



A **flickr** of Hope: Harvesting Social Networking Sites with Archive-It

NDIIPP Partners Meeting

July 21, 2010





DCR Authority

- **§ 121.5 (a). Public records and archives.**

State Archival Agency Designated. – **The Department of Cultural Resources shall be the official archival agency of the State of North Carolina** with authority as provided throughout this Chapter and Chapter 132 of the General Statutes of North Carolina **in relation to the public records of the State, counties, municipalities, and other subdivisions of government.**

- **§ 125-11.7. State Library designated the official depository for all State publications.**

The State Library shall be the official, complete, and permanent depository for all State publications, and shall receive five copies of all State publications in addition to the copies required for the depository system: Provided, however....

- **§ 132-1. "Public records" defined.**

- (a) **" Public record" or "public records" shall mean all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions.** Agency of North Carolina government or its subdivisions shall mean and include every public office, public officer or official (State or local, elected or appointed), institution....



DCR's History With Archive-It

- Pilot partner beginning in 2005
- Subscriber since May 2006
 - Archiving North Carolina State Agencies and Boards & Commissions
 - Static Web pages & anything connected to a static Web page that the harvester can grab
 - Including blogs
 - Schedule varies
- Purchased older harvested materials in 2009
 - Internet Archive harvested NC government sites harvested since 1996

Governor's Office

- Obama in DC; Perdue in NC
 - Interest in e-governance, Web 2.0, & social networking sites (SNS) skyrocket.
 - Outreach opportunity; communication forum
- Documenting social change, social reaction to important events, and a cultural phenomena



A Collaboration Is Born

- Governor's Office approached DCR to discuss SN (how to encourage usage and to comply with public records law).
- Gov. Office, ITS, and DCR created best practices for SN usage for agency staff.
- DCR looked at options to archive SN content and chose AI to help with archiving.



"Terrassa Castellers,"
by Ryan Opaz

Roles & Responsibilities

- Internet Archive/Archive-It
 - Harvesting service
 - Identifies limitations of harvester/crawler
 - Provides scoping help
- State Library of NC & NC State Archives
 - Scope crawls
 - Identify agency “seed” sites
 - Schedule crawls
 - Manage budget
- NC Governor’s Office
 - Provides agency user perspective
 - Provides help in identifying agency social networking sites



Is Social Networking Content Considered a public record?

- We view websites and the various documents posted on them (.doc, .pdf, etc.) as publications.
- Social Networking sites are being used to conduct state business (marketing, information push, etc.) and, therefore, meet the statutory definition of public record.
- We decided not to take the “wait and see” approach.



<http://media.ebaumsworld.com>

A Long Story Made Short





What Are We Harvesting?

- In 2009, we tested harvesting the Governor's facebook, flickr, twitter, and YouTube accounts.
- In 2010, we expanded to include all state agency facebook, flickr, twitter, and YouTube accounts were made aware of.
 - As of spring 2010, these sites are harvestable and immediately available* through our Web archive



*YouTube is currently harvested, but major aspects are not currently renderable.



What Else Are We Interested In Harvesting?

- Harvested a LinkedIn account, but haven't done a full rollout yet.
- Haven't tested any others yet.
- May also be interested in harvesting:



Scribd



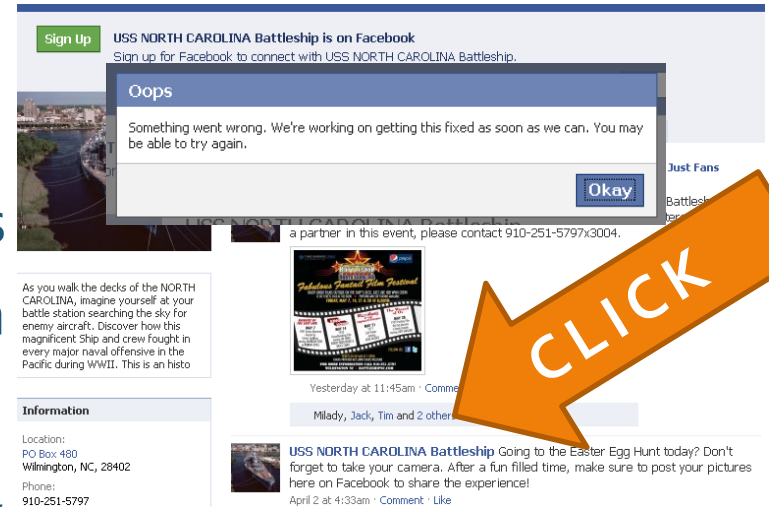


- Easy to archive
 - Captures photostream
 - Captures comments
- Renders very well
 - Renders photostream
 - Renders comments
 - Some icon-type images (logos, etc.) don't render





- Not quite so easy to archive
 - Captures wall information (including comments)
 - Captures images & fans/friends
 - Captures other tab information
- Renders pretty well
 - Basic content renders perfectly
 - Links to more info require turning off javascript in your browser and/or the use of “proxy mode” viewing





- Easy to archive
 - Captures posts from account specified
 - Captures links (if settings adjusted)
- Renders very well
 - Renders posts
 - If set up for it, will render linked content
 - Special scoping rule necessary to capture content behind 'More Info' link





- Not quite so easy to archive
 - Captures channel (all videos)
 - Captures images & fans/friends
 - Captures other tab information
- Renders pretty well
 - Basic content renders perfectly
 - Links to more info require turning off javascript in your browser and/or the use of “proxy mode” viewing





Partner Challenges

- Network configurations
- Scoping the harvests
- Privacy settings/logins
- Capturing new/other social networking sites
- Maintaining list of active NC social networking site users



<http://rockclimbingshoesandgear.com/>



jennifer.ricker@ncdcr.gov



Much credit for this presentation goes to Amy Rudersdorf & Kelly Eubank.

