

SPONSORED BY



Personal Digital Preservation: Issues and Approaches

Randy Wilson
wilsonr@familysearch.org
RootsTech 2012

The Problem

How to preserve precious photos long-term?

Outline:

- Issues
- Standards
 - Arrangements
 - Face-tagging
- Preservation
- Open discussion

The Situation

- We have photos
 - Shoe box
 - Slides
 - Negatives
 - Albums
 - Documents
- We can scan them
 - Flatbed scanner
 - Slide scanner



The Situation

- We can identify faces
- ...but not all of them
- ...and face tags are still proprietary.



Virginia Lee Gray, Rena Garnes,
Salina Cordelia Speaks Gray,
James Kay Polk Gray
Anita Wilson (front)

The Situation

- We can share photos
 - DVD-ROM
 - Online
 - Facebook
 - Flickr
 - Picasa
 - Etc...
- ...but it is ad-hoc
- ...and short-lived.

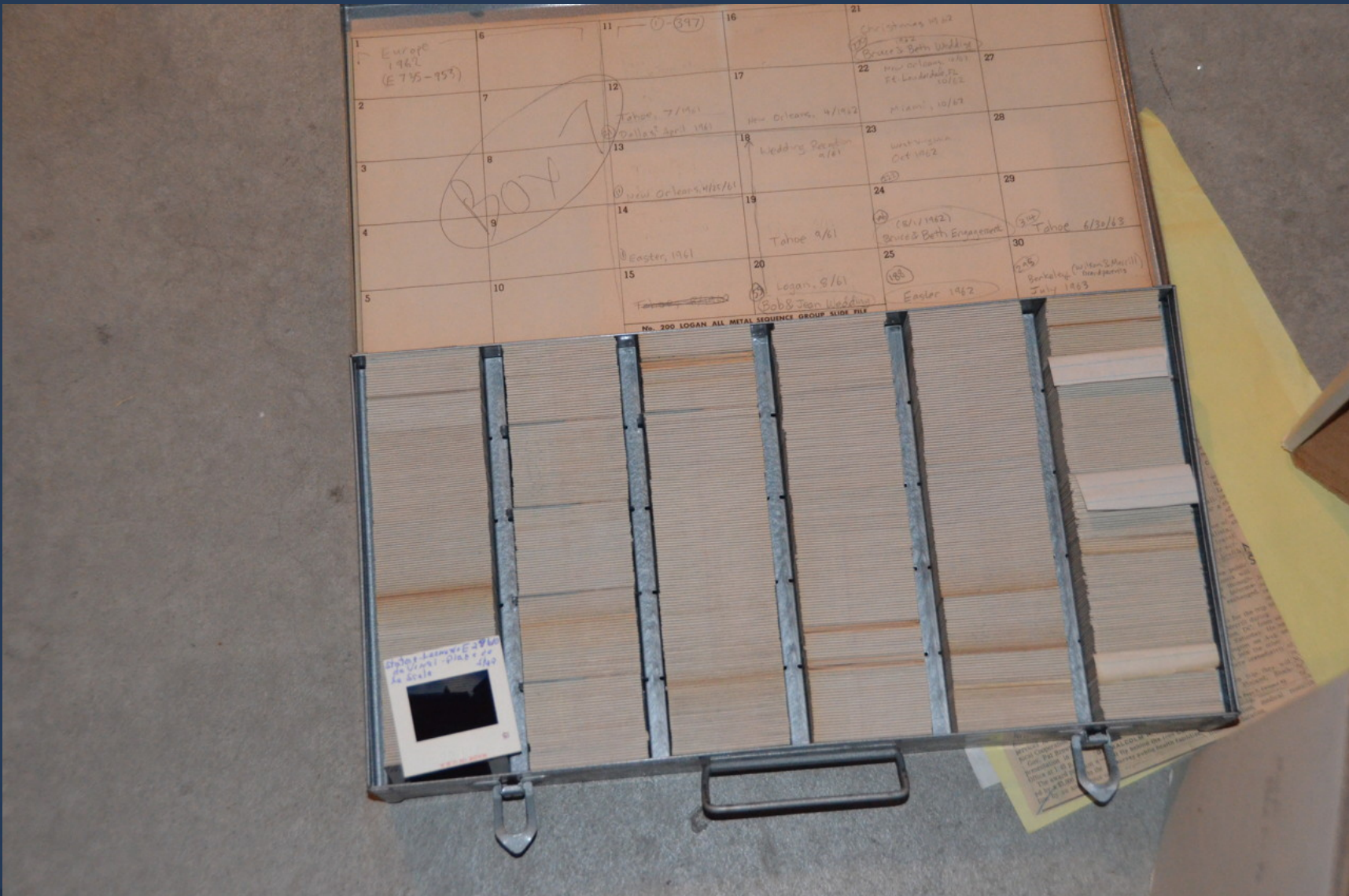


The Situation

Organizing & arranging images is hard

- We can remember how we found them;
- OR we can rearrange them more nicely;
- ...but it is hard to do both, especially long-term.

- Robert's Slides
 - Wooden box
 - Slide 0001.tif
 - Slide 0002.tif
 - ...
 - Small boxes
 - Box 1966A
 - Slide 0001.tif
 - Slide 0002.tif
 - Box 1966B
 - ...





Images and “Artifacts”

Physical Artifacts

- Photograph
- Document
- Journal
- Cassette tape
- Movie Reel

Digital Artifacts

- Image
 - TIFF, JPEG, PDF...
- Audio
 - MP3, WAV...
- Video
 - MOV, AVI, DV...

Archival Principles

http://en.wikipedia.org/wiki/Archival_processing

1. *Respect de Fonds* (Collections/grouping)
2. Respect for Original Order

⇒ Remember the *grouping* and *ordering*.

- Context preserves *meaning* and thus the *value*.
- Groups have similar people, time, place.
- Order preserves time, logical groups.





Physical and Logical Arrangements

Physical Arrangement

Box 00037

Box 00038

Box 00039



Logical Arrangement

1959.08a - Family gathering

1959.08b - Trip to Hawaii

Embedded Arrangement Tags

Embed physical “path” in image metadata, for use when needed.

- As single directory path.

path=“myattic.org/wilsonr/03-MHM/02-Slide_Boxes/Box_07/A0327.tif”

- As nested XML, with “sortKey”

```
<collection title=“Randy’s Photos” uri=“https://myattic.org/ark:12345/047”>  
  <collection title=“Malcolm’s Slides” uri = “https://myattic.org/ark:12345/A7634D8-87”  
    sortKey=“03-MHM”>  
    <arrangement uri=“https://myattic.org/ark:12345/B76FR28” sortKey=“02-Slide_Boxes”>  
      <collection uri=“https://myattic.org/ark:12345/F76R56E-34” sortKey=“Box_07”>  
        <collection uri=“https://myattic.org/ark:12345/H32R56E-34” sortKey=“A0327”>
```

- Can reconstruct arrangement from subset of images.
- Need a standard for portability and longevity of arrangements.

Importance of Standards

Standards needed for

- Interoperability
 - Do work using one tool
 - Migrate to another when needed
 - Work is not lost
- Longevity
 - A proprietary solution only lasts as long as that system.

Face Tagging



Old photos can be:

- *Priceless treasures*
- or
- *Worthless rubbish*

Depending on if you know who it is.

Face Tagging

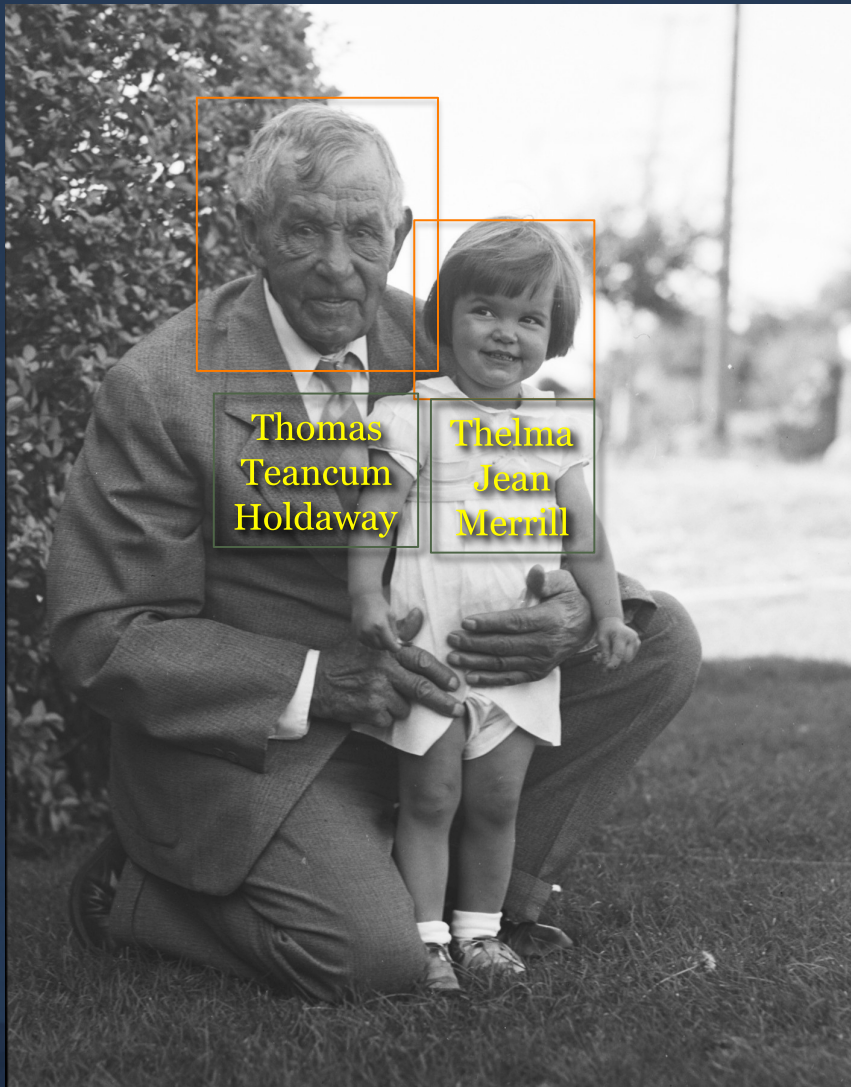


Names are nice:

*Thomas Teancum Holdaway,
Thelma Jean Merrill*

But ambiguous in a
group photo

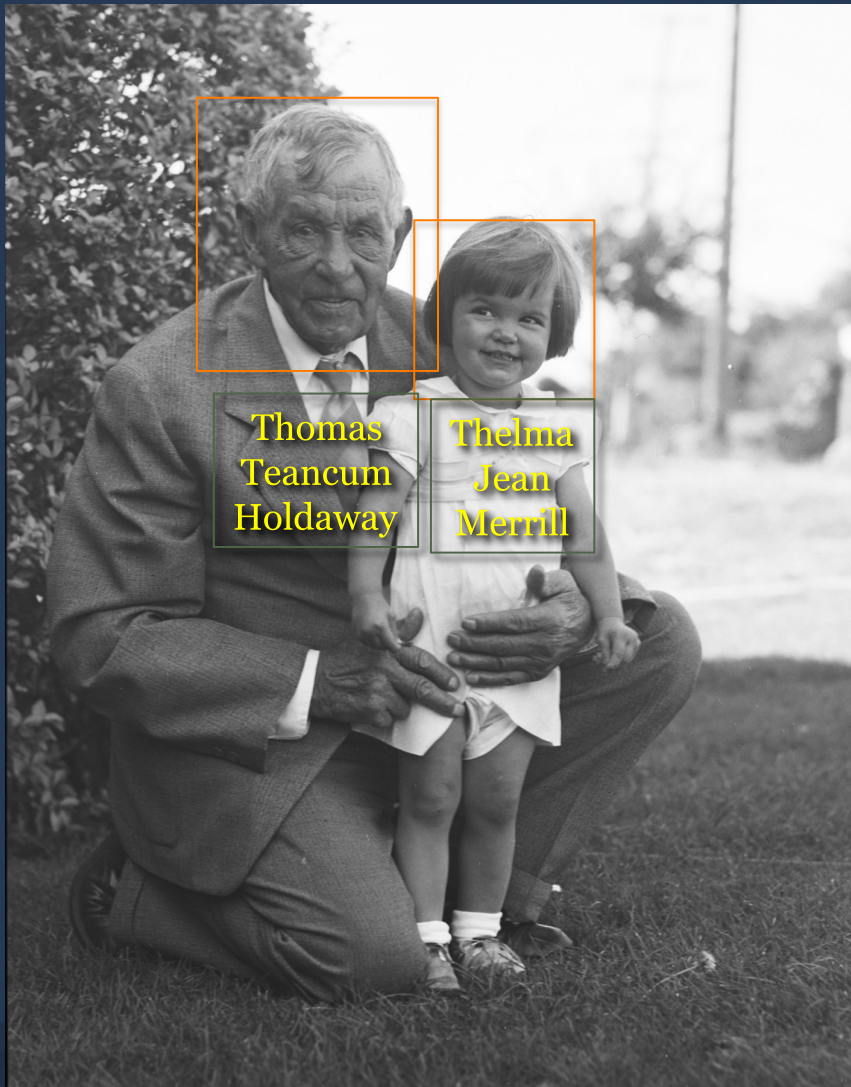
Face Tagging



Face tags are better

- You know which name goes with which person.

Face Tagging



Face tagging systems

- Facebook
- Picasa
- iPhoto
- Photoshop
- Flickr
- Photoloom
- Mundia
- 1000memories.com
- Myheritage
- Heritagecollector
- etc...

Face Tagging

Face tagging systems

- Facebook
- Picasa
- iPhoto
- Photoshop
- Flickr
- Photoloom
- Mundia
- 1000memories.com
- Myheritage
- Heritagecollector
- etc...

- Face recognition
 - Face clusters
- Name vs. Entity
 - Facebook user
 - Ancestor in tree
 - External IDs

Face Tagging Standard

Metadata Working Group (MWG)

- Extension to Adobe XMP
- V2.0: November 2010, includes:
 - Image regions (i.e., face tags)
 - Hierarchical keywords
 - Image collections

MWVG Face Tag Standard

- Define *region* as one of:
 - Rectangle (center, w, h)
 - Circle (center, radius)
 - Pointusing *relative* coordinates (0..1).
- Store original width and height

MWG Face Tag: Handling Edits

- Scaling
 - ⇒ Use *normalized* (0..1) coordinates
- Rotation
 - ⇒ Compliant “changers” rotate regions
- Cropping
 - ⇒ Shift regions
 - ⇒ Shrink and shift regions that are partially cropped.
 - ⇒ Drop regions whose center is cropped

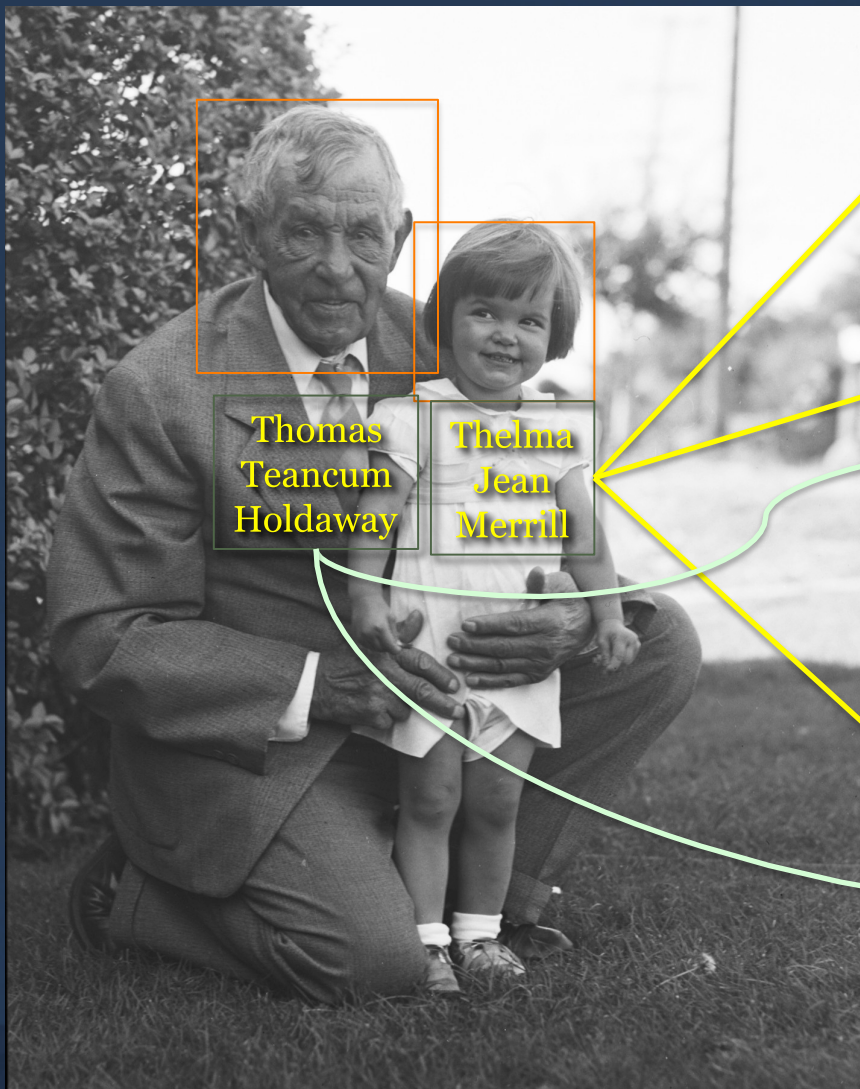
Adopting face-tagging standard

- Metadata Working Group Image Regions
 - No known adopters yet
 - A few adopters would allow users to begin, with hope of future portability.

External Identifiers

- Need extension to handle *external identifiers*.
 - Type: rdf-style URI
 - Facebook User
 - FamilySearch Ancestor
 - Photoloom Person
 - Identifier: URI, usually URL

External Identifiers



Facebook

Jean Wilson

new.familysearch.org

Thelma Jean Merrill

Thomas Teancum Holdaway

Thelma Myrl Holdaway

Mary Eliza White

photoloom.com

Thelma Jean Merrill

Thomas T. Holdaway

Thelma M. Holdaway

FAMILYSEARCH
WHERE GENERATIONS MEET

Preservation Challenges

1. Hard drive crash, fire, theft => *backups*
2. Media degrades (CD-ROM)=>*M-DISC*
3. Obsolete media (5.25” floppies, Zip drive)
4. Obsolete data formats (EBCDIC) => *migrate*
5. Companies go out of business
 - Proprietary formats hard to migrate
6. Dead men don't pay subscription fees
7. Ignorance.
8. Apathy.

Preservation Approaches

- Benevolent Organization
 - Non-profit (e.g., FamilySearch, Internet Archive)
 - Free (e.g., 1000memories), but long-term.
- Prepaid service
 - May need to be backed by “benevolent org.”
- Lots of distributed copies
 - Share with relatives, several online services
 - Unique IDs (URIs) allow sharing of metadata and avoid duplication.
 - Embedded metadata preserves collection info.
 - (LOCKSS—Lots of Copies Keeps Stuff Safe)

Long-lived Links

- Links Break

- Change path

- https://blah.org/v1/books/herman/Grumpy_Dog
 - <https://blah.org/v2/titles/124583>

- => Design paths carefully,

- => Use opaque identifiers

- Change domain

- https://blah.org/ark:/12345/PV7342_34
 - https://next.com/swiped/ark:/12345/PV7342_34

- => Can use “resolver” with long-lived part.

Sharing

- Ad-hoc sharing
 - DVD-ROM, E-mail, Web sites
 - Subset of images, often low resolution
 - No face tags, arrangement info
- Embedded metadata
 - Face tags with external identifiers
 - Help you discover photos of people you care about, and related photos from there.
 - Physical and logical arrangement info/context

Summary

- Organizing, arranging, tagging, preserving, and sharing photos is important to many people.
- Wide adoption of XMP/MWVG face tagging
- Define standards for
 - External IDs on face tags
 - Physical and logical arrangements
 - Unique identifiers embedded in images
- Long-term free or prepaid service; or distributed storage of many copies

Randy Wilson
wilsonr@familysearch.org




FAMILYSEARCH
WHERE GENERATIONS MEET

Thank You.

Sponsored by:



NEW ENGLAND HISTORIC GENEALOGICAL SOCIETY

