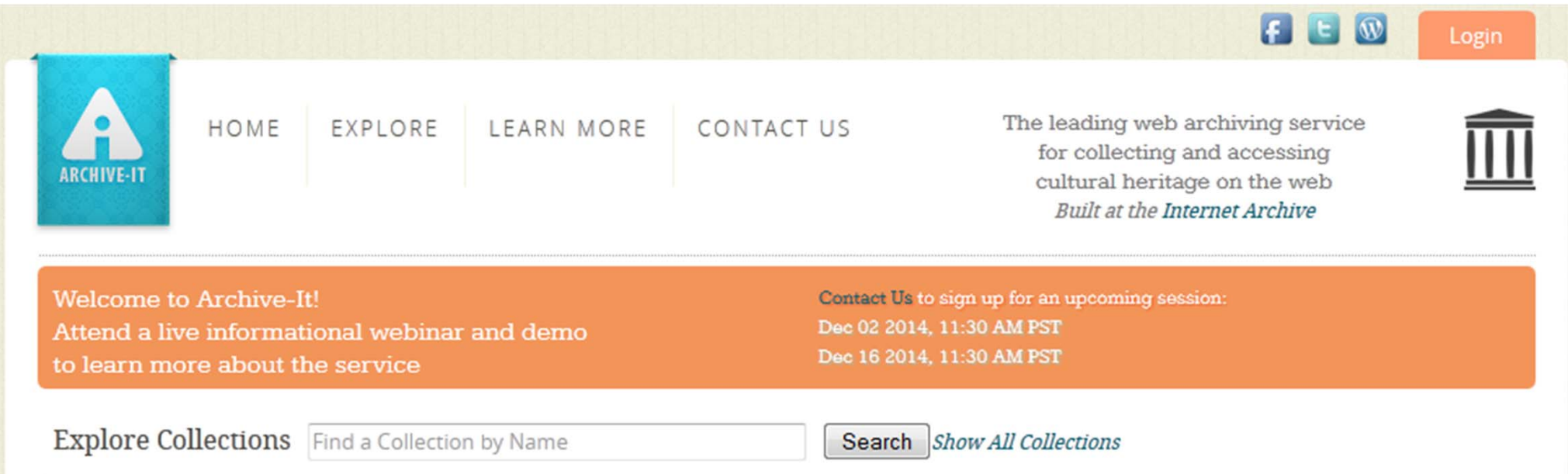


# Web Archiving at K-State: Archive-It



The screenshot shows the Archive-It website interface. At the top right, there are social media icons for Facebook, Twitter, and LinkedIn, and a "Login" button. The main navigation menu includes "HOME", "EXPLORE", "LEARN MORE", and "CONTACT US". Below the navigation, there is a blue banner with the Archive-It logo and the text "ARCHIVE-IT". To the right of the banner, there is a description: "The leading web archiving service for collecting and accessing cultural heritage on the web. Built at the Internet Archive" with a classical building icon. Below this, there is an orange banner with the text: "Welcome to Archive-It! Attend a live informational webinar and demo to learn more about the service". To the right of this banner, there is a "Contact Us" link and two dates: "Dec 02 2014, 11:30 AM PST" and "Dec 16 2014, 11:30 AM PST". At the bottom, there is a search bar with the text "Explore Collections" and "Find a Collection by Name", a "Search" button, and a link "Show All Collections".

4 December 2014  
Web Editors' meeting  
Cliff Hight, University Archivist

# Kansas Archive-It Consortium Intro

- Developed during late 2013 and 2014
- Members include:
  - Kansas Historical Society
  - K-State
  - KU (including Dole Institute of Politics)
  - Washburn University
  - Emporia State
  - Fort Hays State

# Outline for Rest of Presentation

- Why preserve web content?
- What tools are we using?
- How do archived crawls look?
- What will we do next?

# Why Preserve Web Content?

- Content is moving there
  - Paper used to be main medium
  - Some never sees paper today

K-State home » Vice President for Research » News » 2014 » November 20, 2014

## Office of Vice President for Research



### RSCAD Momentum

- Home
- RSCAD Momentum**
- Events
- Profile
- Reporting units

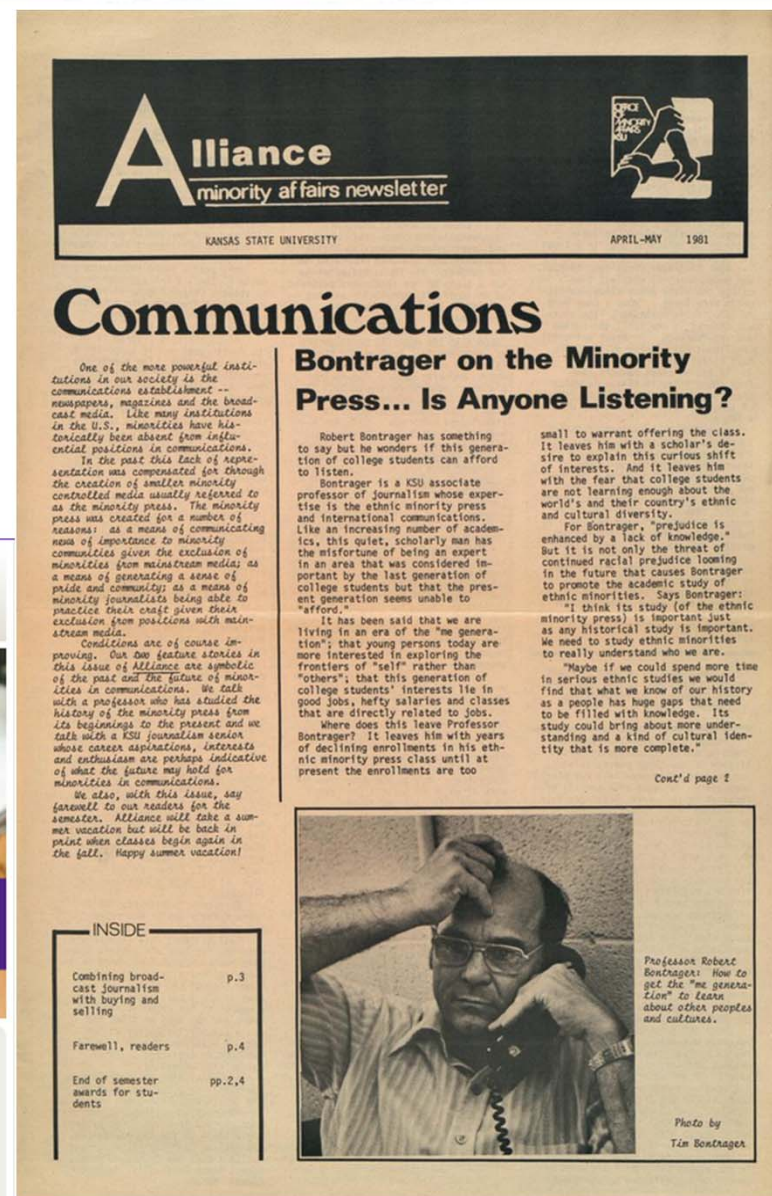
## Research, Scholarly and Creative Activities, and Discovery (RSCAD) News

November 20, 2014

The weekly RSCAD newsletter provides the latest research news, funding opportunities and academic trends.

### Issues

- November 20, 2014
- November 13, 2014
- November 6, 2014
- October 30, 2014
- October 23, 2014
- October 16, 2014
- October 9, 2014
- October 2, 2014



# Why Preserve, II

INTERNET ARCHIVE  
WaybackMachine

1.451 captures  
12 Dec 98 - 1 Dec 14

http://www.k-state.edu/ Go

NOV DEC APR  
1997 12 1998 2000

Close X  
Help ?

Quick Search

Directories

- People
- Offices
- Colleges & Departments
- Organizations

University Calendars

KATS

News, Weather & Sports

Kansas State University  
Manhattan, KS 66506  
(785) 532-6011  
[Text mode](#)

Friday, December 11, 1998

**About K-State**

- [History & Facts](#)
- [Achievements](#)
- [Campus map](#)
- [International & Multicultural](#)

**Prospective Students**

- [Consider K-State](#)
- [Admissions](#)
- [Housing](#)
- [Student Financial Aid](#)

**Academics**

- [Advising](#)
- [Courses](#)
- [Continuing Education](#)
- [Graduate School](#)
- [KSU Libraries](#)

**Outreach**

- [Lifelong Learning](#)
- [Research & Extension](#)
- [Alumni](#)
- [Foundation](#)

**Information Technology**

- [Vice Provost and IT Units](#)
- [IT Policies & Guidelines](#)
- [IT Assistance](#)
- [Year 2000 Readiness](#)

**Service & Support**


- [University Organization](#)
- [Publications & Policies](#)
- [Employment](#)
- [Student Union](#)

**Provost**

- [Faculty Handbook](#)
- [Research & Technology](#)
- [Strategic Planning](#)

**Arts & Events**

- [Calendars](#)
- [Lecture Series](#)
- [McCain Auditorium](#)
- [Beach Art Museum](#)

Search Directories    **K-State**

[Search](#) | [Directories](#) | [Calendar](#) | [Help](#) | [Comments](#)  
[Copyright Information](#)  
December 4, 1998

- Historical interest
  - See websites throughout time
  - Visual history of web design

# Why Preserve, III

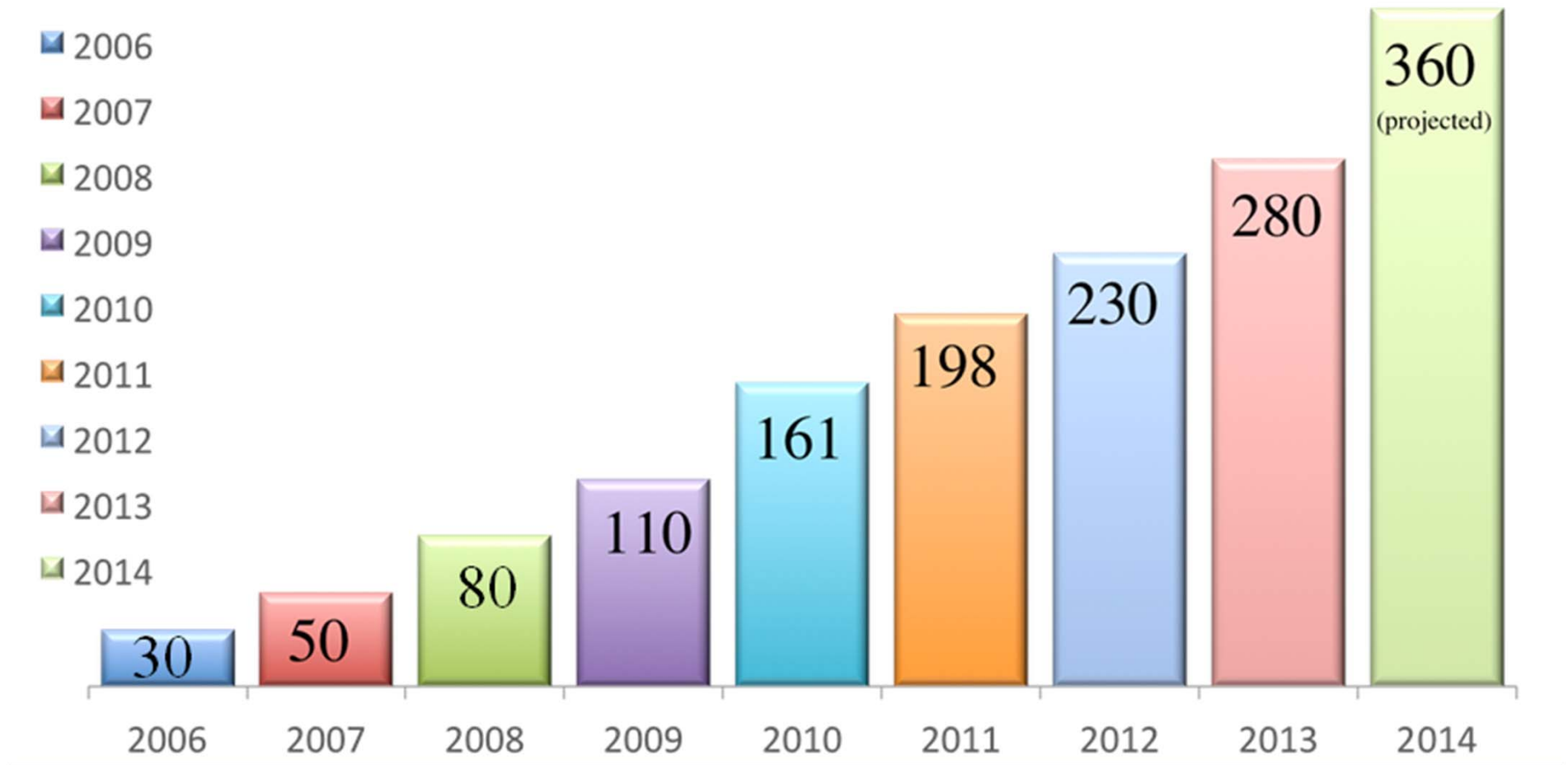
- Potential for research (Think K-State 2025!)
  - Preserves content that often disappears later
  - Government information in the digital era
  - Machine access for types of “big data” analysis

## Sources for further reading:

- Peter Stirling, Philippe Chevalier, and Gildas Illien, “Web Archives for Researchers,” *D-Lib Magazine* 18, no. 3/4 (March/April 2012), see: <http://www.dlib.org/dlib/march12/stirling/03stirling.html>.
- Stanford University Libraries, “Web Archiving, Use cases,” see: <http://library.stanford.edu/projects/web-archiving/use-cases>.
- Emily Reynolds, “If We Capture, Will They Come? Researcher Uses for Web Archive Collections,” *The Signal, Digital Preservation Blog*, Library of Congress, 12 March 2013, see: <http://blogs.loc.gov/digitalpreservation/2013/03/if-we-capture-will-they-come-researcher-uses-for-web-archive-collections/>.
- Oxford Internet Institute, “Using Web Archives: A Futures Perspective,” February-June 2011, see: <http://www.oii.ox.ac.uk/research/projects/?id=85>.

# Why Preserve, IV

- Archive-It partners, 2006-2014



# What Tools Are We Using?

- Internet Archive (<http://archive.org/>)
  - Home of collections for video, live music, audio, texts, TV news, and more
  - Home of software collection (<http://archive.org/details/software>)
  - Home of Internet Arcade (<http://archive.org/details/internetarcade>)
  - Home of Wayback Machine (<http://archive.org/web>)
  - Home of Archive-It service (the one that matters for this presentation, (<http://www.archive-it.org/>))




# What Tools, II

- Archive-It tools
  - Heritrix Web Crawler collects content
    - Written in Java by Internet Archive and others, open and free
    - Writes content to WARC files, dedups documents, etc.
  - Umbra Browser Automation Tool also captures
    - Allows preservation of dynamic components of sites
  - NutchWAX for full text searching
  - Wayback Machine for viewing and access
  - Solr for metadata searching

# How Do Archived Crawls Look?

- Example from Emporia State University
  - <http://www.archive-it.org/organizations/892>

Explore >> Emporia State University



**Emporia State University**  
Archive-It Partner Since: Nov, 2014  
Organization Type: [Colleges & Universities](#)  
Organization URL: <http://emporia.edu>

Narrow Your Results

Subject Sort By: **Count** | [\(A-Z\)](#)

Universities & Libraries (2)

Creator Sort By: **Count** | [\(A-Z\)](#)

ESU Office of the President (1)

Date Sort By: **Count** | [\(A-Z\)](#)

2014 (1)

Language Sort By: **Count** | [\(A-Z\)](#)

Sites and collections from this organization are listed below. Narrow your results at left, or enter a search query below to find a collection, site, specific URL or to search the text of archived webpages.

[Collections](#) | [Sites](#) | [Search Page Text](#)

Page 1 of 1 (2 Total Results)

Sort By: [Collection Name \(A-Z\)](#) | [Collection Name \(Z-A\)](#)

**ESU Office of the President**  
Archived since: Nov, 2014  
Description: Web presence of the Emporia State University Office of the President.  
Subject: [Universities & Libraries](#)

# How Crawls Look, II

- K-State's site on Wayback Machine
  - [http://web.archive.org/web/\\*/k-state.edu](http://web.archive.org/web/*/k-state.edu)



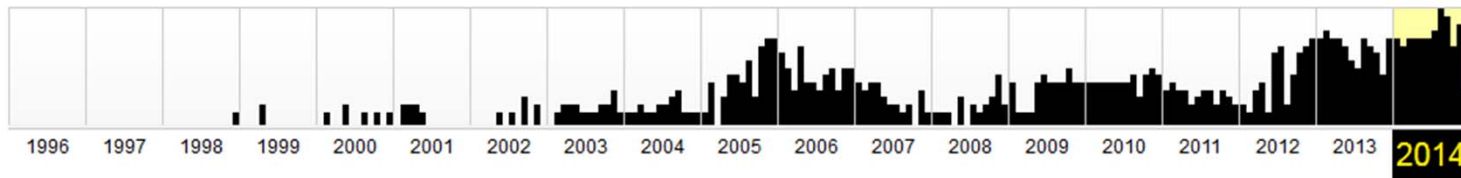
http://k-state.edu

BROWSE HISTORY

<http://k-state.edu>

Saved **1,453 times** between December 12, 1998 and December 2, 2014.

**PLEASE DONATE TODAY.** Your generosity preserves knowledge for future generations. Thank you.



# What Will We Do Next?

- Determine what to crawl
  - K-State sites
  - Collection strength areas (cooking, agriculture, Kansas life and culture, military history, consumer movement, etc.)
  - Possibly create a web-based nomination form
- Make content publicly available
- Publicize availability



Thanks!

Questions?

Library: [www.lib.k-state.edu](http://www.lib.k-state.edu)

Special Collections: [www.lib.k-state.edu/special-collections](http://www.lib.k-state.edu/special-collections)

Archive-It collections: [www.archive-it.org/organizations/890](http://www.archive-it.org/organizations/890)

My email: [chight@ksu.edu](mailto:chight@ksu.edu)