

Digital Asset Management for Preservation, Archiving, and Distribution



Steve D. Bowman, Ph.D., P.E., P.G.
Geologic Hazards Program Manager



From a
Folder/File
Based (*not
always very
clean and
useable*)
Structure to a
Managed
System

The screenshot shows a Windows File Explorer window with the address bar set to <code><< GHP >> ResourceSpace Upload >> 2018 Project >> HB2017-32_SLCO</code>. The left sidebar shows a tree view of folders, with 'HB2017-32_SLCO' selected. The main pane displays a list of files with columns for Name, Date modified, Type, and Size. The status bar at the bottom indicates '17 items' and 'Offline status: Online'.

Name	Date modified	Type	Size
2-98-111_A.pdf	3/28/2017 10:00 AM	Adobe Acrobat D...	32,126 KB
7-817-001430_A.pdf	3/28/2017 10:04 AM	Adobe Acrobat D...	6,761 KB
7-817-001430_B.pdf	3/28/2017 10:05 AM	Adobe Acrobat D...	6,571 KB
7-817-001430_C.pdf	3/28/2017 10:05 AM	Adobe Acrobat D...	6,127 KB
7-817-001430_D.pdf	3/28/2017 10:06 AM	Adobe Acrobat D...	2,165 KB
7-817-001430_E.pdf	3/28/2017 10:06 AM	Adobe Acrobat D...	2,256 KB
7-817-001430_F.pdf	3/28/2017 10:06 AM	Adobe Acrobat D...	5,342 KB
8-817-001576.pdf	3/28/2017 10:07 AM	Adobe Acrobat D...	9,776 KB
8-817-001580.pdf	3/28/2017 10:08 AM	Adobe Acrobat D...	8,760 KB
8-817-001580_B.pdf	3/28/2017 10:09 AM	Adobe Acrobat D...	8,784 KB
8-817-001599.pdf	3/28/2017 10:10 AM	Adobe Acrobat D...	6,897 KB
8-817-001608.pdf	3/28/2017 10:12 AM	Adobe Acrobat D...	17,168 KB
8-817-001863.pdf	3/28/2017 10:13 AM	Adobe Acrobat D...	4,851 KB
98-03.pdf	3/28/2017 10:13 AM	Adobe Acrobat D...	3,234 KB
983232.pdf	3/28/2017 10:13 AM	Adobe Acrobat D...	1,242 KB
61987027.pdf	3/28/2017 10:15 AM	Adobe Acrobat D...	18,793 KB
E95-2052.pdf	3/28/2017 10:16 AM	Adobe Acrobat D...	14,926 KB

Geologic Data Preservation: The Overwhelming Types and Volume of Digital Data!

File Formats

- Open-Source – best option for archiving, no patent encumbrances, likely usable in the future.
- Common or Custom Closed-Source – generally a poor choice for archiving, may require specialized software that may not be usable in the future, encumbered by patents, and possibly, user costs.
- Legacy (old) – increasing difficult to use and/or convert to modern formats.

Compression or the Lack of

- LZW – lossless, patent expired in 2003, older method.
- Deflate (zip) – lossless, in the public domain after 1989; newer, higher compression ratio than LZW.
 - Images: JPEG – lossless or lossy, widely used.
 - Images: JPEG2000 – lossless or lossy wavelet-based compression based on JPEG.



Geologic Data Preservation: The Overwhelming Types and Volume of Digital Data!

Storage Hardware/Media

- Legacy Media (Floppy Disks, 9-Track Tapes, Tape Cartridges, CDs, DVDs, etc.) bad
- Local Computers (hard drives) bad
- Portable Hard Drives – not so good
- Network Attached Storage (NAS) Systems
- Storage Area Network (SAN) Systems
- Cloud Storage Providers
 - Traditional Methods (file/folder structures, many file systems available [NTFS, etc. through NFS, etc.])
 - Object-Based Methods (Amazon S3, etc., requires FUSE connector or URL calls with an API)
 - No guarantee of data permanent existence, other limitations possible on data availability, etc.
 - Costs (data storage service levels, data download/usage), can be significant and difficult to estimate. Typically, billed on a monthly basis.



Part of the Solution: Digital Asset Management (DAM) Systems

Computer system to manage archiving, ingestion, annotation, cataloguing, storage, retrieval, and distribution of digital assets and their metadata.

Should handle the downloading, renaming, backing up, rating, grouping, archiving, optimizing, maintaining, thinning, and exporting files functions.



Open-Source Software

Open-source software is software whose source code is published and made available to the public, enabling anyone to copy, modify and redistribute the source code without paying royalties or fees.

Open-source code evolves through community cooperation composed of individual programmers, as well as, commercial organizations.

It is often found that some of the individual programmers who start an open-source project may end up establishing commercial organizations offering products or services incorporating open-source programs. This may include offering custom modifications to the open-source code, offering *Software as a Service* (Saas) hosting, user support, etc.



Digital Asset Management (DAM) Systems

Commercial or Closed-Source

- Documentum – very expensive and complex
- CONTENTdm
- Atlassian – mainly documents, mixed license model, complex
- many others

Open-Source

- ResourceSpace – started 2006, Linux or Windows + PHP, Apache
- Dspace – started 2002
- Razuna
- Fedora – started 2003, can be used to build a management system
- many others (open-source systems usually have a commercial supporter for SaaS, support, etc., with varying open-source licensing)



How to Choose a Digital Asset Management (DAM) System

Minimum Requirements

- Widely used and supported, so that a path to the future exists (do not get into legacy or near-legacy solutions).
- Low cost – open-source often the best, over commercial or in-house developed solutions.
- Easy export of the data to common, open formats. Do not want to get into closed, proprietary data solutions and formats.
- Handles common data formats (PDF, TIF, JPEG, etc.).
- Ease of use by data preservation staff and users (your customers!).

Additional Requirements

- Installation – server, storage, network, security
- Maintenance – long-term, how difficult will the system be to maintain?
 - Special dependencies, are they updated often? Do they work with the latest operating systems? Are “hacks” needed to make parts of the system work? If so, likely security and continuity issues.



How to Choose a Digital Asset Management (DAM) System

Open-Source Server

- LAMP – Linux operating system, Apache webserver, MySQL database server, and Perl/PHP/Python scripting languages (other combinations possible)

Open-Source Data Processing

- ImageMagick – graphics file conversion and processing.
- Exiftool – read and write metadata to graphics and other file formats.
- FFmpeg – audio and video file conversion and processing (transcoding). Must use the correct codec (many proprietary).
- Ghostscript – Postscript and PDF file conversion and processing.

Open-Source Data Web Display

- OpenLayers or Leaflet – map processing and display of various raster and vector layers on a webpage.



ResourceSpace (open-source DAM system)

Open-Source Server

- Developed in 2006 for Oxfam (UK NGO organization) and open-sourced licensed (BSD permissive free software).
- Full source code available.
- Active, open-source community providing security, new features, and bug fix updates.
- Two to four major updates per year released.
- Based on a LAMP server system; however, can use Windows servers.
- Easy to install, maintain, update, and use.
- No limitation to the number of items (resources) managed or the size of those resources – only limited by the hardware and network used.
- A backup of resource metadata is stored in a XML file specific to each resource.
 - No proprietary dependencies or formats used, easy to export data should the need arise in the future. Full access to data in MySQL.



Use of ResourceSpace at the Utah Geological Survey

UGS GeoData Archive System
Utah Seismic Safety Commission Website



ResourceSpace Demonstration

Virtual Ubuntu Linux (LAMP) Server Appliance Using
VirtualBox on a Windows 10 Workstation



Utah Aerial Imagery Collection and Web Application

Collection

- Contains over 130,000 individual frames (mostly stereo coverage; some low-sun-angle and oblique imagery) from across Utah and a 0.25 degree buffer.
- Frames collected by federal agencies before 1955 generally only available from the National Archives in College Park, Maryland. Our collection contains a significant amount of pre-1955 frames, along with UGS collected frames.
- High-quality digital scans of paper (800 dpi) and film (1200 dpi) frames using color calibrated Epson 10000XL (Silverfast control) and Creo EverSmart Select II scanners.
 - JPEG in TIF for web access, deflate lossless compressed TIF for archival original.

Database and Web Application

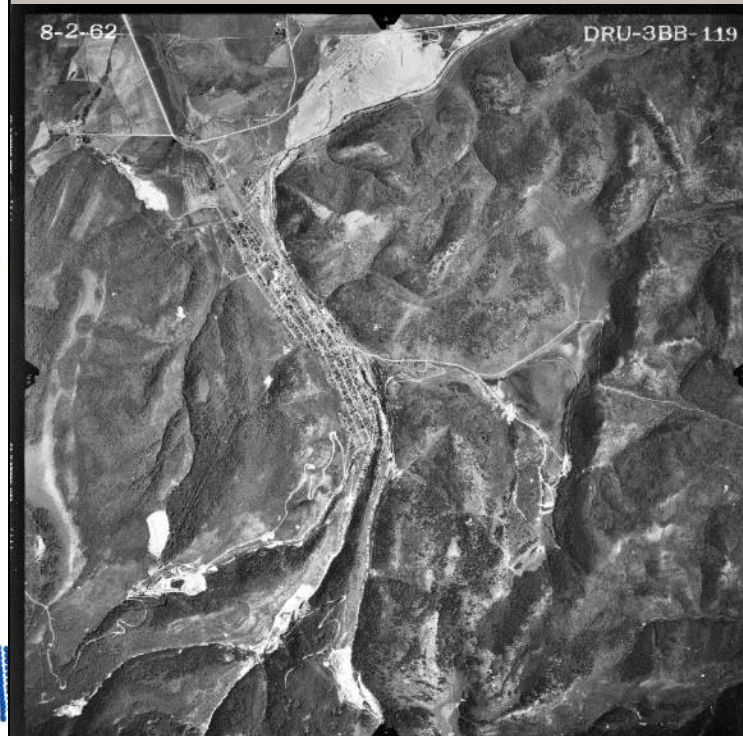
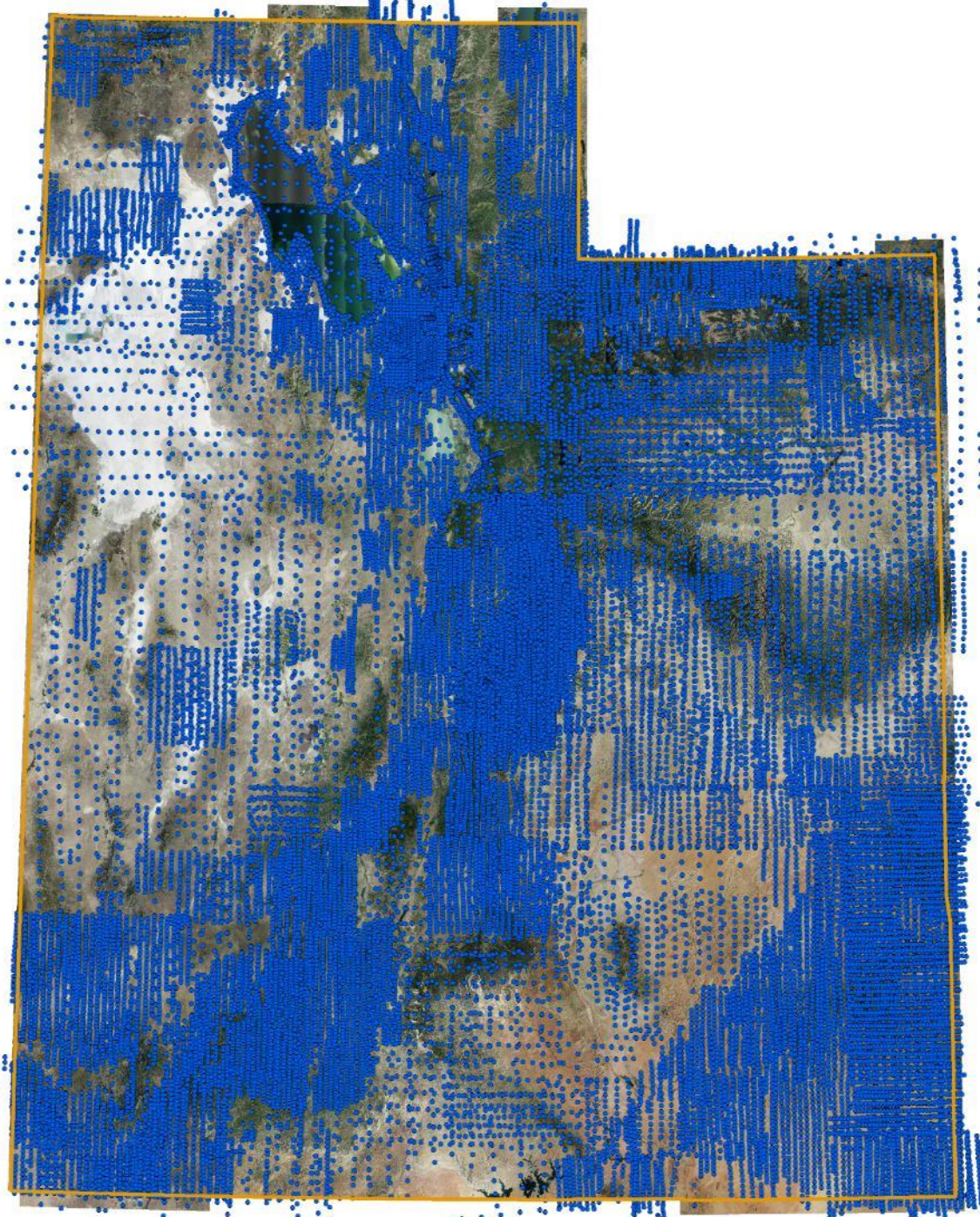
- Microsoft SQL database, PHP web application, and C# management program.
 - In need of migration, due to PHP MSSQL driver not included in PHP7 (PHP5 being depreciated 12/31/2018).
 - ResourceSpace is a good replacement candidate.



UGS Historical Aerial Imagery Collection

Over 97,000 images from across the state are currently available in the UGS Aerial Imagery database.

<https://geodata.geology.utah.gov/imagery/>



UGS Aerial Imagery Collection

Aerial Photography Frame Center Point Locations (you may need to zoom in to see points)

Login

1935-1959 1960-1989 1990-present

Search Results Info Data Sets Help

This database contains 88,792 individual photographs. Low-resolution JPEG images can be viewed online and high-resolution TIFF images can either be downloaded by the user or transferred to a user's portable drive by the [Natural Resources Map & Bookstore](#) (click the [Help](#) tab for more information).

Click on an individual photograph point on the map to show a popup containing basic metadata and a small preview image.

To search for photographs, users can create a search-bounding box by moving the map markers, by using the Draw Box, by entering latitude and longitude coordinates, or by typing in an address (street address, city, state) and region size. Enter additional search criteria to narrow your search.

Click on the Search button to display list of selected photographs.

Search by Region

NE Corner	SW Corner
Latitude: <input type="text" value="38.75"/>	Latitude: <input type="text" value="38.5"/>
Longitude: <input type="text" value="-111.75"/>	Longitude: <input type="text" value="-112.25"/>

Search by Address

Type address here

Region Size (in miles): Width: Height:

Additional Criteria

From	To	Project Code:
Year: <input type="text" value="1930"/>	<input type="text" value="2012"/>	<input type="text"/>
Scale: <input type="text"/>	<input type="text"/>	Search Limit: <input type="text" value="250"/>

