

# Welcome!

The webinar will begin at 2:00 Eastern/11:00 Pacific

### Audio broadcast—volume

Today's audio is streaming over your computer's speakers or headphones.

You can adjust the volume on the Audio broadcast box:



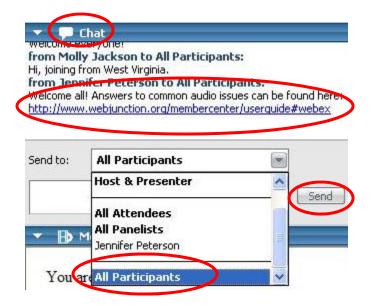
If you hear an echo or are disconnected from the audio:

Select Leave Audio Broadcast or click X to close box(es). Then rejoin.



# **Chat Option**

- If you have technical difficulties, please chat privately with "WJ Support."
- Please post your questions to "All Participants."
- Please paste to chat your links to relevant resources.





**Step 1:** Type in the dialog box.

**Step 2:** Select **recipient(s)** from dropdown option.

Step 3: Click "Send" button.



# If needed, access audio by telephone

If you cannot or do not wish to listen to today's audio via your computer, please follow these steps to join by telephone.

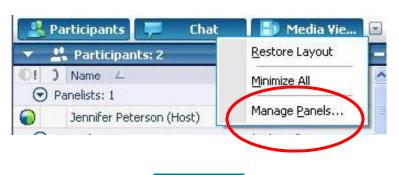


Click on **Request** under the participants list.

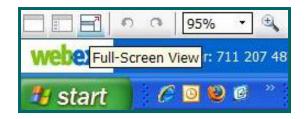
You will be provided with a toll-free number. Be sure to enter # following each of the operator's requests for **Access Code** and **Attendee ID**.

# Customize your experience

- Panels can be minimized or expanded
- Hover over edge of panels to drag and resize
- Size can be adjusted or presentation can be viewed "full screen"
- Use this menu at bottom right to return to panel view



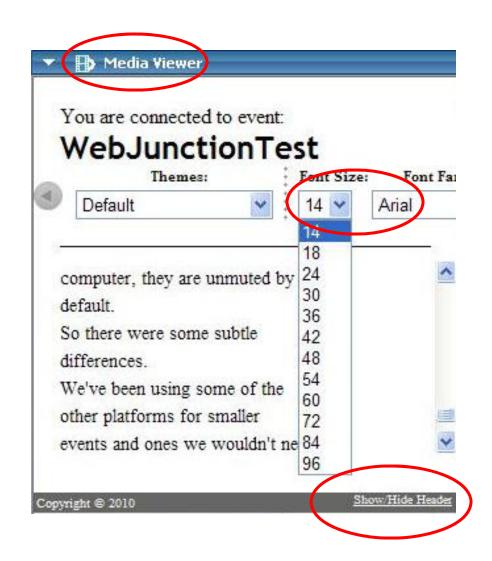








- Access via the Media
   Viewer on your Panel
   options.
- You may adjust the font size and type for captions.
- Then select
   Show/Hide Header in the bottom right corner.





Remember to chat privately with **WJ Support** if you need technical assistance.

Other Technical problems?

**Contact WebEx support** 

**Event Number: 716 004 413** 

Phone: 1-866-229-3239

Email: support@webex.com



# Today's Producers



Jennifer Peterson

Host/Moderator

WebJunction
Community Manager



Kendra Morgan 'WJ Support' WebJunction Program Manager



# Today's Panelists

Sarah McHugh, Statewide Projects Librarian, Montana State Library



Amy Rudersdorf,
Director of the Digital
Information
Management
Program, State
Library of North
Carolina



Taylor Surface, Senior Product Manager, Digital Collection Services, OCLC





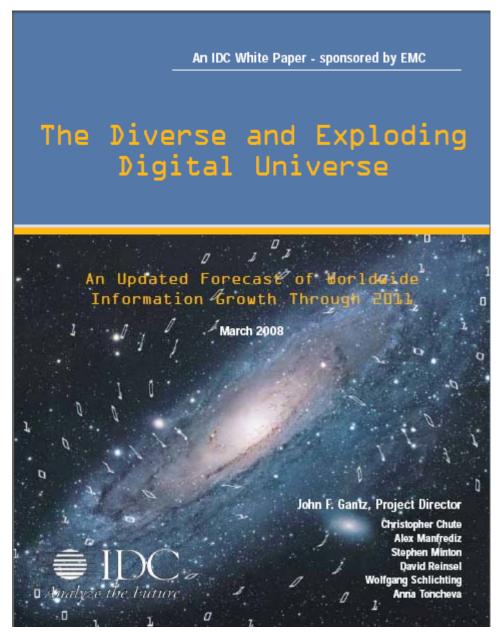
# Digital Preservation: Basics, Best Practices, Choices for Libraries

WebJunction
August 2010

# **Taylor Surface**

OCLC Digital Collection Services
Senior Product Manager





http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf

## Agenda



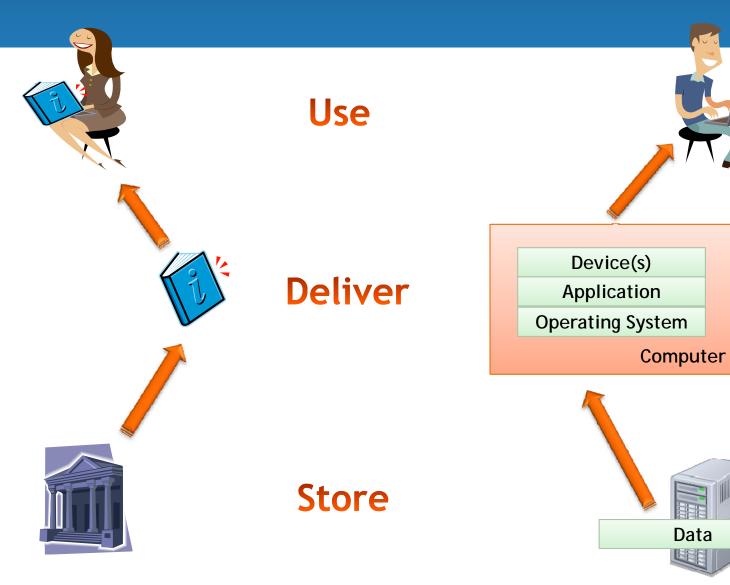
• What is digital preservation?

• Who provides this service?

How can you do this in your library?

# Digital preservation / changing needs





### Operational preservation questions



How will you store & maintain these for 5, 10, 50 years?

• What is in your collection?

When will you know to take preservation actions?

• Where have you put them?

### Defining the frontier - 1994



#### Preserving Digital Information

Report of the Task Force on Archiving of Digital Information

commissioned
by
The Commission on Preservation and Access
and
The Research Libraries Group

May 1, 199

- Major Findings
- Responsibility: creators, publishers, owners
- Long-term requires deep infrastructure
- Need organizations capable of
  - Storing
  - Migrating
  - Accessing
- Need certification for trust
- Need legal rights and capacity for rescue

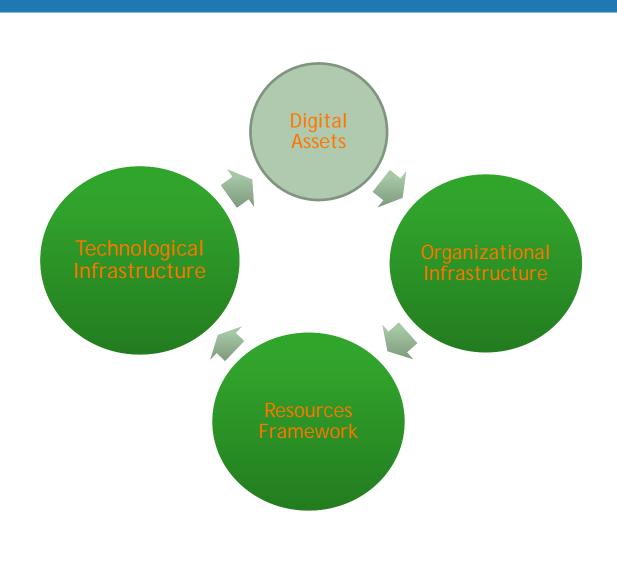
### Growing family of standards & practices



- OAIS our vocabulary for structure, roles, terminology (ISO 14721) ... when did you start using the term "ingest"?
- Trusted Digital Repositories (CRL implementing)
- Enabling preservation actions
  - Registries of data formats: Pronom, GDFR now UDFR
  - Data format verification tools: JHOVE, DROID
- Metadata elements PREMIS v2.0, April 2008

## Digital Preservation Framework





### Where are you applying your resources?



**Digital Program Costs - JISC study** 

Estimates from UK National Data Centres

Acquisition & Ingest ~42%

Archival Storage & Preservation

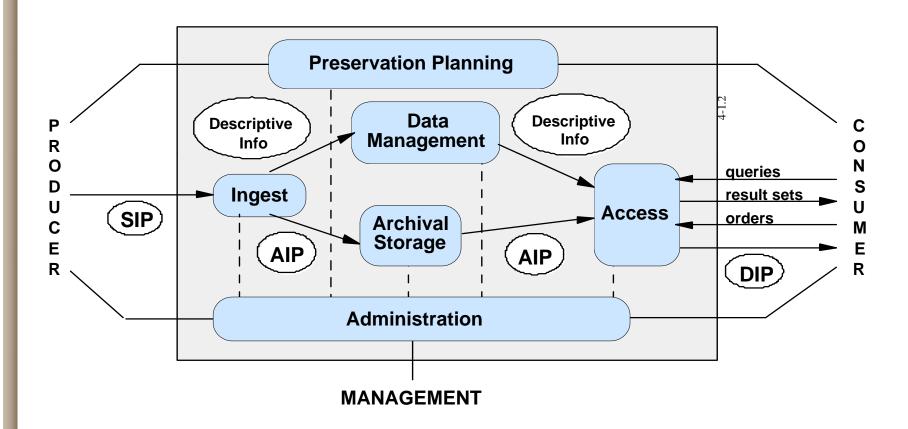
~23%

Access ~35%



### **OAIS Functional Model**





Key concept: Designated community - who are you serving?

# Preserving common collections



- National library of the Netherlands, E-Depot (journals)
- LOCKSS (journals)
- Portico (journals)
- CLOCKSS (journals)
- Hathi Trust (books)

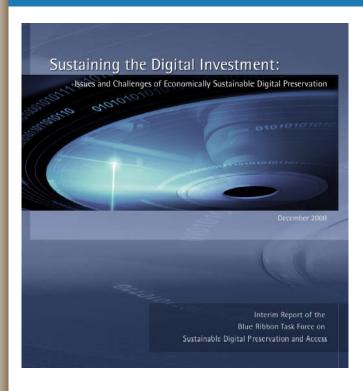
### Preserving unique, local collections



- DuraCloud (software as a service)
- Ex Libris Rosetta (software)
- Internet Archive (services)
- MetaArchive Cooperative (LOCKSS private networks)
- OCLC Digital Archive (service)
- Tessella (software, services)
- "Open Source" (software)

# Blue Ribbon Task Force on Sustainable Digital Preservation and Access





Preliminary lessons learned

- "Sell" outcomes
- Avoid discounting benefits
- It's hard to separate costs
- Diversify funding
- Non-monetary incentives
- Consider many options when choosing economic model

http://brtf.sdsc.edu/biblio/BRTF\_Interim\_Report.pdf



# Digital Preservation: Basics, Best Practices, Choices for Libraries

WebJunction
August 2010

# **Taylor Surface**

taylor\_surface@oclc.org +1 614 761 5145

# www.oclc.org/digitalarchive





# Post your questions to chat



# Providing Access

Digitization and Long Term
Preservation Efforts in
Montana



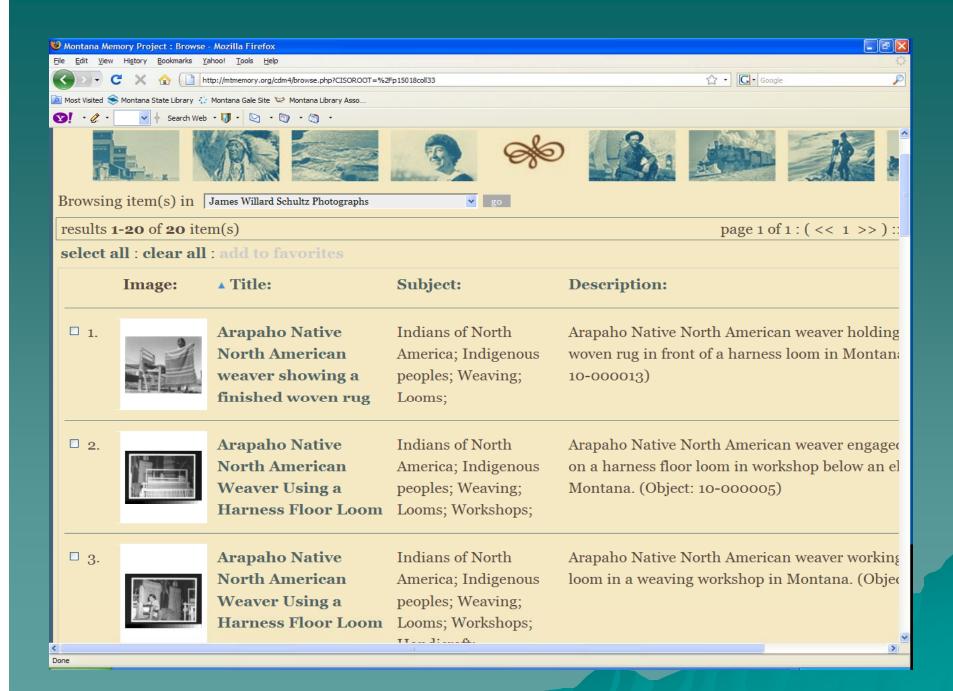
navigation bar.

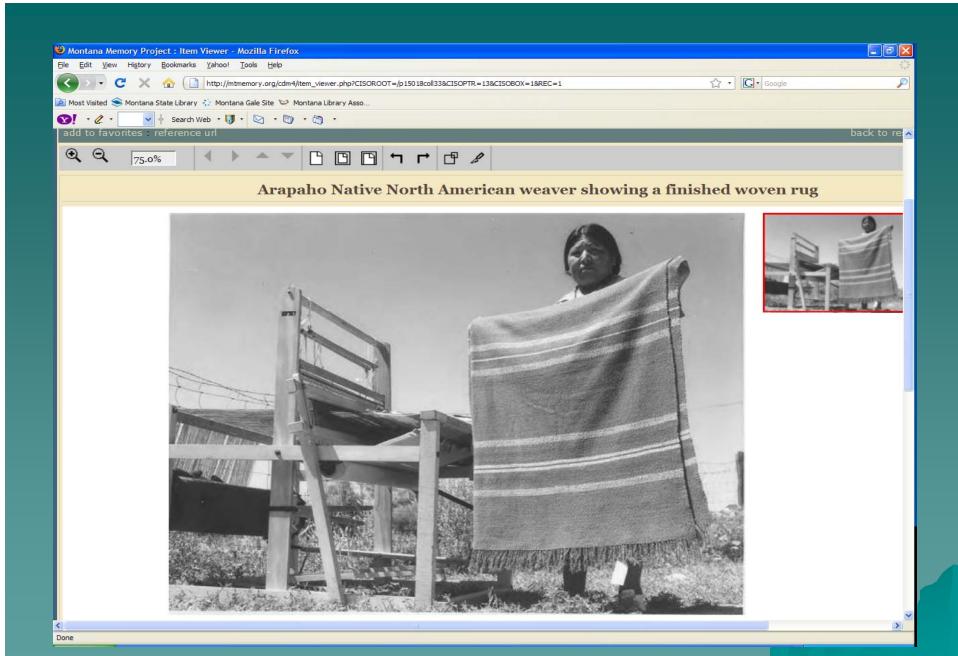
**Browse** 

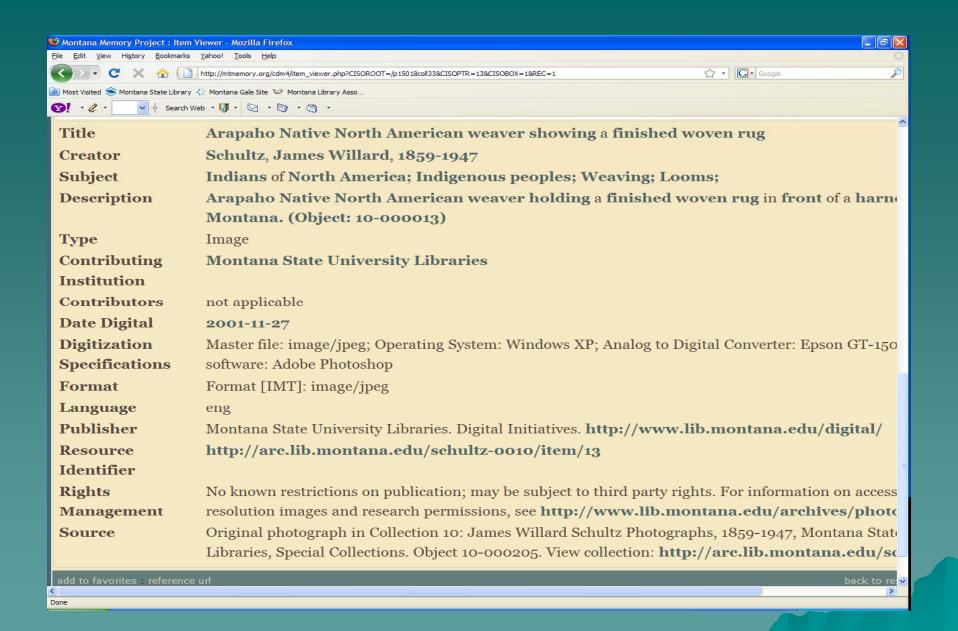
go >>

Browse through the collection without formulating a specific search.

Done







# Achieving Understanding

Facilitated discussions among major stakeholders

Identifying digitization efforts for access versus digitization efforts for long term preservation/access

Agreement that statewide efforts should include both

Discussion of options for both

Decision on long term preservation tool

# **Building Documentation**

- Statewide Digitization Registry
- Montana Memory Project Planning Documents
- IMLS Connecting2Collections
   Grant/Montana Survey















Talking Book Library

For the Public

For State Employees

For Librarians

About MSL

MSL Home > For Librarians > Library Directory > Edit Path > Main Menu > Digitization > Digitization Menu

Jobs

- ☑ Library Directory
- Browse Path
- Edit Path
- Tutorial
- Download
- Print (pdf)
- Montana Shared Catalog
- For All Librarians
- Tor Public Librarians
- For School Librarians ☑
- For Trustees
- ☐ Grants
- Online Publications
- Statewide Projects
- ☑ Continuing Education and Certification
- ☑ Wired-MT Listserv

#### **Event Calendar**

#### Search Catalog

Words or Phrase >

Search

#### Search Website



#### Search Other Libraries



Enter title, subject or author

#### iGoogle Gadget

This gadget adds a search box to your iGoogle portal allowing you to search the Montana State Library Catalog.



### **Library Directory**

#### Montana Historical Society Research Center Digitization Menu

Montana Memory Project Project Planning Document Add/Edit

To propose a new collection to the Montana Memory Project Steering Committee please complete the Montana Memory Project - Project Planning Document. This will be reviewed by the MMP Steering Committee to evaluate and plan for your training and support needs.

Information Only Data and Collections Add/Edit

To document digitization projects not currently planned for addition to the Montana Memory Project please complete the Information Only Data and Collections. This will help the State Library track digitization projects that exist around the state and serve as a statewide registry of digital projects. This information used to evaluate opportunities for new statewide projects, support future grant Project Planning Documents and will make the State Library aware of opportunities for collaborative projects.

Please contact the Statewide Projects Librarian for questions regarding the Digitization process.















Talking Book Library

For the Public

For State Employees

For Librarians

MSL Home > For Librarians > Library Directory > Edit Path > Main Menu > Digitization > Projects >

About MSL

Jobs

Home

- Library Directory
- Browse Path
- Edit Path
- Tutorial
- Download
- Print (pdf)
- Montana Shared Catalog
- Tor All Librarians
- For Public Librarians
- For School Librarians 

  ■
- ☐ Grants
- Online Publications
- Statewide Projects
- ☑ Continuing Education and Certification
- ☑ Wired-MT Listserv

#### **Event Calendar**

Search Catalog

Words or Phrase

Search

Search Website

Search

#### Search Other Libraries

Enter title, subject or author

Search

#### iGoogle Gadget

This gadget adds a search box to your iGoogle portal allowing you to search the Montana State Library Catalog.

Add to Google

Montana Memory Project Planning Document Menu

#### **Library Directory**

Montana Historical Society Research Center Montana Memory Project Planning Document Menu

Digitization Environment	Edit	Approved
Digitization Hardware	Menu	Complete
Digital Collection Planning	Menu	Complete

Print Proposal Menu
Project Planning Document Status Approved

Complete

Your Montana Memory Project Planning Document has been approved and is available for editing. Approved collections will also be available for editing.

Return to Digitization Menu















Talking Book Library

For the Public

For State Employees For Librarians

About MSL

Jobs

Home

■ Library Directory
■ Browse Path
■ Edit Path
□ Tutorial
Download Download
Print (pdf)
Montana Shared Catalog
■ For All Librarians
For Public Librarians
For School Librarians
□ For Trustees

E	Grant	s
Ξ	Online	<ul> <li>Publications</li> </ul>

Statewide Projects
 Continuing Education and
 Certification

■ Wired-MT Listserv

#### **Event Calendar**

Search Catalog

Words or Phrase

Search Website

Search

#### Search Other Libraries

Enter title, subject or author

Search

#### iGoogle Gadget

This gadget adds a search box to your iGoogle portal allowing you to search the Montana State Library Catalog.

Add to Google

<u>MSL Home > For Librarians > Library Directory > Edit Path > Main Menu > Digitization > Projects > Digitization Environment</u>

#### **Library Directory**

Montana Historical Society Research Center Add Digitization Environment Information

#### Contact Person

Please select the name of the employee at your library who is responsible for maintaining and reporting your digitization information. If he or she is not listed here, return to "<a href="Employee List" and add the employee to the directory.">Employee List</a>" and add the employee to the directory.

What computer operating system do you use?	
MS Windows XP	
If other, please explain	
What type of Internet connectivity does your organization have?	
High Speed 💌	
ow many of your staff will need CONTENTdm software training? The Montana Memory Project uses the CONTENTdm software ackage.)	
0	
Digitization is being handled on a for-fee basis by Mansfield Library, UM <u>Missoula</u> , contact person Steve <u>McCann</u> .	

Submit

















Talking Book Library

For the Public

For State Employees

For Librarians About MSL Jobs

Edit / Delete

Edit / Delete

Home

□ Library Directory

- Browse Path
- Edit Path
- □ Tutorial
- Download Print (pdf)
- Montana Shared Catalog
- Tor All Librarians
- Tor Public Librarians
- □ For School Librarians □
- For Trustees
- ☐ Grants
- Online Publications
- Statewide Projects
- ☑ Continuing Education and Certification
- ☑ Wired-MT Listserv

#### **Event Calendar**

#### Search Catalog

Words or Phrase >

Search

#### Search Website

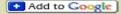
Search

#### Search Other Libraries

Enter title, subject or author Search

#### iGoogle Gadget

This gadget adds a search box to your iGoogle portal allowing you to search the Montana State Library Catalog.



MSL Home > For Librarians > Library Directory > Edit Path > Main Menu > Digitization > Projects > Digitization Hardware > Digitization Hardware Menu

#### **Library Directory**

#### **Montana Historical Society Research Center** Digitization Hardware Menu

Hardware Needed	What kinds of digitization hardware do you need to complete your project?		
	Туре		
	Click here to declare that you do not know what will be needed		
	Click here to add a needed digitization hardware item		
	Please note that you must either declare that you do not know what will be needed or add a needed digitization hardware item before you can submit your final Project Planning Document.		
Hardware Owned	What types of digitization hardware do you currently own?		
	Type Make Model Status		
	Click here to add currently owned digitization hardware item		

n/a

1640 XL

n/a

Epson

Return to Montana Memory Project Planning Document Menu

Digital Camera

Scanner



Complete

Complete















Talking Book Library

For the Public

For State Employees

For Librarians

About MSL

Jobs

Home

- Library Directory
- Browse Path
- Edit Path
- Tutorial
- Download
- Print (pdf)
- Montana Shared Catalog
- ☐ For All Librarians
- Tor Public Librarians
- For School Librarians 12
- For Trustees
- ☐ Grants
- Online Publications
- Statewide Projects
- Continuing Education and Certification
- Wired-MT Listserv

#### **Event Calendar**

#### Search Catalog



Words or Phrase 💌



#### Search Website



#### Search Other Libraries



Enter title, subject or author

Search

#### iGoogle Gadget

This gadget adds a search box to your iGoogle portal allowing you to search the Montana State Library Catalog.



 $\underline{\mathsf{MSL}}$  Home >  $\underline{\mathsf{For}}$  Librarians >  $\underline{\mathsf{Library}}$  Directory >  $\underline{\mathsf{Edit}}$  Path >  $\underline{\mathsf{Main}}$  Menu >  $\underline{\mathsf{Diqitization}}$  >  $\underline{\mathsf{Projects}}$  >  $\underline{\mathsf{Collections}}$  > Digital Collection Plan Menu

#### **Library Directory**

#### Montana Historical Society Research Center Digital Collection Plan Menu

#### **Montana County Histories**

		Status
Title	Edit	Complete
Partner Libraries	Menu	Optional
General Information	<u>Edit</u>	Complete
Collection Materials	Edit	Complete
Collection Formats	Edit	Complete
Metadata Programs	Edit	Complete
Media Storage	Edit	Complete
Time Periods	Menu	Complete
Topics	Menu	Complete
Timeframe	Menu	Complete
Sample File Upload	<u>Info</u>	
Metadata Sample	Edit	Complete

#### Print Collection

Overall Collection Status: Approved

Return to Digital Collection Planning Menu

# Making it possible for all to play

Partnerships between major stakeholders in Montana

Partnerships between local library and local museum/archive

Managed as a statewide project open to all

Governing committee that assists the project

Shared technical and training support

# Next Challenges

- Educating efforts
- Identifying and prioritizing valuable and at risk local collections
- Defining the scope of the Montana Memory Project
- Expanding funding sources and partnerships
- Increasing training and support

- Montana Memory Project: <a href="http://mtmemory.org/">http://mtmemory.org/</a>
- Montana State Library: <a href="http://msl.mt.gov/">http://msl.mt.gov/</a>

Sarah McHugh Statewide Projects Librarian Montana State Library

Email: <a href="mailto:samchugh@mt.gov">samchugh@mt.gov</a>

Phone: 406-444-9816



## Post your questions to chat



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

# Digitization & Preservation Symposium August 25, 2010





#### PART 1: The Set Up - 1987

§ 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, the State Library shall receive only five copies of any State publication offered for sale by a State agency at a price at least high enough to recover production costs: Provided, further, the State Library, notwithstanding the definition of "State publication" contained in this Article, shall have authority to exclude from required deposit in the State Library any items or materials which it finds are not appropriate for deposit. (1987, C. 771, S. 2.)





#### PART 1: The Set Up - 1987

§ 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, the State Library shall receive only five copies of any State publication offered for sale by a State agency at a price at least high enough to recover production costs: Provided, further, the State Library, notwithstanding the definition of "State publication" contained in this Article, shall have authority to exclude from required deposit in the State Library any items or materials which it finds are not appropriate for deposit. (1987, C. 771, S. 2.)



## The Set Up: 2002-2007

- 2002: The problem of digital preservation is identified
- 2004ish:
  - 1. Federal grants fund two Web harvesting projects
    - Web Archives Workbench\* (OCLC) and based on the "Arizona Model" - LOC
    - Capturing Electronic Publications (State Library of Illinois and UIUC) – IMLS
  - 2. Internet Archive
    - Pilots their subscription-based Web harvesting service called "Archive-It"
  - State Library and State Archives of NC joins all pilot projects
- 2007: State budget includes a line item for digital preservation program at the State Library
  - Digital Information Management Program (DIMP) is born
  - Funds participation in ongoing Archive-It service with sister division, North Carolina State Archives



## The Set Up: 2008-2009

- 2008: Society begins to recognize digital preservation as an issue; Social networking "takes off"
  - Obama in DC; Perdue in NC
    - Interest in e-governance, Web 2.0, & social networking sites (SNS) skyrockets.
    - Outreach opportunity; communication forum
  - Documenting social change, social reaction to important events, and a cultural phenomena
- 2009: NC selects Internet Archive's Archive-It
   (AI), collocates all harvests from 1996 on, begins
   considering social networking issues



## A Long Story Made Short





#### Hold On. Two Very Loose Definitions



"Web harvesting involves using web crawling software to automatically copy and collect internet content such as website pages."

- National Library of New Zealand

**Social networking sites** display your network (LinkedIn), tell you who you should network with (twitter), allow you to have conversations (facebook), or build and nurture a network (BrazenCareerist).

 From BrazenCareerist.com, "a social network for Generation Y professionals"





## What Are We Harvesting?

- 2005-2009.
  - North Carolina State Agencies and Boards & Commissions
    - Static Web pages & anything connected to a static Web page that the harvester can grab
    - Schedule varies
- 2009.
  - All of the above, plus the Governor's social networking sites
- 2010.
  - All of the above, plus all (known) North Carolina State Agency social networking sites



#### Other Options?

- Cheaper solutions, but IOHO:
  - Not as stable as Al
  - Not as tested/reviewed/open as Al
- Other solutions:
  - Web Archiving Services (WAS)
    - Same crawler as AI (Heritrix)
  - ContextMiner (social networking specific)
  - OCLC Web Harvester
- Host our own harvester/Ignore
  - Not an option













- Archive-It developed functionality based on (our) needs assessment
  - 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.



- Trial & Error, Quality Review, Collaboration
  - AI, Department of Cultural Resources, NC Governor's Office
- Documenting detailed changes for partners



De Grazia, Victoria. "The secretarial pool at the Perugina Candy Works, 1928." *How Fascism Ruled Women: Italy, 1922-1945*. Berkeley, CA: University of California Press, 1992.



- Al testing/interested in testing LinkedIn, Slideshare, Vimeo, and MySpace
  - Public Record? Relevance?
- WAS partners looking at delicious
- Will continue to beta new social networking site platforms as they emerge













## PART 3: A Tour of What Works & Doesn't













- Comments & photostreams are harvested and presented easily
  - The rendered page looks & acts remarkably similarly to the original
- How big is it?
  - The larger "seeds," such as the Governor's site return about 12,000 documents.









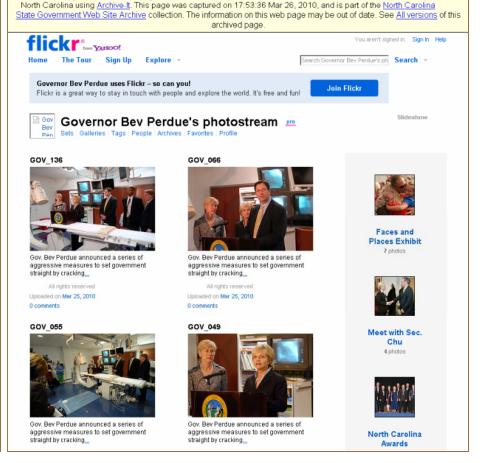






You are viewing an archived web page, collected at the request of North Carolina State Archives and State Library of















- Content can be captured and rendered back
  - Javascript galore
    - A growing issue with social networking sites (and Web in general)
    - Doesn't always render properly
    - Fix in new release
- NC Gov's page returns about 1,800 documents









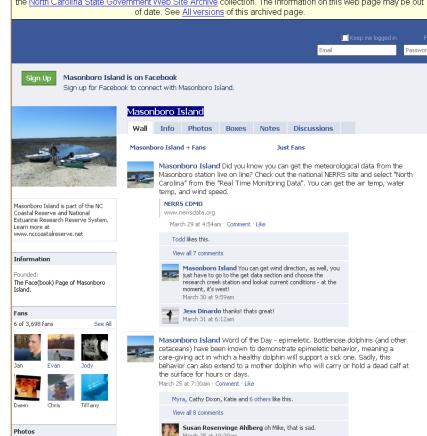


#### Live





**Archived** 

















#### **Archived Site**

Sign Up

#### USS NORTH CAROLINA Battleship is on Facebook

Sign up for Facebook to connect with USS NORTH CAROLINA Battleship.



#### Information

Location:

PO Box 480

Wilmington, NC, 28402

Phone:

910-251-5797



USS NORTH CAROLINA Battleship Going to the Easter Egg Hunt today? Don't forget to take your camera. After a fun filled time, make sure to post your pictures here on Facebook to share the experience!

April 2 at 4:33am / Comment / Like

Milady, Jack, Tim and 2 others like this











- Capturing and rendering tweets is easy
  - Links and retweets pose a bit more of a problem
  - We choose to include links (one hop off) and then manually assess the crawl results so that we don't harvest inappropriate sites in *future* crawls.
  - You only see the retweet, not the original
- VisitNCWine twitter crawl returns about 120 documents without the hop and 2000 with it







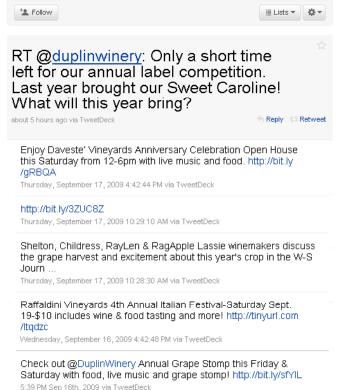




#### **Archived**

#### Live

#### **VisitNCWine**

















- Harvesting, but rendering problems abound
- Specifically, videos don't play unless you use a proxy site
  - Due to code running the page
  - CDL, the Library of Congress, Harvard, IA, and "other NDIIPP partners" will be meeting with YouTube representatives at the end of July to discuss.





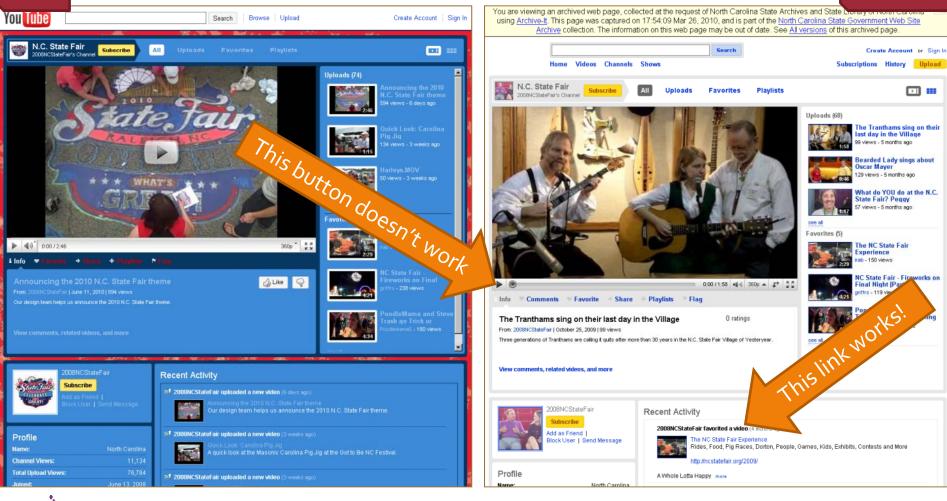














## Challenges remain

- "proxy mode" viewing
- Scoping the harvests
- Privacy settings/logins
- Capturing new/other social networking sites
- Maintaining list of active NC social networking site users
- Just when you think you've got it figured out, Mark Zuckerberg\* goes and changes everything on you





## amy.rudersdorf@ncdcr.gov



Much credit for this presentation must be given to Jennifer Ricker & Kelly Eubank at the NC Department of Cultural Resources, and to Molly Bragg, formally of the Internet Archive, but currently enjoying the wonder that is Chapel Hill, NC.





## Post your questions to chat



# Stay Involved

On WebJunction

webjunction.org/processing-andpreservation

Crossroads (monthly newsletter) webjunction.org/crossroads

Upcoming Events and Archives webjunction.org/events/webinars