

Archiving IU Social Media with Archive-It

...

Digital Library Brown Bag | September 6, 2017

Molly Wittenberg & Tyler Davis

University Archives

#dlbb

@IUBArchives

Overview

**what and
why?**



value



**using
Archive-It:**
collection & research

**what we
learned**



**moving
forward**

The Internet Archive & Archive-It

Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.



305B



15M



3.6M



3.6M



1.4M



191K



1.5M



180K



292K



GO

[Advanced Search](#)



WayBackMachine



The Internet Archive & Archive-It



Indiana University

Archive-It Partner Since: Sep, 2005

Organization Type: Colleges & Universities

Organization URL: <http://www.libraries.iub.edu/>

Web archives at IU: 2005-present

Why archive the web?

- Publication of current missions, objectives, policies, activities
- Official copy of University records
- Consistent archiving of web content
- Common issue across all IU campuses

Web archives at IU: 2005-present



Digital Library Program

INDIANA UNIVERSITY

[Home](#)

[Collections](#)

[Exhibits & Highlights](#)

[Reference Tools](#)

[Research & Development](#)

[About Us](#)

[Related Links](#)

[Search](#)

[Site Map](#)

The Indiana University Digital Library Program (DLP) is dedicated to the production, maintenance, delivery, and preservation of a wide range of high-quality networked resources for scholars and students at Indiana University and elsewhere.

The Digital Library Program is a collaborative effort of the

- [Indiana University Libraries](#), the
- [Office of the Vice President for Information Technology](#),

and the university research faculty with leadership from the

- [School of Library and Information Science](#) and the
- [School of Informatics](#).

What's new

- [Variations3](#)
 - A new three-year IMLS grant
- [Fall '05 Brown Bag Workshops](#)
- [Sound Directions](#)
- [Chymistry of Isaac Newton](#)
- [Job Openings](#)

This collaboration capitalizes on the institutional capabilities of this university, focusing university resources on digital library services and projects that support the teaching and research of IU faculty, support the learning and research of IU students, and foster research about the digital library.

Last updated: 14 October 2005
URL: <http://www.dlib.indiana.edu/>
Comments: diglib@indiana.edu

Copyright 1997-2005, The Trustees of Indiana University

IU Social Media collection, est. May 2017

Title: Facebook: African Studies Program (Bloomington)

URL: <https://www.facebook.com/African-Studies-Program-367538175107/>

Captured 2 times between Jun 29, 2017 and Jun 30, 2017

Group: Facebook

Title: Facebook: Center on Aging and Community

URL: <https://www.facebook.com/Center-on-Aging-Community-Indiana-University-107042329334011/>

Captured 2 times between Jun 30, 2017 and Jul 1, 2017

Videos: 58 Videos Captured

Group: Facebook

Title: Facebook: Department of French and Italian (Bloomington)

URL: <https://www.facebook.com/FRITIUI/>

Captured 3 times between Jun 29, 2017 and Jun 30, 2017

Videos: 124 Videos Captured

Group: Facebook

Facebook (136)

Flickr (17)

Google+ (17)

Instagram (34)

Pinterest (18)

Twitter (118)

Youtube (52)

Value of web archives

Research Value

- Internet History as a burgeoning field
- Cultural heritage research
 - What were people talking about? What were they sharing and seeing?

Administrative Value

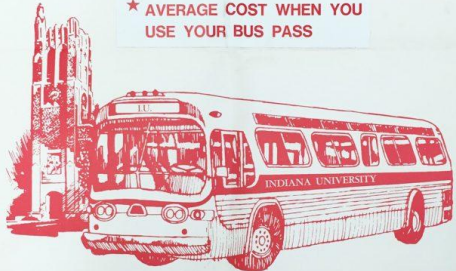
- How was an office's website designed?
- How has the interface progressed over time?
- An ongoing archive of relevant images, documents, staff information

Promotion and information-sharing

BIGGEST BARGAIN ON CAMPUS

6¢ A RIDE!! *

★ AVERAGE COST WHEN YOU
USE YOUR BUS PASS



...IS ALL IT COSTS YOU TO RIDE AN INDIANA UNIVERSITY CAMPUS BUS! RIDE IT ANYTIME...AND AS OFTEN AS YOU LIKE WHEN YOU PURCHASE A \$25 SEMESTER PASS. FAST AND CONVENIENT BUSES RUN FROM 7:00 A.M. TO MIDNIGHT, MONDAY THRU SATURDAY, AND 8:00 A.M. TO 11:00 P.M. SUNDAY.

PASSES AVAILABLE AT:

- Registration
- Married Housing
- Residence Halls
- IU Ticket Office (Assembly Hall)
- Bursar's Office
- Campus Bus Office 315 West Sixth Street

BUY YOUR PASS TODAY!!

Campus Bus promotional poster, circa 1970's
Institute for Urban Transportation records

http://webapp1.dlib.indiana.edu/findingaids/view?doc.view=entire_text&docId=InU-Ar-VAD9210

INDIANA UNIVERSITY
CAMPUS BUS SERVICE
Bloomington

campus bus | parking operations | parking enforcement | motor pool

INDIANA UNIVERSITY BLOOMINGTON
campus bus

The time is: 3:38:51 PM

- View Live Bus Locations
- Bus Services
- Bus Schedules
- Bus Tips
- Night Owl
- Student Transportation Board
- Advertising
- All Routes Map
- Student Employment
- Campus Bus Information
- Bloomington Transit Bus Pass
- Bloomingtontransit.com



CAMPUS BUS

Campus Bus Service
Employee Portal
Photos by A. M. Spierer
120 W. Gates Lane
Bloomington, Indiana 47403
Phone: (317) 850-6304
E-Mail: campusbus@iub.edu

IU Campus Bus @IUBCampusBus · 18 Feb 2016

Last week we upgraded our website. What do you think? Big upgrade from our website for the past 11 years.

Campus Bus promotional Twitter post, February 2016

Campus event announcements

"The great mass meeting in the Men's Gymnasium next Monday afternoon will give every student an opportunity to show interest in the history of Indiana. In the presence of Governor McCray there has been an opportunity to talk over the details of the campaign and to hear the opening of next Monday be a grand success."

SEVEN

"I realize the all-Indiana mass meeting next Friday at one of the high points in the history of the State. It is desirable to have a representative meeting on the history of the subject. Every student should feel free to come to participate in the same that will be reported at the meeting to have proved to be great."

—WILLIAM C. B. BROWN/INDIANA

MEANS MARCH 7

DATE OF THE
GREATEST STUDENT MASS MEETING
2:30 P. M., TUESDAY, MEN'S GYM
EVER HELD AT INDIANA UNIVERSITY

One Hour After The Meeting Active Solicitation
Among Students Will Start

No Money Is To Be Pledged At The Mass Meeting
None Will Be Accepted If Voluntarily Pledged

*Governor McCray Will Make The Opening
Speech Of The Meeting—He Has Side-Trackd
Everything To Come To Bloomington—HE BE-
LIEVES IN INDIANA*

**Most Of The Talks Will Be By Students. The Program
Will Make Every Man And Woman Proud
Of INDIANA**

**Students, Faculty, Townspeople—Everyone Who
Loves Indiana Will Be There—All Eyes Are Turned
On Indiana Students**

THEY WILL
Show The Governor Real Indiana Spirit

Secure | <https://www.facebook.com/iuill/photos/a.197454769352478.40268.12375360772594/122316436114841/?type=3&theater>

Student Life and Learning, Indiana University Bloomington (IUSLL)
Like This Page · March 28 ·

#CoxScholars#IU

You are invited to the

Cox Civic Scholar End of Year Poster Presentations

**Friday, March 31, 2017
3:30PM - 5PM
IMU - Hoosier Room**

⏪ ⏩

This presentation will provide an opportunity to hear about what the Cox Civic Scholars have learned this year through their work in the community. Students, faculty and staff are welcome to stop in anytime during the event.

**Reserve space by contacting Brandon Shurr at
bshurr@indiana.edu**

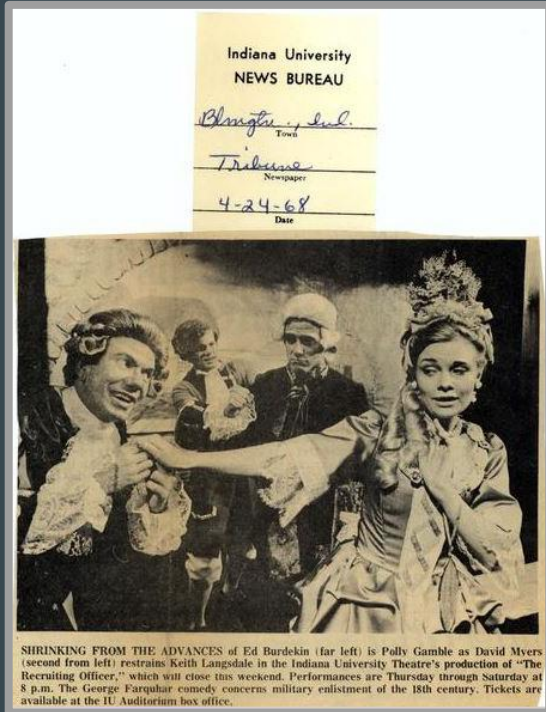
Timeline Photos

Student Life and Learning digital event flyer,
Facebook, March 2017

Memorial Fund event flyer, paper, circa 1920's
Memorial Fund Campaign records

<http://webapp1.dlib.indiana.edu/findingaids/view?brand=general&docId=InU-Ar-VAD5939.xml>

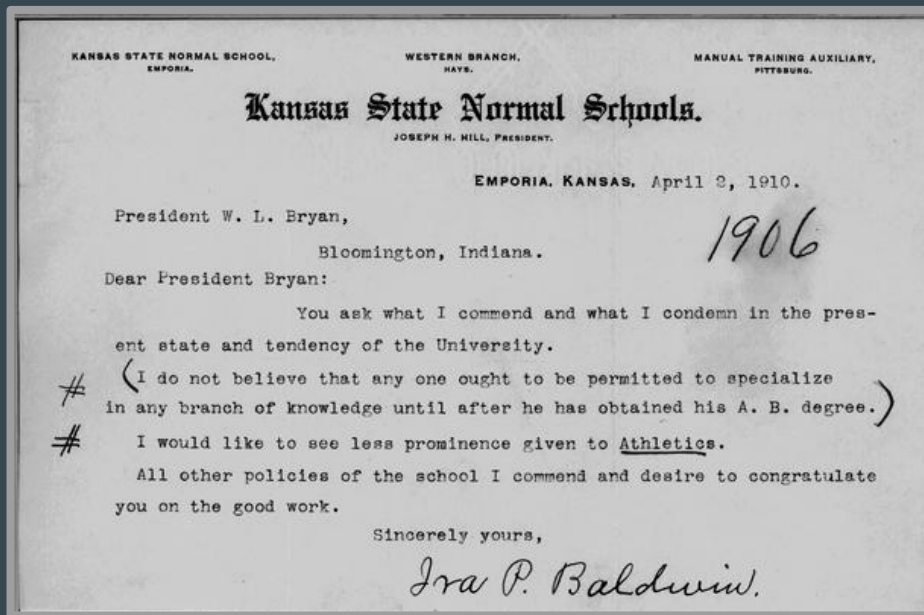
Photographs and visual history



Promotional photo for *The Recruiting Officer* at IU Theatre, April 1968
Department of Theatre and Drama records
<http://webapp1.dlib.indiana.edu/findingaids/view?brand=general&docId=InU-Ar-VAC0946.xml>

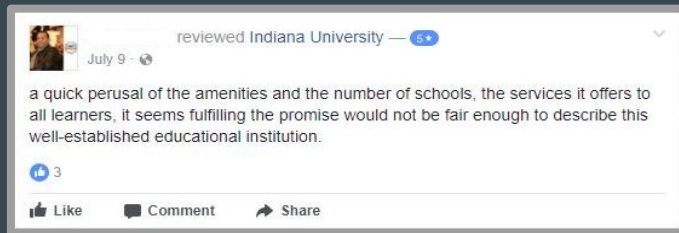
Promotional photo for *Love's Labour's Lost* at IU Theatre via Instagram, July 2017

Feedback from students, alumni, and community members



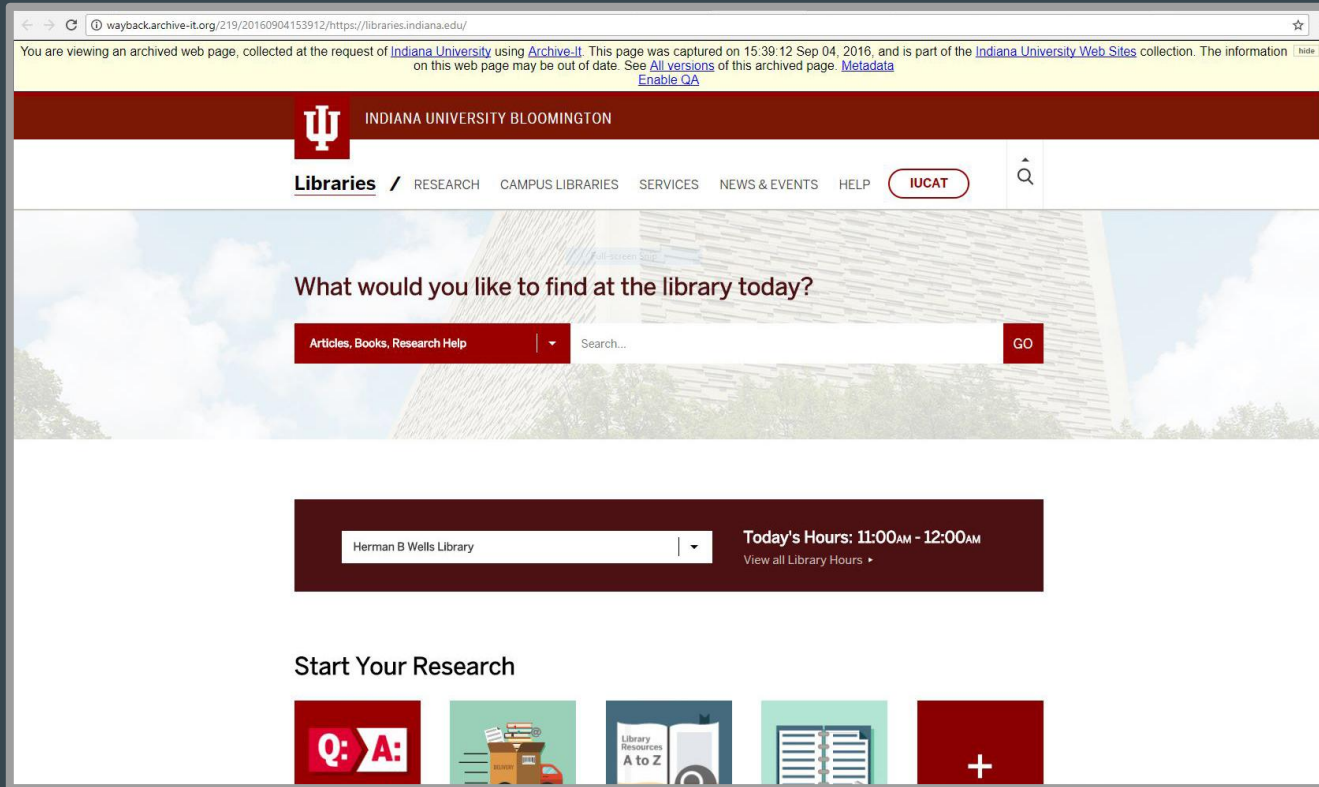
Letter from IU alumni on the state of the University, 1910
President's Office Correspondence records

<http://webapp1.dlib.indiana.edu/findingaids/view?brand=general&docId=InU-Ar-VAA3710.xml>



Posts on IU Bloomington Facebook
page, 2014-2017

Web design and interface over time



Indiana University Libraries homepage, September 2016
from the Indiana University Web Sites collection on Archive-It.org

Library social media accounts we are currently capturing



- Black Film Center/Archive

- Facebook
- Twitter

- Center for Innovative Teaching and Learning (CITL)

- Facebook
- Twitter
- YouTube

- IU Bloomington Libraries

- Facebook
- Flickr
- Twitter
- YouTube



- Lilly Library

- Facebook
- Twitter



- Department of Information and Library Science (ILS)

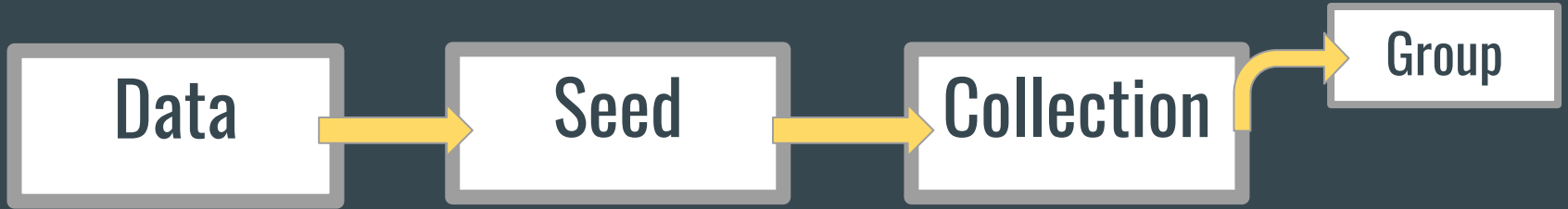
- Facebook
- Instagram
- Twitter

- University Archives

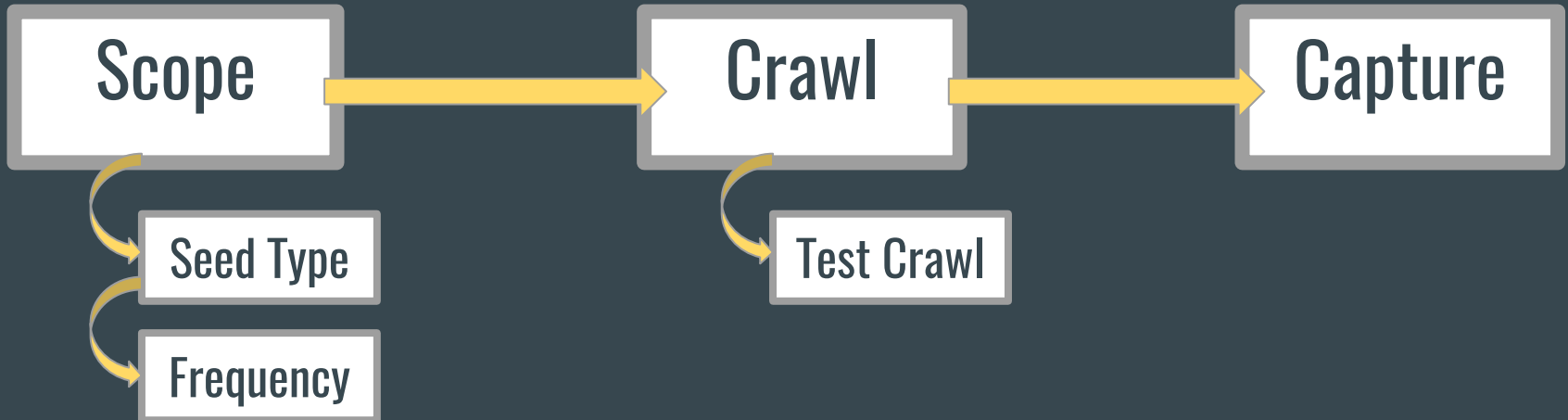
- Facebook
- Instagram
- Twitter



The Basics



Using Archive-It



Test Crawls

Using Archive-It

- Test crawls do not count against data subscription
- Faulty test crawls may be deleted
- Good test crawls should be saved
- Quality Assurance (**QA**) to improve quality of crawls

How much to crawl?

Crawl types:

Using Archive-It

Standard



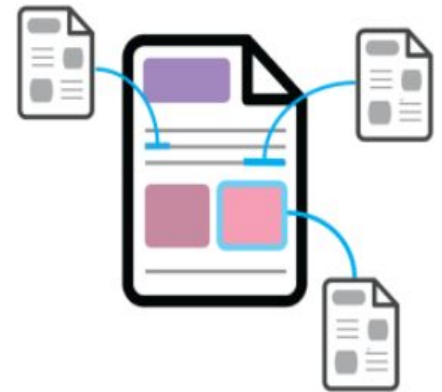
Standard +



One Page



One Page +



How often to crawl?

Office of Sustainability
Website crawls,
2016-2017

2016	2017
6 pages	3 pages
Mar 4, 2016 *	Mar 4, 2017
Jun 4, 2016 *	Jun 4, 2017
Jun 4, 2016	Sep 4, 2017
Sep 4, 2016 *	
Sep 4, 2016	
Dec 4, 2016 *	

Using Archive-It

Office of Sustainability,
Facebook crawls, 2017

**Found 3 Captures between Jun 30, 2017 -
Jul 14, 2017**

2017

3 pages

[Jun 30, 2017](#) *

[Jun 30, 2017](#) *

[Jul 14, 2017](#) *

How much data to manage?

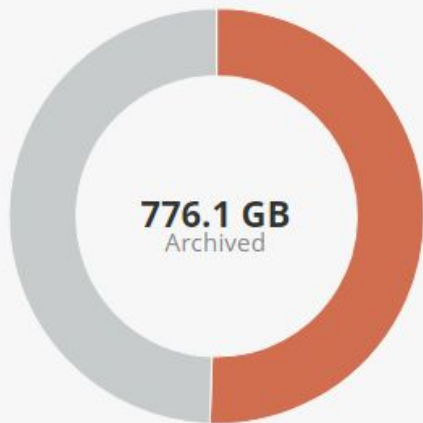
Using Archive-It

Data usage as of 9/1/2017:

Indiana University

11.7 TB archived since Sep 8, 2005

Current Subscription



Data Budget Usage

Active Collection List (5 Active Collections)

[Create a Collection](#)[Download Collection List](#)

Collection Name	Data (this period) ▾	Docs (this period)	Active Seeds	Last Crawl
Indiana University Social Media	566.1 GB	1,566,671	427	Aug 24, 2017
Indiana: State and Local Docum	161.2 GB	2,260,153	92	Aug 3, 2017
Indiana University Web Sites	48.8 GB	2,309,104	210	Aug 28, 2017
National Government Statistical	0 bytes	0	217	Nov 13, 2016
Open Folklore Legacy Web Arch	0 bytes	0	3	Mar 29, 2011

Ignore robots.txt exclusions?

Using Archive-It

1. **www.facebook.com** - in order to archive Facebook-hosted content
2. **fbcdn.net** - in order to archive important page styling elements
3. **akamaihd.net** - in order to archive important page styling elements

Archive-It Help Center: Archiving Facebook.
<https://support.archive-it.org/hc/en-us/articles/208333113-Archiving-Facebook-pages>

Limitations

- Functional limitations of crawls
- Technical limitations of Archive-It

Not in Archive

The page you requested has not been archived in Archive-It.

This could be for a number of reasons.

Most likely the page you are requesting was outside of the crawler's scope. Try another request or click to see other pages from <https://>.

It is also possible that this site is currently being crawled and the archived pages are not yet available in the Wayback Machine. It can take up to 24 hours after a crawl has been completed for the site to appear in the Wayback Machine. Please try again after the allotted time.

Security Check

Please enter the text below

Hit reload on your browser to refresh this page if this image doesn't load.

Can't read the text above?

Try another text or an audio captcha

Text in the box:

What's this?

Submit

Searching & locating the resources

Using Archive-It

USEFUL LINKS

Blogging Hoosier History

Guide to research at the IU Archives

IU Chronology

IU Board of Trustees minutes

Access to student records

Bloomington Faculty Council records

Arbutus (yearbook)

IU web sites on Archive-It.org

Archives of Institutional Memory

1. Indiana University web sites

Publication Date: 2006

Language: English

Format:  Journal, Periodical  Online Resource

Copies At: Bloomington, Kokomo, Columbus, East, Fort Wayne, Indianapolis, Northwest, South Bend, and Southeast

IUCAT search results: Indiana University web sites

Indiana University Web Sites on Archive-It.org

<https://archive-it.org/collections/219>

Indiana University Social Media Accounts on Archive-it.org

<https://archive-it.org/collections/8920>

Searching & locating the resources

Using Archive-It



Indiana University

Archive-It Partner Since: Sep, 2005

Organization Type: [Colleges & Universities](#)

Organization URL: <http://www.libraries.iub.edu/>

Description: The IU Bloomington Libraries are one of the leading academic research library systems in North America, providing strong collections, quality service and instructional programs, and leadership in the application of information technologies. The Bloomington campus library holdings include more than 6 million bound volumes and more than 17 million other materials, including manuscripts, maps, music, microforms, and films. The IU Bloomington Libraries comprise the Main Library (which houses collections in the humanities and social sciences), subject-specific campus libraries, and other units such as the Wylie House Museum and University Archives.

Narrow Your Results

Subject

Sort By: **Count** | (A-Z)

Universities & Libraries (3)

Arts & Humanities (1)

Blogs & Social Media (1)

Folklore (1)

Government (1)

More ▾

Creator

Sort By: **Count** | (A-Z)

Open Folklore Project (1)

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 (5 Total Results)

Sort By: **Collection Name (A-Z)** | Collection Name (Z-A)

Indiana University Social Media Accounts

Archived since: May, 2017

Enter a search term on the right to search the text within the archived pages. Or for more search options, use the Advanced Search options below.

Advanced Search

Contains all of:

Exact phrase:

Not containing:

From the Host:

ex. www.archive-it.org

Results per host:

1 (default) ▾

File format:

All formats ▾

Capture date range:

From: ▾ ▾

To: ▾ ▾

Advanced Search

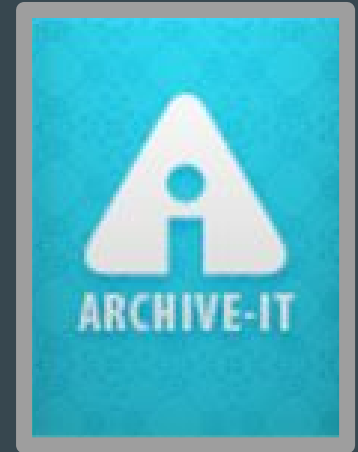
[? Help with Search](#)

What we learned

- All accounts **do not** need equal amounts of data collection
 - Data is not free or unlimited, so scoping crawls to conserve data is essential
- Run test crawls on new seeds -- save good test crawls
- Technical errors can have a variety of causes and solutions
- Keep notes on the workflow -- and keep them current!
- The [Archive-It Help Center](#) is a wonderful resource!

Moving forward

- Identify accounts outside of the [IU social media index](#)
- Fine-tune approach to data collection
- Metadata for social media collection
- Outreach
 - Expand use in reference services



Contact information

Molly Wittenberg
Records Manager
University Archives
molwayne@indiana.edu

Tyler Davis
Graduate Student
Master of Library Science
davistyl@indiana.edu

[Indiana University Social Media Sites at Archive-It.org](#)