Archiving Michigan State University's Website:
Appraisal, Inventory, and Selection of University Web Properties

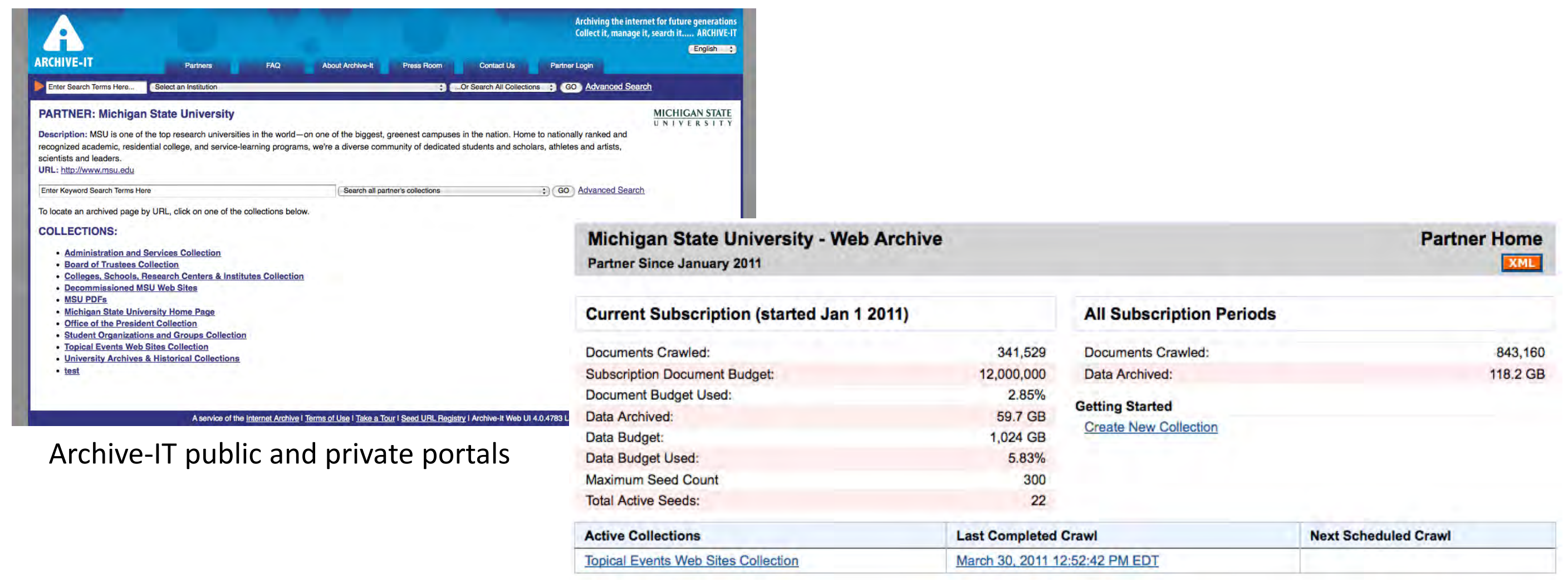
By Ed Busch, Julia Corrin, Cynthia Ghering

The Project

Michigan State University Archives & Historical Collections contracted with Internet Archive to use the Archive-It service to crawl, preserve, and provide access to University web sites. Since the university's web presence is too large to archive in totality, MSU undertook both a self study and a peer review to determine what web content should be archived, identify the frequency of the crawls, and explore existing best practices for web site preservation.

The Internet Archive

Founded in 1996, the non-profit Internet Archive is building a digital library of Internet sites. Their Wayback Machine allows users to see archived version of web pages and their associated data (images, source codes, documents, etc.). With the use of data from Alexa Internet and the Heritrix web crawler, the Internet Archive provides an on-demand archiving service, "Archive-IT," to institutional subscribers.


ARCHIVE-IT
Archiving the internet for future generations
Collect it, manage it, search it..... ARCHIVE-IT
English
Partners
FAQ
About Archive-It
Press Room
Contact Us
Partner Login
Enter Search Terms Here...
Select an Institution
...Or Search All Collections
GO
Advanced Search
MICHIGAN STATE UNIVERSITY
Enter Keyword Search Terms Here
Search all partner's collections
XML

PARTNER: Michigan State University

Description: MSU is one of the top research universities in the world—on one of the biggest, greenest campuses in the nation. Home to nationally ranked and recognized academic, residential college, and service-learning programs, we're a diverse community of dedicated students and scholars, athletes and artists, scientists and leaders.

URL: http://www.msu.edu

To locate an archived page by URL, click on one of the collections below.

COLLECTIONS:

- Administration and Services Collection
- Board of Trustees Collection
- Colleges, Schools, Research Centers & Institutes Collection
- Decommissioned MSU Web Sites
- MSU PDFs
- Michigan State University Home Page
- Office of the President Collection
- Student Organizations and Groups Collection
- Topical Events Web Sites Collection
- University Archives & Historical Collections
- test

A service of the Internet Archive | Terms of Use | Take a Tour | Seed URL Registry | Archive-It Web UI 4.0.4783 L

Michigan State University - Web Archive
Partner Since January 2011
Partner Home

Current Subscription (started Jan 1 2011)	
Documents Crawled:	341,529
Subscription Document Budget:	12,000,000
Document Budget Used:	2.85%
Data Archived:	59.7 GB
Data Budget:	1,024 GB
Data Budget Used:	5.83%
Maximum Seed Count	300
Total Active Seeds:	22

All Subscription Periods	
Documents Crawled:	843,160
Data Archived:	118.2 GB

Getting Started
Create New Collection

Active Collections	Last Completed Crawl	Next Scheduled Crawl
Topical Events Web Sites Collection	March 30, 2011 12:52:42 PM EDT	

Archive-IT public and private portals

The Methodology

The project consisted of eight major activities:

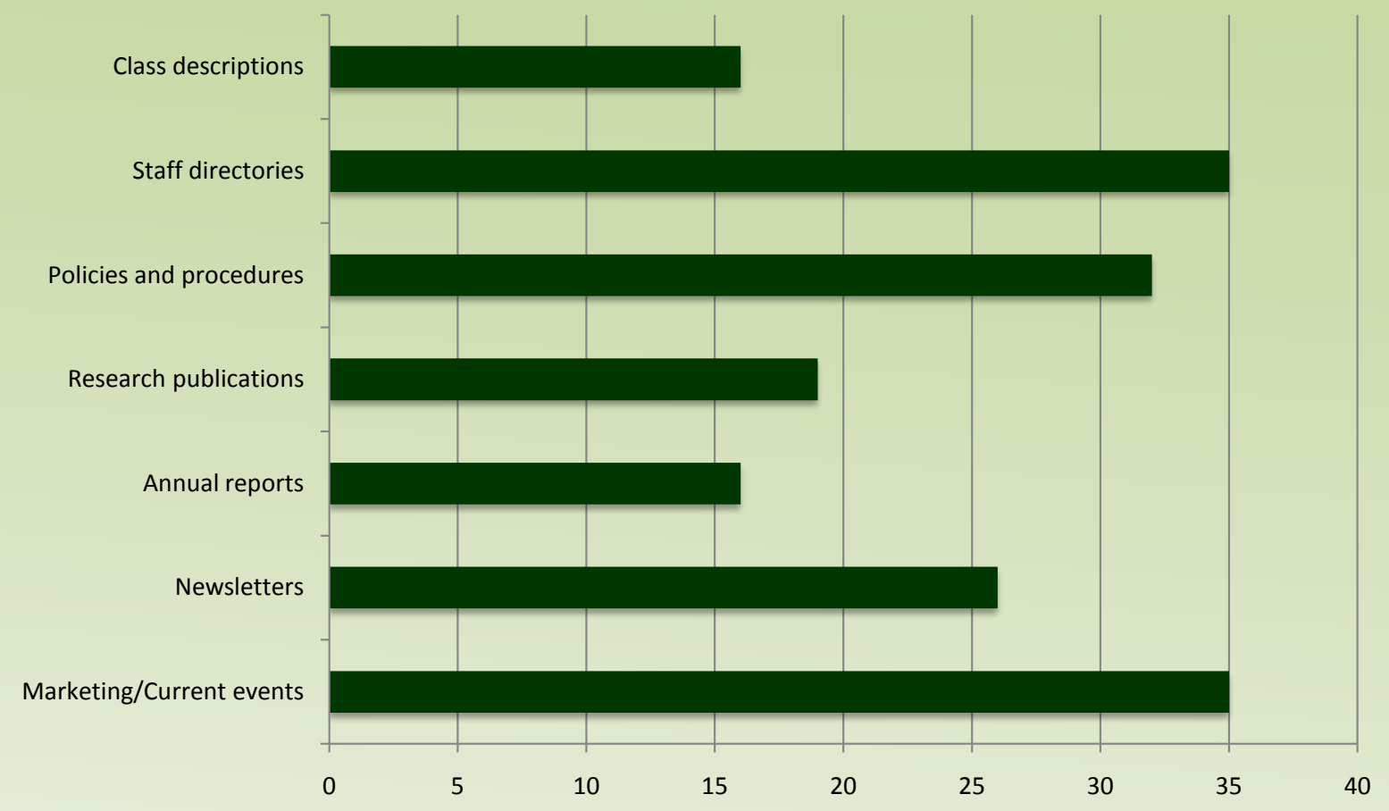
- •Literature Review
- •Peer review of other web archiving projects
- •Online survey of MSU technology staff
- •Inventory of the MSU domain
- •Creation of a web site collecting plan
- •Update of university retention schedule(s)
- •Creation of MSU Collections to crawl
- •Update of University Archives web site for user access

The first four activities resulted in a comprehensive report, authored by University of Michigan intern Julia Corrin, that offered recommendations for the future of the project and a summation of findings related to the organization, size, and content of the web site. Using the final report as a starting point, University Archives staff developed a web site collecting plan and began archiving MSU web sites and preserving web pages and documents.

What is actually in the MSU Web Domain?

The IT Perspective

A web survey was administered to 800 members of an IT staff email list. From the 67 respondents, it was determined that MSU web properties are substantial and updated frequently. Respondents also indicated that the content of their web sites was, for the most part, not duplicated in print. Staff directories, marketing and current events, policies and procedures and newsletters are posted on many MSU web sites.



Website Inventory

There are at least	3,816,112	URLs on the MSU.edu domain.
There are at least	3,670,000	PDFs on the MSU.edu domain.
There are	1,273	known subdomains.
That is almost	3,000	URLs per known subdomain.

Of the 439 subdomains, close to **30%** were redirected to another subdomain. There were 102 subdomains, or **23%,** that included password protected or dynamic content and cannot be archived by current methods.

Retention Schedule Update

Schedule Number: 116.98

Schedule Approved Date: General

Title: MSU Publications

Disposition: Permanent

Disposition Description: Retain one copy in office of creation permanently. This office should send copies to archives either as published or on an annual basis. *The University Archives is capturing most web sites within the msu.edu domain and identified MSU related external sites. Offices should inform the University Archives before retiring old websites and when creating new web sites.*

Description: These are publications created at MSU including pamphlets, brochures, newsletters, magazines, guide-books, bulletins, programs, announcements, videos, *web sites,* electronic publications (one-time and serial publications), for on-campus and off-campus audiences.

(Revisions in *italics*.)

Metadata (Dublin Core)

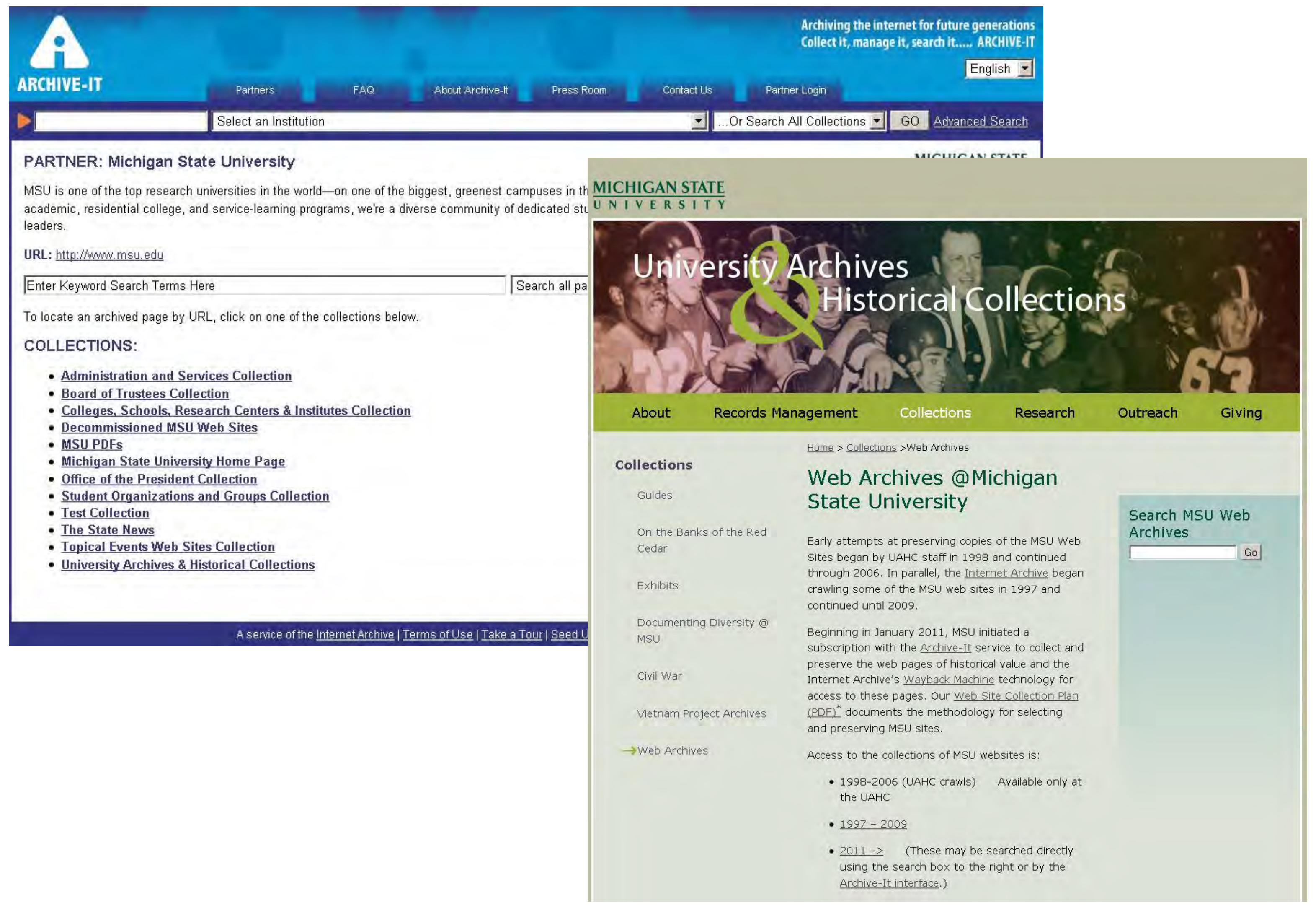
- Title
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Source
- Rights
- Collector

The Results

- Creation of a Web Site Collection Plan (http://archives.msu.edu/documents/CollectionPlan_v1.pdf)
- Update of general retention schedule
- Creation of crawls
  - •Administration and Services Collection
  - •Colleges, Schools, Research Centers and Institutes Collection
  - •Student Organizations and Groups Collection
  - •*The State News* (student newspaper)
  - •Topical Events Web Sites Collection
- User access to these collections through the University Archives site, archives.msu.edu, and the Archive-It Partners page

Recommendations

- Collections
  - Refine MSU Collections in Archive-It as we learn more about the MSU web sites.
  - Perform quality checks on captured sites and modify crawls to better represent the web presence.
- Metadata
  - Develop more robust metadata capture. Benchmark other institutions' metadata models. Add more metadata at Seed and Document levels as practical.
- Outreach
  - Be proactive. Market new web site archiving to MSU colleges and offices. Educate staff during records management training.


ARCHIVE-IT
Archiving the internet for future generations
Collect it, manage it, search it..... ARCHIVE-IT
English
Partners
FAQ
About Archive-It
Press Room
Contact Us
Partner Login
Select an Institution
...Or Search All Collections
GO
Advanced Search
Enter Keyword Search Terms Here
Search all pa
MICHIGAN STATE UNIVERSITY
University Archives & Historical Collections
About
Records Management
Collections
Research
Outreach
Giving
Search MSU Web Archives
Go

PARTNER: Michigan State University

MSU is one of the top research universities in the world—on one of the biggest, greenest campuses in th academic, residential college, and service-learning programs, we're a diverse community of dedicated stu leaders.

URL: http://www.msu.edu

To locate an archived page by URL, click on one of the collections below.

COLLECTIONS:

- Administration and Services Collection
- Board of Trustees Collection
- Colleges, Schools, Research Centers & Institutes Collection
- Decommissioned MSU Web Sites
- MSU PDFs
- Michigan State University Home Page
- Office of the President Collection
- Student Organizations and Groups Collection
- Test Collection
- The State News
- Topical Events Web Sites Collection
- University Archives & Historical Collections

A service of the Internet Archive | Terms of Use | Take a Tour | Seed U

Collections
Guides
On the Banks of the Red Cedar
Exhibits
Documenting Diversity @ MSU
Civil War
Vietnam Project Archives
→Web Archives

Home > Collections >Web Archives

Web Archives @Michigan State University

Early attempts at preserving copies of the MSU Web Sites began by UAHC staff in 1998 and continued through 2006. In parallel, the Internet Archive began crawling some of the MSU web sites in 1997 and continued until 2009.

Beginning in January 2011, MSU initiated a subscription with the Archive-It service to collect and preserve the web pages of historical value and the Internet Archive's Wayback Machine technology for access to these pages. Our Web Site Collection Plan (PDF)* documents the methodology for selecting and preserving MSU sites.

Access to the collections of MSU websites is:

- 1998-2006 (UAHC crawls) Available only at the UAHC
- 1997 – 2009
- 2011 -> (These may be searched directly using the search box to the right or by the Archive-It interface.)

INSTITUTE of Museum and Library SERVICES
ARCHIVE-IT
MICHIGAN STATE UNIVERSITY
University Archives & Historical Collections
Digital PRESERVATION National Digital Information Infrastructure & Preservation Program: A Collaborative Initiative of the Library of Congress
school of information M