

Preservation Systems

NDIIPP All Partners Meeting

December 2011

Carol Kussmann

Preserving Electronic Records

- ▶ What options are available for preservation?
 - Content specific e.g. Web Archiving
 - Collaborative redundant storage – LOCKSS
 - Consortium repository
 - Contract service repository
 - Open source and/or off the shelf solution
 - Develop your own
- ▶ What features or benefits best fit your current and/or future environment?

• Explored web archiving and micro-services based contract service repositories

Web Archiving

Web Archiving

▶ Questions to Ask:

- What information/content do you hope to capture?
What is your purpose?
- Is this the best way to capture the information for long-term use?

▶ Things to think about:

- Required investments
 - Time, staffing, costs
- Technical details
 - Skills, equipment, crawl settings
- Administrative details
 - Long-term access, public/private

Web Archiving Service

- ▶ Create a collection
 - Select seeds
 - Set scope – parameters
 - Choose length of time and frequency for crawl
 - Add metadata

Create Site

Capture Settings | Scheduling | Descriptive Data

* Required field

* Site Name:

i * Seed URLs:
Ex.: http://www.example.com

i Scope: Host site

Capture Linked Pages: No Yes

Max. Time: Brief Capture (1 hour)

Web Archiving Service

- ▶ Review and evaluate captures
 - Make changes as necessary
 - Don't want too much or too little

3. View captures



The screenshot displays two capture records within a dashed border. Each record consists of a thumbnail image on the left and a text block on the right. The top record's thumbnail shows a collage of images, and its text block contains the following information:

- Title: http://www.usdoj.gov/inspector/ethics/management_programs.ppt
- Captured: Sun May 27 10:55:35 -0700 2007
- URL: http://www.usdoj.gov/inspector/ethics/management_programs.ppt
- Show Detailed Record

The bottom record's thumbnail shows a document and a CD, and its text block contains the following information:

- Title: http://www.usdoj.gov/inspector/ethics/management_programs.ppt
- Captured: Sun May 27 10:55:35 -0700 2007
- URL: http://www.usdoj.gov/inspector/ethics/management_programs.ppt
- Show Detailed Record

Search, browse and display archived content.
Review capture reports.

Web Archiving Service

- ▶ Entire process documented
- ▶ List of things to think about
- ▶ Earlier version compared with Archive IT

1. Create site



Provide URLs, capture settings and metadata for the sites you plan to capture.

2. Capture site



| SITE INFO | STATUS | ACTIONS |
|---|--|-------------------------|
| Bureau of Reclamation Mid-Pacific Region The Mid-Pacific Region is one of five Reclamation... | Running Started 1 minute ago 9 files captured Stop this job | Capture |
| California Department of Water Resources DWR operates and maintains the State | Finished Last completed 22 days ago | Capture |
| California State Water Resources Cont... | Finished Last completed 22 days ago | Capture |

Start capturing the site content.
WAS will email you when complete.

3. View captures



Title: http://www.water.ca.gov/graphics/interim/watermanagement_programs.pdf
19.2 KB | [img/png](#)
Captured: Sun May 27 10:33:10 -0700 2007
URL: http://www.water.ca.gov/graphics/interim/watermanagement_programs.pdf
[Show Detailed Record](#)

Title: http://www.water.ca.gov/graphics/interim/watermanagement_planning.pdf
19.2 KB | [img/png](#)
Captured: Sun May 27 10:33:10 -0700 2007
URL: http://www.water.ca.gov/graphics/interim/watermanagement_planning.pdf
[Show Detailed Record](#)

Search, browse and display archived content.
Review capture reports.

4. Public access



http://www.water.ca.gov/graphics/interim/watermanagement_planning.pdf
Public Access

California's DROUGHT

Precipitation
In California, most precipitation falls from winter and spring through fall, winter and into spring.

- Increased average precipitation for the last 7 years by from about 10% of average.
- Some parts of Southern California

Configure your project for public access.
Set the description, URL or banner images.

Contract Service Repositories

Contract Service Repositories

- ▶ Systems vary: wanted hands on experience with preservation systems
- ▶ Merritt
 - CDL wanted to know how non-university customers would use repository
- ▶ Safety Deposit Box
 - Explored multi-state tenancy
 - Tessella was interested in learning about the needs of state governments

Repository Testing

| Functions | Records |
|----------------|---------------------------------|
| Ingest | Public |
| Storage | Permanent/non-changing |
| Management | Various formats / content types |
| Preservation | Large and small quantities |
| Authentication | |
| Export | |
| | |



Archive Organization

| Merritt | Safety Deposit Box |
|---|--|
| Collection: Group of objects | Collection: One or more deliverable units |
| Object: One or multiple files | Deliverable Unit: One or multiple files |
| View 10 most recent entries to a collection | Collections and deliverable units viewed within a folder structure |

Collection: Demo Merritt collection
 Merritt > Collection: Demo Merritt collection

| Primary Id | Object info |
|--------------------------------------|-------------------------|
| ark:/99999/fk4ww7r1s | mn. hf0002 |
| ark:/99999/fk4ht2wnn | Margaret Low. DPR stats |
| ark:/99999/fk4x06fcn | Margaret Low. DPR stats |
| ark:/99999/fk4m04cp5 | |

- ▶ Large File Test
- ▶ Original Word Documents
- ▶ 2011 Bills 1 to 10
 - ▶ 2011 HF0001
 - ▶ 2011 HF0002
 - MN-201100HF0002-0
 - MN-201100HF0002-1
 - ▶ MN-201100HF0002-2
 - ▶ 2011 HF0003
 - ▶ 2011 HF0004
 - ▶ Transformations upon Ingest
 - ▶ 2007-2008 Bills part 2

Ingest

| Merritt | Safety Deposit Box |
|-------------------------------|---|
| Upload objects | Upload collections and/or deliverable units |
| Manual or automatic processes | Manual or automatic processes |
| Single or batch upload | Single or batch upload |

- ▶ Submission Information Packets (SIPS)
- ▶ Think about:
 - Level of technical skills required
 - How the ingest process fits into current workflows

Metadata

| Merritt | Safety Deposit Box |
|---|--|
| Creates its own system metadata. | Creates its own system metadata. |
| Allows users to add four metadata elements (author, title, date, ID). | Allows users to add any amount of metadata, conforming to any schema |
| | Metadata can be added after ingest to any entity. |

- ▶ Metadata becomes searchable
- ▶ Not requiring metadata upon ingest lowers the barrier for participation.
- ▶ Accepting any and all metadata adds flexibility to a system and increases ability to find relevant content.

Search and Navigation

| Merritt | Safety Deposit Box |
|---|--|
| System and provided metadata. (Author, title, date, or ID) | System and provided metadata. (Anything) |
| | Keyword searching on indexed materials – pdf, word, xml... |
| | Narrow search using metadata fields filters. |
| Searches a single collection. | Searches the entire archive. |
| Navigate through a single collection 10 files at a time in order of ingest. | Navigate through entire archive using a folder structure. |

Object Lookup

Search object author, title, date or identifier.

Search

 Only search top level records

Retrieval / Export

| Merritt | Safety Deposit Box |
|--|--|
| Single file. | Single file. |
| Single version of an object. | Single deliverable unit. |
| Complete object (all versions). | [Entire collection.] |
| | [Entire archive.] |
| Uses fixity to verify export packages. | Uses fixity to verify export packages. |

- ▶ Must be easy to get materials out.
- ▶ Batch access is important for many uses – disaster recovery, public use.
- ▶ Flexibility and options for export increase system value.

Manage

| | Merritt | Safety Deposit Box |
|------------------|--|--|
| Reports | Not directly available | Standard and Custom* |
| Edits / Versions | If changes are made to single object in a group submission; must resubmit the ENTIRE group | Upload a new version to an existing collection / deliverable unit. |
| Storage | UC3 systems | Anything* |
| Access | Persistent URLs with login permissions | Login permissions. |
| | | Designed to interface with other systems to provide public access. |
| Documentation | Background information and user guides are available. | User guides and system information available to users.* |
| | | Specific user guides created for project* |

Preservation

| | |
|--------------------------|---|
| Merritt | Safety Deposit Box |
| Fixity checks regularly. | Fixity checks on set schedule. |
| | Ability to migrate from one format to another. |
| | Technical registry for holding preservation policy information for preservation planning. |

JPG transformation test

Original (Active)



DSC_0425.JPG



Manifestation 2 (Active)



DSC_0425.pdf

User Interface

| | Merritt | Safety Deposit Box |
|------------------|---|--|
| Home Page: | Screen displays the Primary Id and provided metadata (if any) | Folder structure of archive, when opened shows individual deliverable units and files. |
| | View the 10 most recent files. | Controlled by user with folder structure. |
| Object/DU Level: | Metadata and version information. | Technical properties, title, format, fixity, thumbnail. |

Merritt User Interface

[Collection home](#)[Add object](#)[Change collection ▾](#)

Collection: Demo Merritt collection

Merritt > Collection: Demo Merritt collection

| Primary Id | Object info | — Previous 1 2 3 4 5 6 7 8 9 ... 16 17 Next — |
|--------------------------------------|--|---|
| ark:/99999/fk4ww7r1s | mn. hf0002 | |
| ark:/99999/fk4ht2wnn | Margaret Low. DPR stats | |
| ark:/99999/fk4x06fcn | Margaret Low. DPR stats | |
| ark:/99999/fk4m04cp5 | | |
| ark:/99999/fk4vh5w36 | MerrittPrez2 | |
| ark:/99999/fk4086cjw | merrittPrez | |
| ark:/99999/fk43x8dxq | | |
| ark:/99999/fk48d03hm | Computer Science Division. A GREMLIN Tutorial for the SUN Workstation | |
| ark:/99999/fk4hx1kx6 | Morrish, J. Elmer, 1886-1957, banker. Correspondence from J. E. Morrish to Pacific Gas and Electric Company; Redwood City Public Library | |
| ark:/99999/fk4d50v6v | Gelbart, Nina Rattner. The king's midwife : a history and mystery of Madame du Coudray | |

Object Lookup

Search object author, title, date or identifier.

Overview

| | |
|-----------------------|----------|
| object count: | 168 |
| version count: | 210 |
| file count: | 1,902 |
| total size: | 848.4 MB |

Help Topics

- [About Merritt](#)
- [Adding Objects](#)
- [Retrieving Objects](#)
- [Merritt and metadata](#)
- [Using a manifest](#)
- [Editing objects](#)
- [Glossary](#)
- [Sample data](#)

Merritt User Interface

Object: ark:/99999/fk48w3kmw

[Merritt](#) > [Collection: Demo Merritt collection](#) > [Object: ark:/99999/fk48w3kmw](#)

MN Revisors Office. **MN-200900HF0001-0 zip container test 1**

primary identifier: ark:/99999/fk48w3kmw

title: MN-200900HF0001-0 zip container test 1

creator: MN Revisors Office

date: 2009

local id:

last modified: 2010-12-02 06:31 am

created: 2010-12-02 06:31 am

total size (all versions): 148 KB

total versions: 1

[Download object](#)

Object Lookup

[Go](#)

Search object author, title, date or identifier.

Current Version

Version 1: 2010-12-02 06:31 am

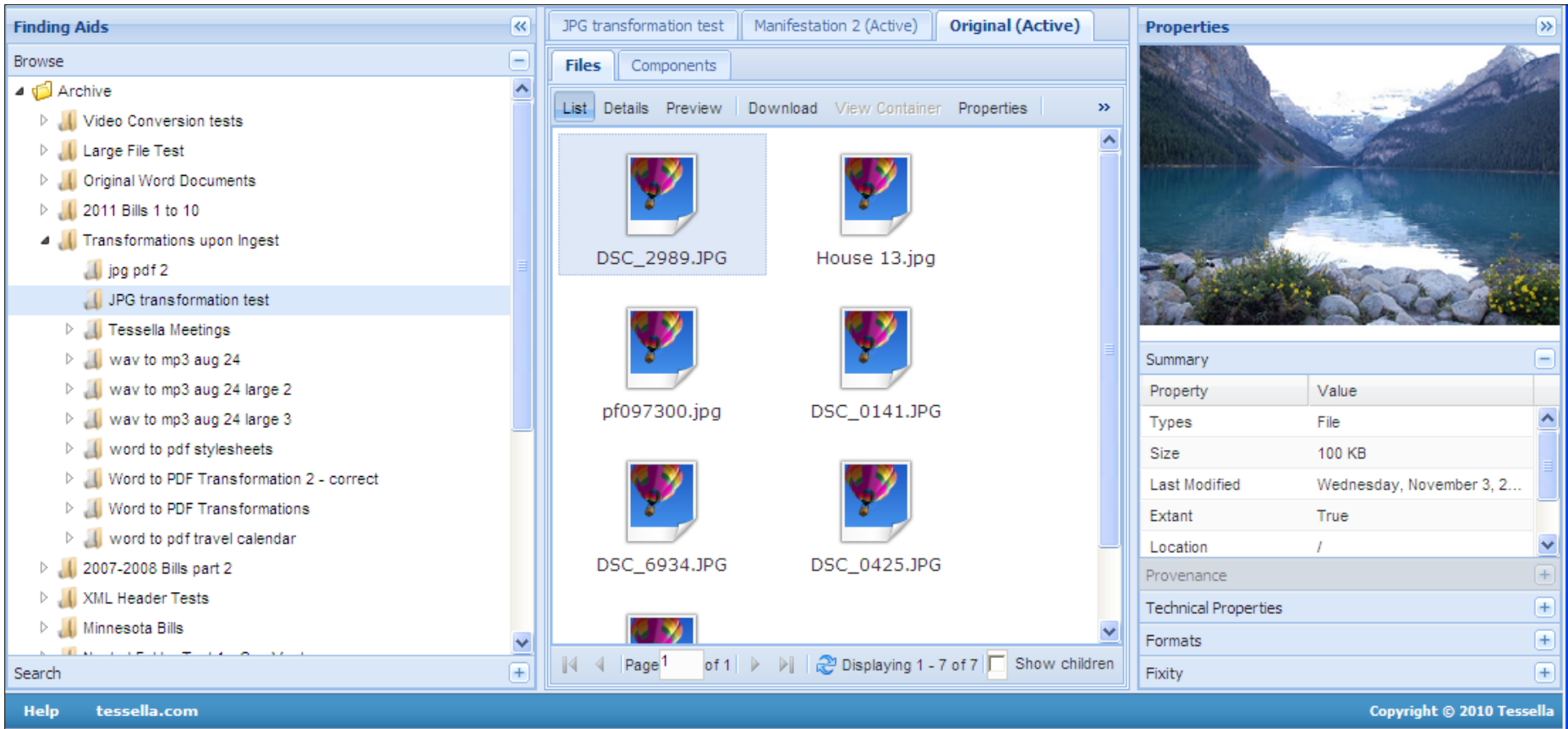
- [MN-200900HF0001-0.md5](#)
- [MN-200900HF0001-0.sha1](#)
- [MN-200900HF0001-0.xml](#)

Merritt is a service of the [University of California Curation Center](#) of the [California Digital Library](#)

Copyright © 2009-2011 The Regents of the University of California

[Image credits](#)

User Interface



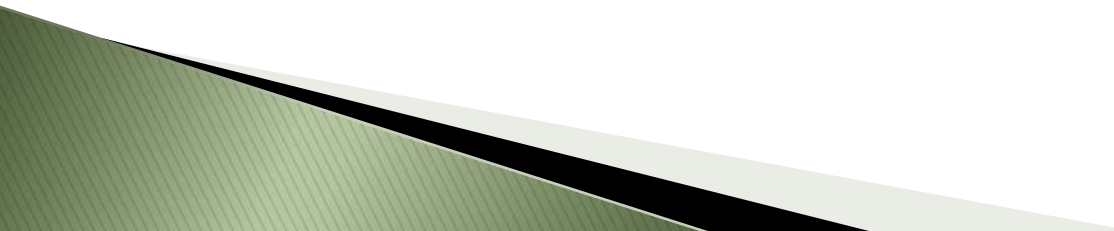
The screenshot displays the Tessella user interface with the following components:

- Finding Aids:** A sidebar on the left with a search bar and a tree view of folders. The selected folder is "JPG transformation test" under "Transformations upon Ingest".
- File Browser:** The main area shows a grid of files:
 - DSC_2989.JPG
 - House 13.jpg
 - pf097300.jpg
 - DSC_0141.JPG
 - DSC_6934.JPG
 - DSC_0425.JPG
- Properties Panel:** On the right, a "Properties" window shows a preview of a landscape image and a "Summary" table.

| Property | Value |
|----------------------|-----------------------------|
| Types | File |
| Size | 100 KB |
| Last Modified | Wednesday, November 3, 2... |
| Extant | True |
| Location | / |
| Provenance | |
| Technical Properties | |
| Formats | |
| Fixity | |

At the bottom of the interface, there is a status bar showing "Page 1 of 1" and "Displaying 1 - 7 of 7" files.

Lessons Learned

- ▶ Know what you want to preserve and how you want to use them
 - ▶ Understand the important elements of a system and how they fit into current routines
 - ▶ Understand the required level of technical skill and experience users will need
- 

Safety Deposit Box

- ▶ Feedback given to Tessella about SDB
- ▶ Changes were made during pilot
- ▶ Future releases to include suggestions
- ▶ Demonstration