Developing a Web Archiving Program at Rollins College

Angelina Altobellis

Digital Archivist, Olin Library

Society of Florida Archivists Annual Meeting | May 9, 2014

ROLLINS

A bit of context

- Small, independent liberal arts college established in 1885

- Archives established 1933

- Mission: to "assemble, preserve, and make accessible the official records of Rollins College and its special, historical collections."

- Systematic acquisition of electronic records (defined broadly) began in 2010

- Digital archivist hired 2013

- College website redesign 2013-2014

Why preserve websites and web content?

"The average lifespan of a Web page today is 100 days. This is no way to run a culture."

(Brewster Kahle, quoted in 2003)

"Unfortunately and disturbingly, the Supreme Court appears to have a vast problem with 'link rot'... We found an alarmingly high number of websites cited by the Supreme Court that are no longer working, almost one-third (29%) of the time."

(Liebler and Liebert, 2013)

Academics

Rollins features small classes taught by renowned faculty. Our academics are enhanced by study-abroad and internship opportunities as well as ways to be engaged on campus and in the community.

▶ MAJORS & MINORS

▶ SPECIAL PROGRAMS

▶ SCHOOLS/COLLEGES

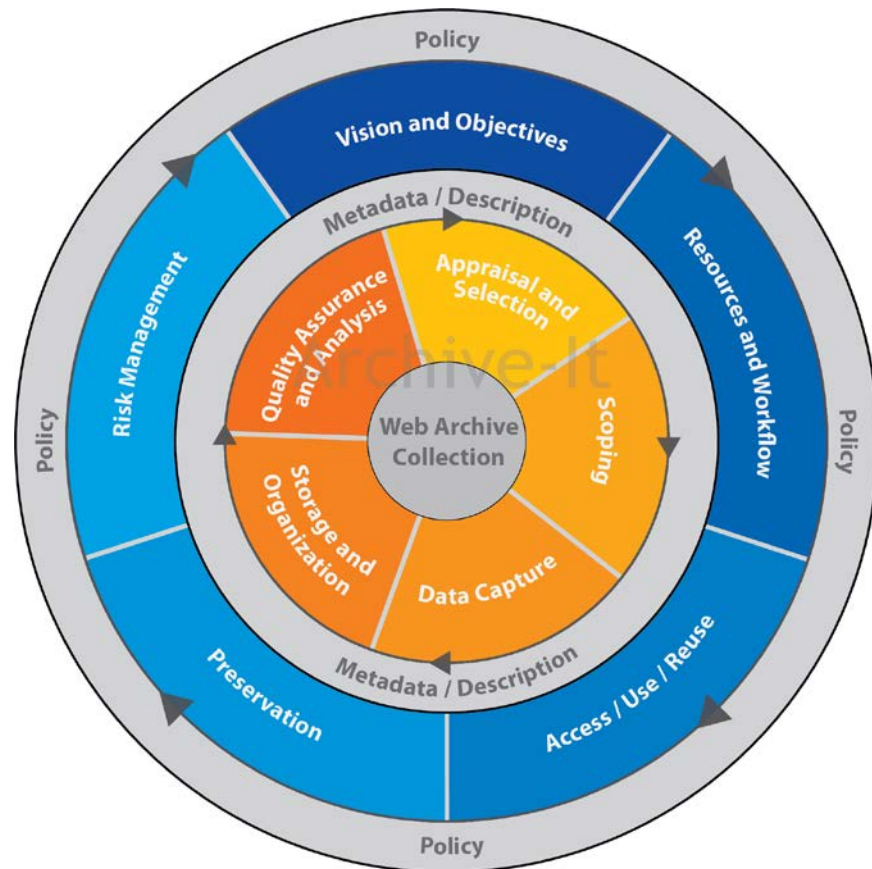▶ CONTINUING EDUCATION

▶ INTERNATIONAL LEARNING

▶ STUDENT SUCCESS

▶ ACADEMIC RESOURCES

▶ ACADEMIC CALENDARS

The World Comes to Rollins

The diverse student body totals more than 3,000 students, representing nearly every state as well as countries in Africa, Asia, Australia, Central America, Europe, and South America.

Web archiving program components

- Appraisal and selection
- Organization and description
- Crawling (Acquisition)
- Storage
- Quality Control
- Access
- Preservation

https://archive-it.org/static/files/archiveit_life_cycle_model.pdf
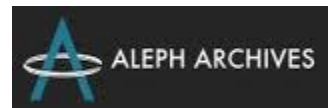
The technical piece: homegrown or vendor?

- We opted to work with a vendor because:
  - Time was a factor
  - Considerations for program sustainability
- …but which vendor?

Do you use an external service or organization to archive, or do crawling in-house?

Both, 9, 14%

In-House, 16, 26%

External, 38, 60%
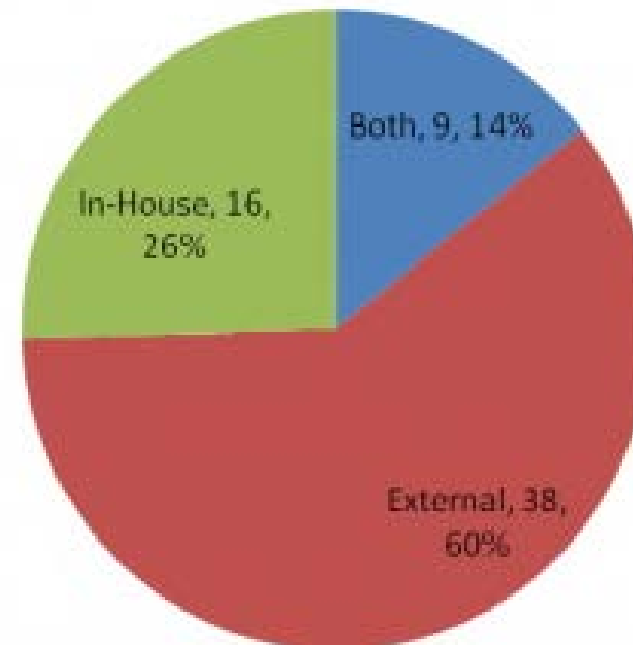
Source: NDSA Web Archiving Survey Report (2012), p. 11

Surveying the options (summer 2013)

Surveying the options (summer 2013)

Policy development (September 2013)

- Program goals

  - "Through its Web Archiving Program, Olin Library selects, captures, describes, preserves, and makes accessible the historic web presence of Rollins College. Its purpose is to support records management and to document the history of the College."

- Collecting scope

- Preservation and access measures

Policy development (September 2013)

- Collecting scope:
  - Administrative units
  - All academic units and the departments within
  - Student support units
  - Important cultural and artistic endeavors of the College
  - Athletics
  - College publications
  - Social networking sites maintained by specific College units

Policy development (September 2013)

- Left open the possibility of thematic collections that support the curriculum and/or complement existing special collections:

> "As the Web Archiving Program develops, the Library may explore the addition of the websites of affiliated organizations (e.g., the Bach Festival Choir and the Genius Preserve), local organizations (e.g., the Hannibal Square Heritage Center), thematic collections that support the curriculum, and websites that complement existing archives and special collections."

Organization & description (November 2013-present)

- Campus Publications (November 2013)

- Rollins College (December 2013)

  - One large collection or individual units?

- Course Catalogues and Schedules (March 2014)

- Apopka Farmworkers Collection (March 2014)

  - First thematic collection, complements existing special collection

Organization & description (November 2013-present)

What's next?

- Quality Control guidelines & implementation

- Web Archiving Policy redux

- Add web collections to Archon finding aids

  - Indexing by Summon for discovery

- Preservation planning

- Explore CINCH (Capture Ingest and Checksum Tool) to package & authenticate selected content

Citations

- Weiss, Rick. "On the Web, Research Work Proves Ephemeral: Electronic Archivists Are Playing Catch-Up in Trying to Keep Documents From Landing in History's Dustbin." *Washington Post* Nov 23, 2003, available from http://faculty.missouri.edu/~glaserr/205f03/Article_WebPub.html

- Liebler, Raizel and Liebert, June. "Something Rotten in the State of Legal Citation: The Life Span of a United States Supreme Court Citation Containing an Internet Link (1996-2010)." *Yale Journal of Law and Technology*: Vol. 15: Iss. 2, Article 2. http://digitalcommons.law.yale.edu/yjolt/vol15/iss2/2

- Archive-It Team, Internet Archive, "The Web Archiving Lifecycle Model." March 2013. https://archive-it.org/static/files/archiveit_life_cycle_model.pdf

- National Digital Stewardship Alliance, Content Working Group, "Web Archiving Survey Report." June 2012. http://www.digitalpreservation.gov/ndsa/working_groups/documents/ndsa_web_archiving_survey_report_2012.pdf

- Michael Shallcross, "On the Development of the University of Michigan Web Archives: Archival Principles and Strategies." Society of American Archivists Campus Case Study 13. April 2011. http://files.archivists.org/pubs/CampusCaseStudies/Case13Final.pdf