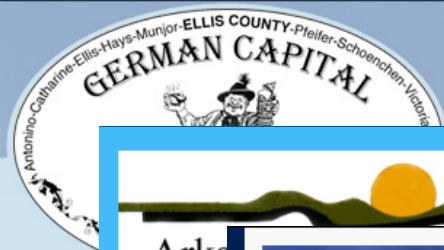


Ella Bird Lott's 80th Birthday, 1941.
Courtesy KansasMemory.org.

KSHS Web Collections

Collection	Frequency	Seeds
Community Organizations	Semi-annual	82
	Weekly	1
Collections of KSHS	Semi-annual	99
	Annual	1
Government Agencies	Annual	86
Political Organizations	Monthly	29
	Semi-annual	1
	One-time	1
Historical/Genealogical Orgs	Annual	83
	One-time	1

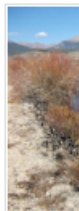




Home Page
Contact us
2016 German
Events
Ellis County
Villages
Antonino,
Catharine,
Ellis, Kansas
Munjor, K
Pfeifer, K
Schoenchen
Victoria, K
Vincent, K
Walker, K
News/Press
Photo Gall
German Ca
German At
Historic Ch
German Fo
Sponsors
Groups/Fa
Maps
Shop/Buy
City of Hay
Links

Ark
Riv
C

Home



Welco
WELCOME to the
professions: A
professions.

ANNOUNCE

FALL 2016

The revis
We also u
The **KSB**
available
newslette

NOTICE

The dead
applicati
now bein
custome

Neewollah 2016 Starts In: 0 days, 0 hours, 0 mins, 0 secs

Home News Sponsors

**Neewollah**
INDEPENDENCE, KANSAS
SINCE 1919

Home Events Ask the General Headquarters Photos

1-877-Neewollah (877-633-9644)

NEEWOLLAH 2016, Oct 21st - Oct 29th

CALENDAR OF
EVENTS 

[Contact Us](#)

[Neewollah on Facebook](#)

Neelah Deb/Stage Escort Application

Logo Contest Information

Queen Neelah LXXIV

Anna Miller



Collaboration.



LibGuides@WU

[Washburn University](#) / [LibGuides](#) / [Washburn University Archives and Special Collections](#) / Archives

Washburn University Archives and Special Collections: Archives

Search All Guides

Search

Washburn University Archives and Special Collections

[Getting Started](#)

[Archives](#)

[Collections](#)

[Articles of Association](#)

[Washburn Coat-of-Arms](#)

[1966 Tornado](#)

[Washburn 150](#)

[Institutional Repository](#)

Access

The Washburn University Archives and Special Collections offer access Monday through Friday by appointment.

Please note all of the materials in the Archives and Special Collections are non-circulating.

Appointments

[washburn.edu/mabee/](#) make an

Archives Index

To search the Archives Index click below.

- [Archives Index](#)

Washburn University Presidents History

- [WU History and Presidents Slideshow](#)

Student Newspapers

[Washburn Press Box 1885 - Box 47 1888](#)

Subject Guide



Martha Imparato

YOUR
STORIES
OUR
HISTORY | KANSAS
HISTORICAL
SOCIETY

Galle Family, Moundridge, Kansas, 1998. Courtesy KansasMemory.org.

KAIC Documentation

My Drive - Google Drive

Kansas Archive-It Consortium

Master Seed List - Kansas Archive-It Consortium

Secure | https://docs.google.com/spreadsheets/d/1uz1w6DVJHiNii7YVGwIffGt1-cRH_d6yaQo_vAGXVKk/edit#gid=105346054

Master Seed List - Kansas Archive-It Consortium

File Edit View Insert Format Data Tools Add-ons Help Last edit was made on February 21 by Letha Johnson

Collection Title

Collection Title	Seed Title	URL	Notes
Kansas State University Web Archive	Kansas State University	www.k-state.edu/	Still figuring out frequency
	Kansas State University	www.ksu.edu/	Still figuring out frequency
	Landon Lecture Series on Public Issues	ome.ksu.edu/lectures/landon/index.html	One-time before change
	K-State Research and Extension	www.ksre.ksu.edu/	Quarterly
	K-State Research and Extension - About Us	www.ksre.ksu.edu/p.aspx?tabid=39	One-time before change
	Kansas 4-H	www.kansas4-h.org/	Semiannual
	Twelfth International Phytotechnologies Conference	conferences.k-state.edu/phytotech2015/	One-time
	Kansas State University President Instagram	https://www.instagram.com/kstate prez/	Still testing
	Kansas State University President Twitter	https://twitter.com/kstate_pres/	Still testing
	Kansas State University First Lady Twitter	https://twitter.com/kstate_1stlady/	Still testing
	Kansas State University First Dog Twitter	https://twitter.com/search?q=%40kstate_1stdog&src=typ	Still testing
	Day of Ecuador at Kansas State University	http://global.k-state.edu/dayofecuador/	Annual
	Kansas State University Libraries	http://lib.k-state.edu/	Quarterly
Kansas State University Affiliated Organizations Web Archive	Flint Hills History Conference	flinthillshistconf.wordpress.com/	One-time
	K-State Athletics	www.kstatesports.com	Quarterly
	Collegian Media Group	https://www.collegianmedia.com/	Annual
	We Are K-State	http://www.wearekstate.com/	One-time
	The Fourum	http://www.thefourum.net/	One-time
	Royal Purple Yearbook	http://www.royalpurpleyearbook.com/	Annual
	Kansas State Collegian Newspaper	http://www.kstatecollegian.com/	One-time & then Weekly
	UFM	http://tryufm.org/	Quarterly
Kansas State University Student Organizations Web Archive	Black Student Union at Kansas State University	https://ksusankofa.wordpress.com/	Semi-annual
	Black Student Union Twitter	http://twitter.com/ksusbu/	Bimonthly
	Black Student Union Twitter #InherentTheRevolution	https://twitter.com/hashtag/InherentTheRevolution/	One-time
	Black Student Union Twitter #MyBlackIsBeautifulWeek	https://twitter.com/hashtag/MyBlackIsBeautifulWeek/	One-time
	Black Student Union Twitter #TheBlackExperience	https://twitter.com/hashtag/TheBlackExperience/	One-time
	Black Student Union Instagram	https://www.instagram.com/ksusbu/	Bimonthly
	Black Student Union Facebook	https://www.facebook.com/groups/101433146571409/	Still testing

KAIC Portal

Kansas Archive-It Consortium

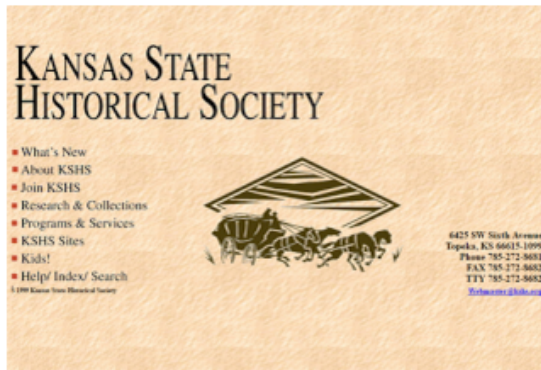
[Introduction](#)

[Search Web Archives](#)

[Publications](#)

[Member login](#)

Search Web Archives



Kansas Historical Society website, April 20, 1999. WayBack Machine.



Kansas Archive-It Consortium, <http://sites.google.com/sites/kansaswebarchives>

Search this site

Kansas Historical Society and State Archives

Search Web Archives

Go

Browse Web archives

<https://archive-it.org/organizations/606>

Collecting Areas

State Government Agencies of Kansas, Kansas Political Organizations, Kansas Community Organizations, Websites pertaining to collections of the Kansas Historical Society, Historical and Genealogical Societies of Kansas, Kansas Digital Newspapers

Kansas State University

Search Web archives

Go

Browse Web archives

<https://archive-it.org/organizations/890>

Collecting Areas

Consumer movement, Cookery, Kansas Life and Culture, Kansas State University History

Lessons learned...



Millennium paperweight, 1999. Courtesy KansasMemory.org.

Contact & Evaluation

Kansas Archive-It Consortium: <http://sites.google.com/site/kansaswebarchives>

Megan Macken
Digital Archivist
Kansas Historical Society
mmacken@kshs.org
785-272-8681, ext. 280

Annual meeting and session
evaluation form:
bit.ly/OMAMAC2017



Solomon grain elevator, Solomon, Kansas, 1998. Courtesy KansasMemory.org.

YOUR
STORIES
OUR
HISTORY | KANSAS
HISTORICAL
SOCIETY