Anticipating Storage Requirements: Management and Archiving

Greg Schneider
UMMZ Collections Manager
Division of Reptiles and Amphibians



- Management and Archiving
 - Best Practices
 - Redundancy
 - Manual vs. Automated
- Storage and Space Requirements
 - File Types and Folders Examples
 - Access
 - Space Requirements
- Concluding Remarks





guidelines.pdf - Google Chrome

U.S. National Archives and Records Administration (NARA)

Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images

For the Following Record Types-Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts

June 2004

Written by Steven Puglia, Jeffrey Reed, and Erin Rhodes

Digital Imaging Lab, Special Media Preservation Laboratory, Preservation Programs U.S. National Archives and Records Administration 8601 Adelphi Road, Room B572, College Park, MD, 20740, USA Lab Phone: 301-837-3706 Email: preserve@nara.gov

Acknowledgements: Thank you to Dr. Don Williams for target analyses, technical guidance based on his extensive experience, and assistance on the assessment of digital capture devices. Thank you to the following for reading drafts of these guidelines and providing comments: Stephen Chapman, Bill Comstock, Maggie Hale, and David Remington of Harvard University; Phil Michel and Kit Peterson of the Library of Congress; and Doris Hamburg, Kitty Nicholson, and Mary Lynn Ritzenthaler of the U.S. National Archives and Records Administration.

SCOPE:

The NARA Technical Guidelines for Digitizing Archival Materials for Electronic Access define approaches for creating digital surrogates for facilitating access and reproduction; they are not considered appropriate for preservation reformatting to create surrogates that will replace original records. The Technical Guidelines presented here are based on the procedures used by the Digital Imaging Lab of NARA's Special Media Preservation Laboratory for digitizing archival records and the creation of production master image files, and are a revision of the 1998 "NARA Guidelines for Digitizing Archival Materials for Electronic Access", which describes the imaging approach used for NARA's pilot Electronic Access Project.



Journal - JDI Home Groups **IIP/CIIP Career** Events Membership

Archiving, Chapter 5: Data Storage Management

More in this Section...











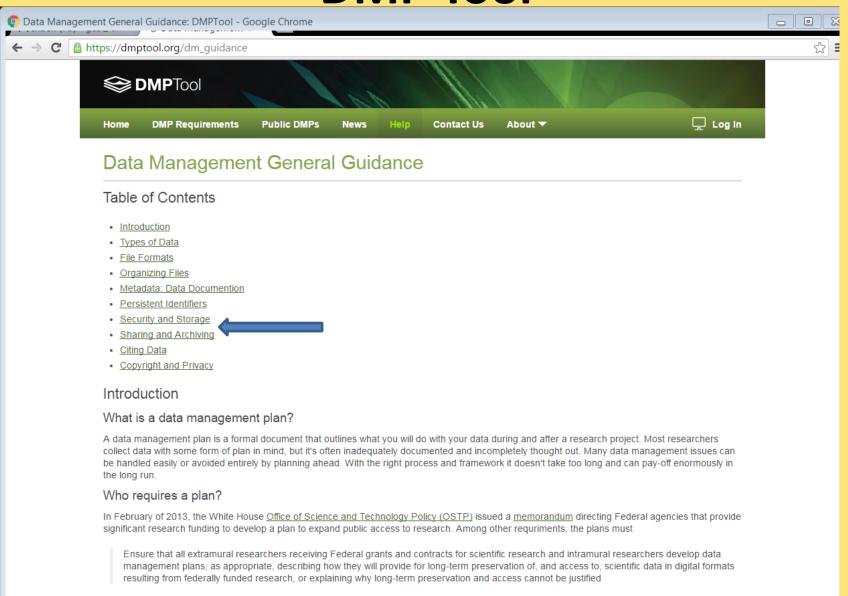
Author:

Robert Glicksman, Medical Imaging Technical Specialist (Retired)

Introduction

Medical image archives can grow to contain enormous amounts of data. Radiology, Cardiology, or Pathology image sets can consume several gigabytes of storage. The sum total of image data associated with many patients and over many years can run into the petabyte (10¹⁵bytes) range. Obviously, a practical medical image archive must economize on the cost of raw storage.

DMP Tool



The National Science Foundation (NSF) already requires a 2-page plan as part of the funding proposal process. Soon most or all US Federally funded

We can help

grants will require some form of data management plan.

File Formats

Examples of preferred format choices:

- · Image: JPEG, JPG-2000, PNG, TIFF
- . Text: plain text (TXT), HTML, XML, PDF/A
- · Audio: AIFF, WAVE
- · Containers: TAR, GZIP, ZIP
- Databases: prefer XML or CSV to native binary formats

Examples of discouraged format choices and better alternatives:

Discouraged Format	Alternative Format
Excel (.xls, .xlsx)	Comma Separated Values (.csv)
Word (.doc, .docx)	plain text (.txt), or if formatting is needed, PDF/A (.pdf)
PowerPoint (.ppt, .pptx)	PDF/A (.pdf)
Photoshop (.psd)	TIFF (.tif, .tiff)
Quicktime (.mov)	MPEG-4 (.mp4)

DMPTool Background

DMPTool Background

The original DMPTool was a grassroots effort, beginning in January 2011 with eight institutions partnering to provide in-kind contributions of personnel and development. The effort was in direct response to demands from funding agencies, such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), that researchers plan for managing their research data. By joining forces the contributing institutions are able to consolidate expertise and reduce costs in addressing data management needs.

The original contributing institutions were:

- The University of California Curation Center (UC3) at the California Digital Library
- DataONE
- <u>Digital Curation Centre (UK)</u>
- Smithsonian Institution
- University of California, Los Angeles Library
- University of California, San Diego Libraries
- University of Illinois, Urbana-Champaign Library
- University of Virginia Library

Redundancy

Backups and storage

Making regular backups is an integral part of data management. You can backup data to your personal computer, external hard drives, or departmental or university servers. Software that makes backups for you automatically can simplify this process considerably. CDs or DVDs are not recommended because they are easily lost, decay rapidly, and fail frequently. The UK Data Archive provides additional guidelines on data storage, backup and security.

Backup Your Data

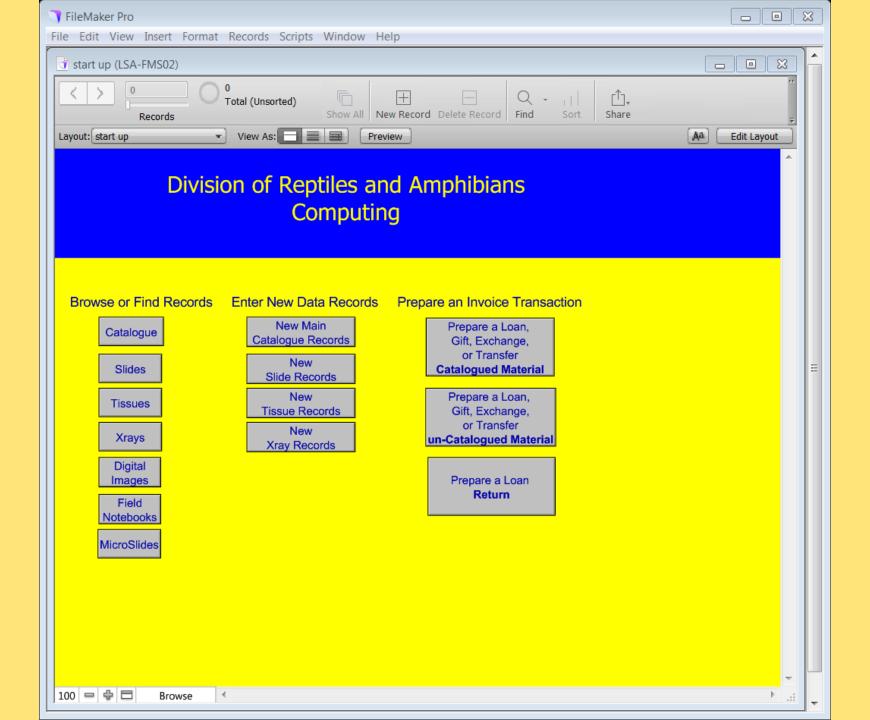
- Good practice is to have three copies in at least two locations (e.g. original + external/local backup + external/remote backup)
- Geographically distribute your local and remote copies to reduce risk of calamity at the same location (power outage, flood, fire, etc.)

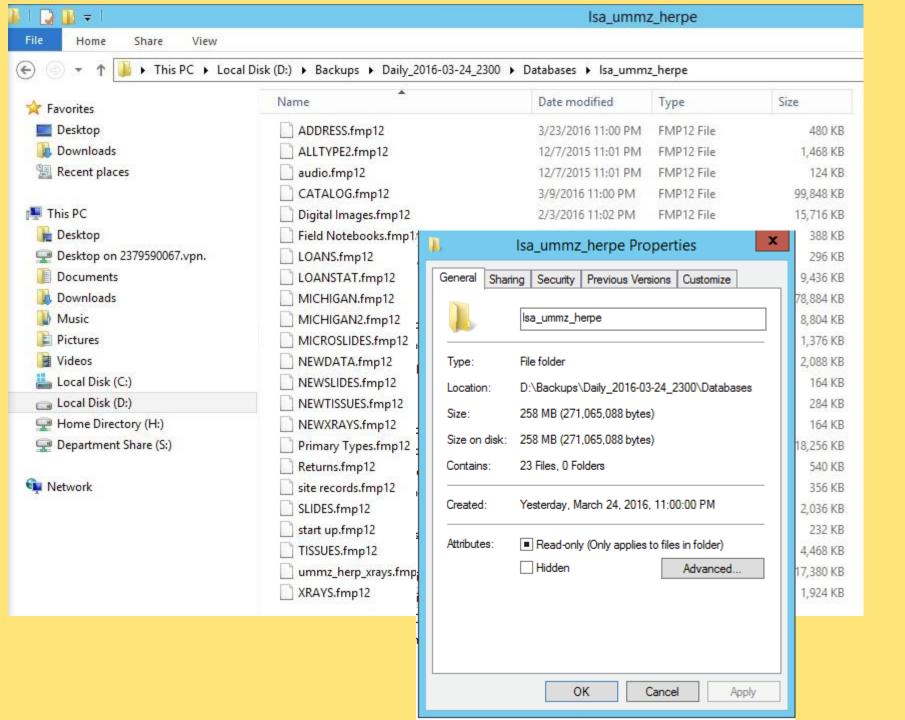
Data Backup Options

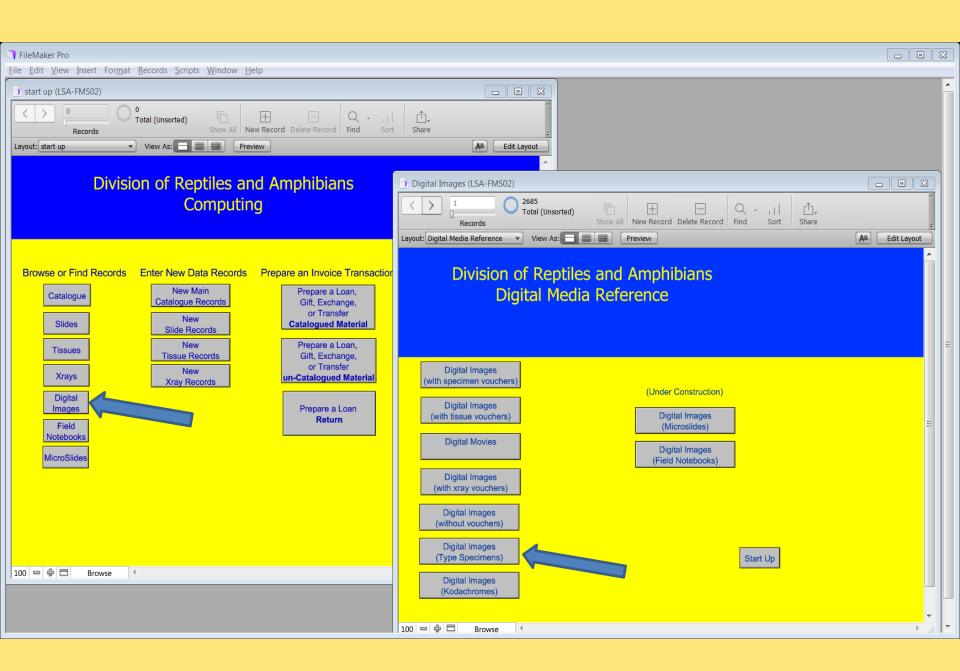
- · Hard drive using software like:
 - · Windows 8 File History
 - OS X Time Machine
 - Linux/UNIX rsync)
- · Tape backup system
- Many institutions provide a service similar to <u>UCBackup</u> at UC Berkeley. Check with your campus IT support to see if backup service is available.
 Althernately, your academic department may provide storage space and backup services.
- · Cloud storage some examples of private sector storage resources include:
 - · Amazon S3 and Glacier-Requires client software, no encryption support
 - S3-based Remote Hard Drive Services such as <u>Elephant Drive</u> and <u>Jungle Disk</u>.
 - · Mozy (from EMC) Free client software, 448-bit Blowfish encryption or AES key
 - Carbonite Free

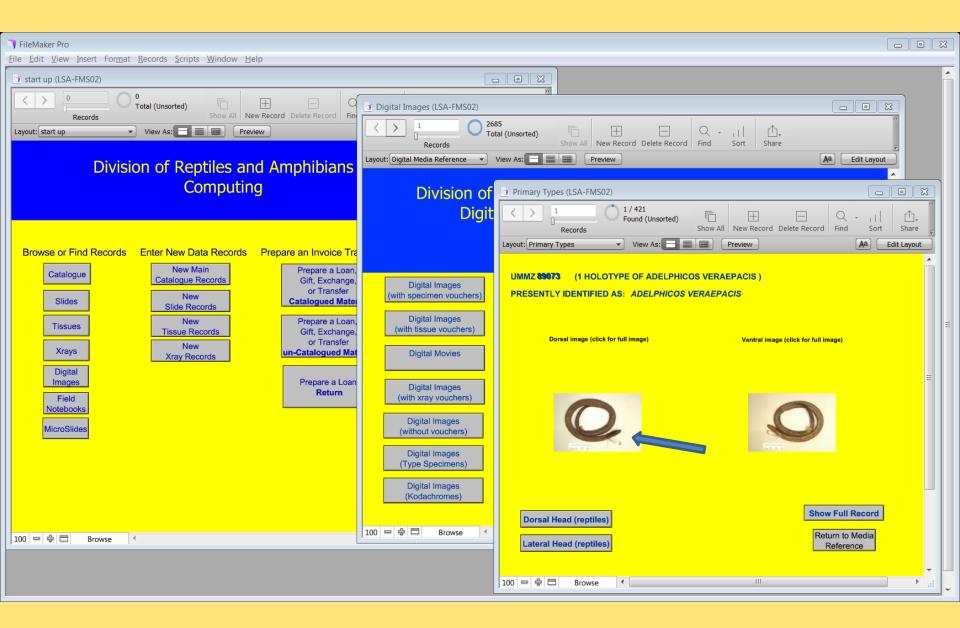
Storage and Space Requirements

- File Types and Folders Examples
- Access
- Space Requirements

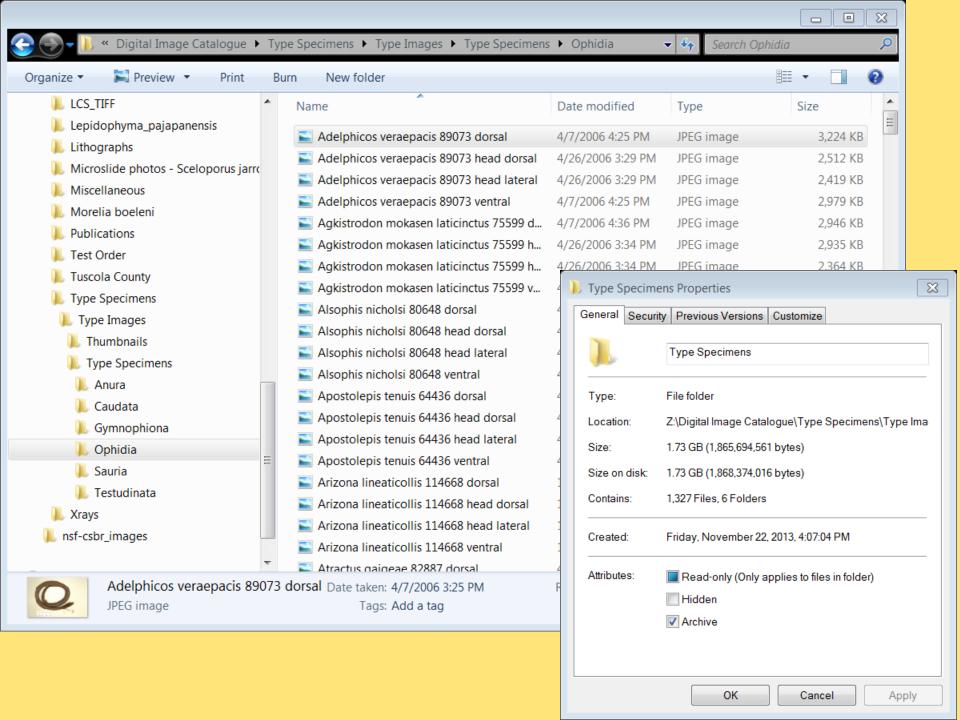


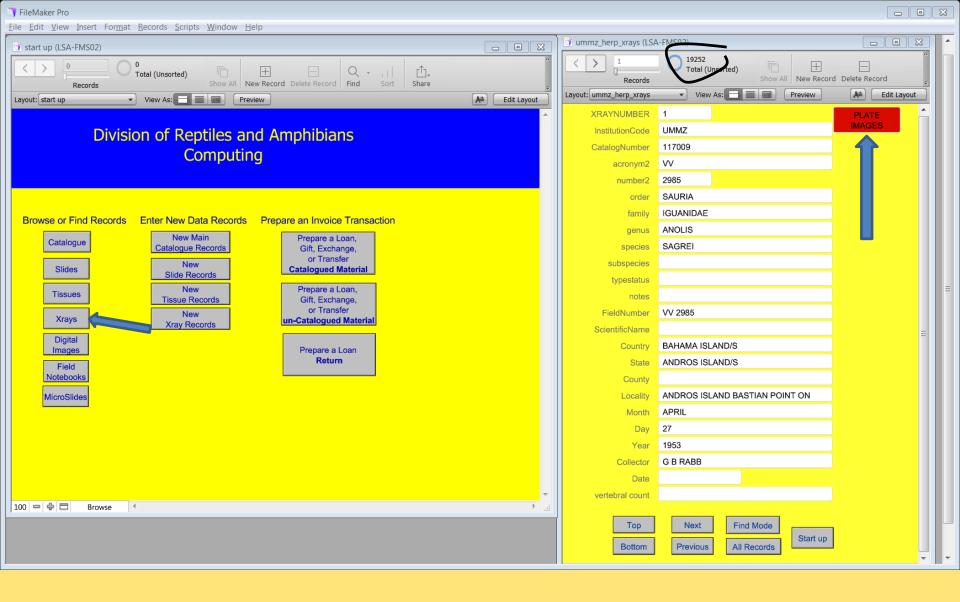




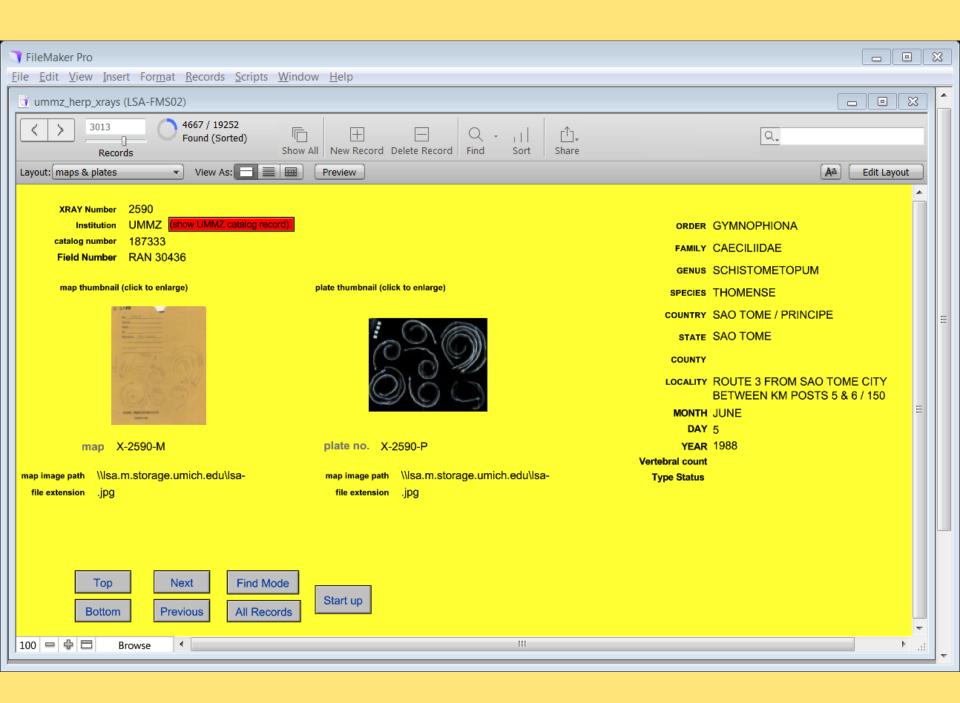


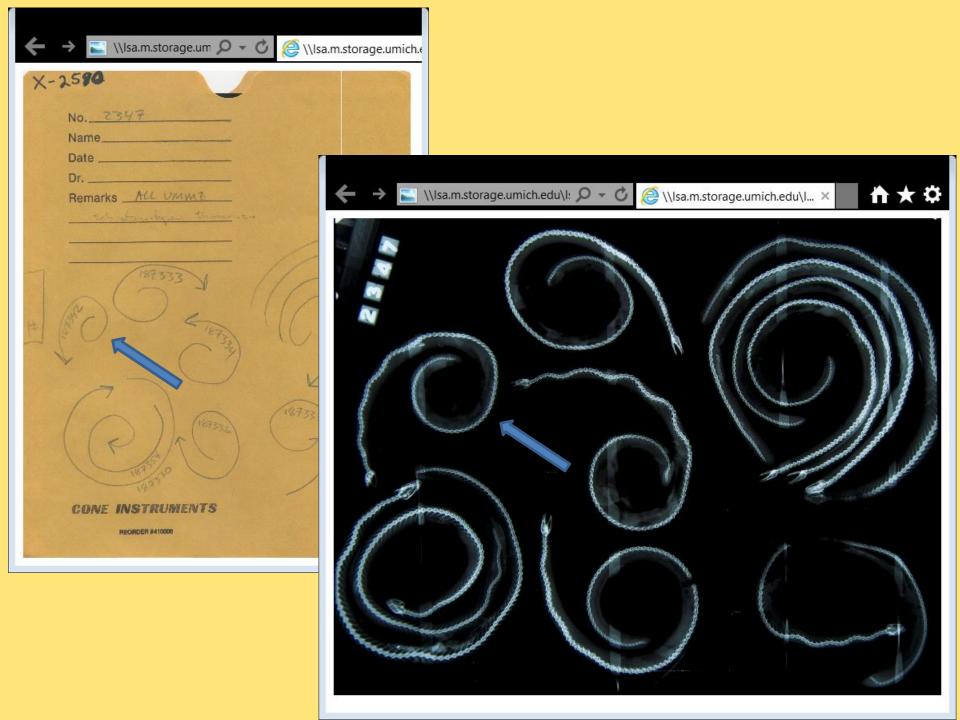


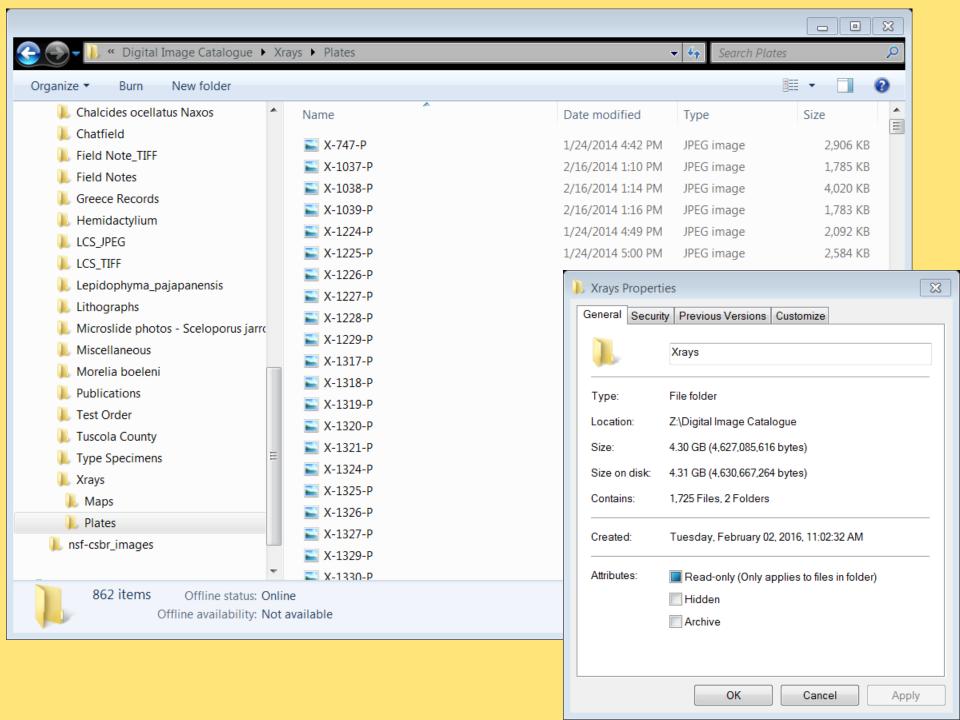


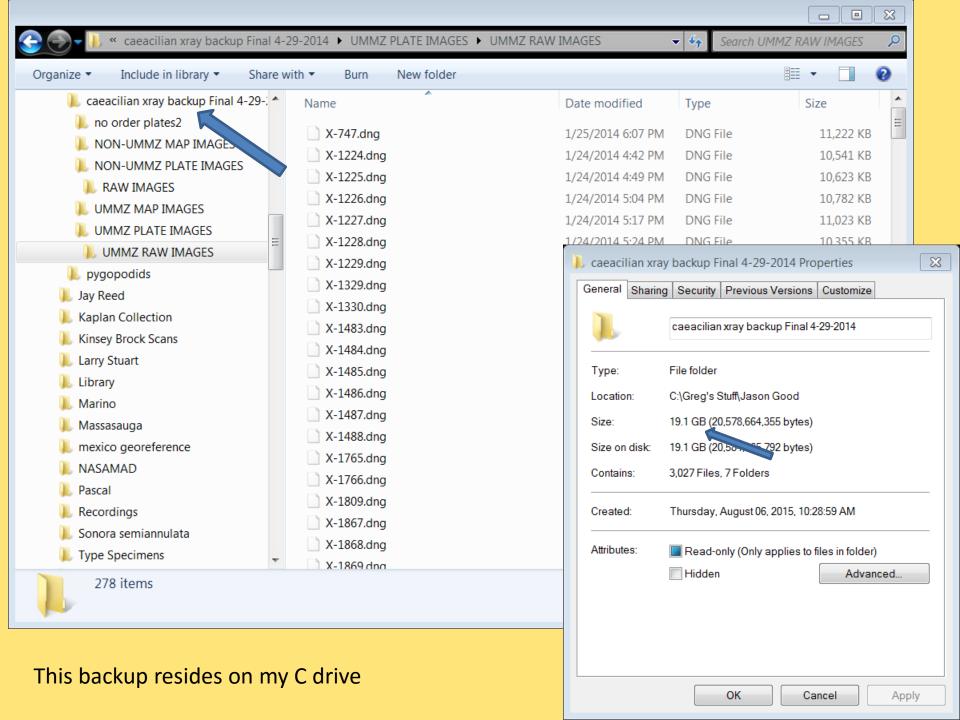


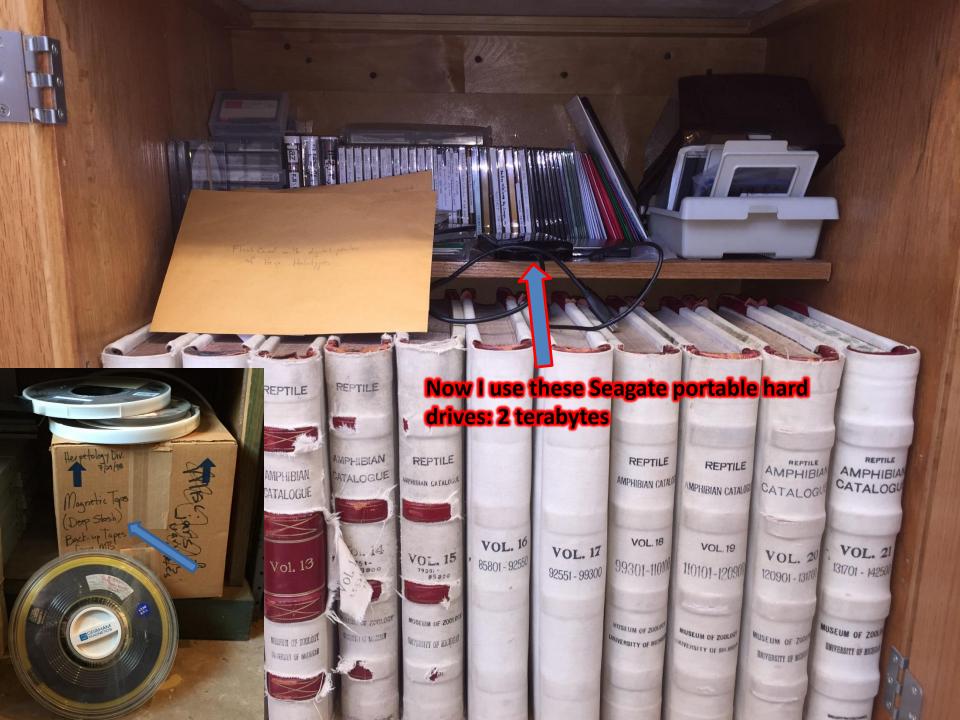
2,939 radiograph plates containing 19252 images (records) 4,667 have digital images

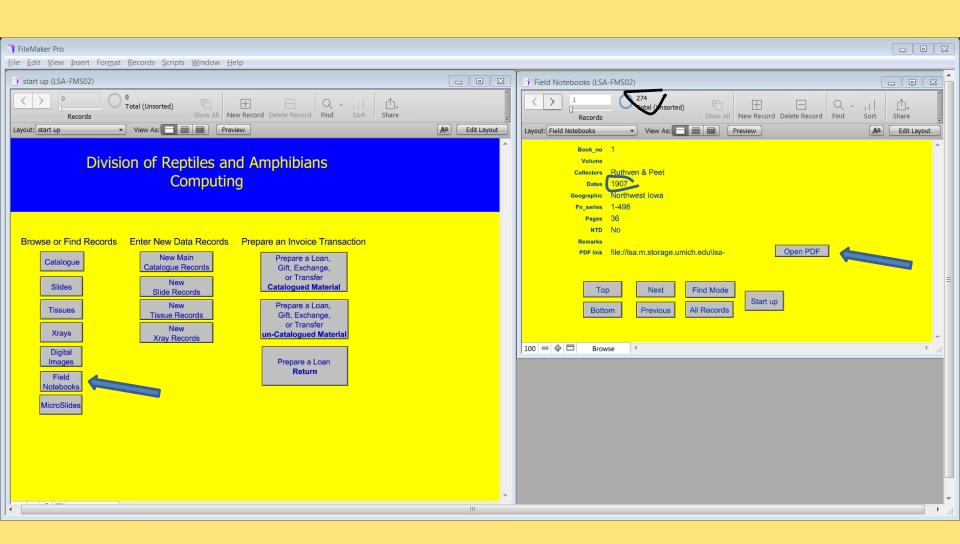




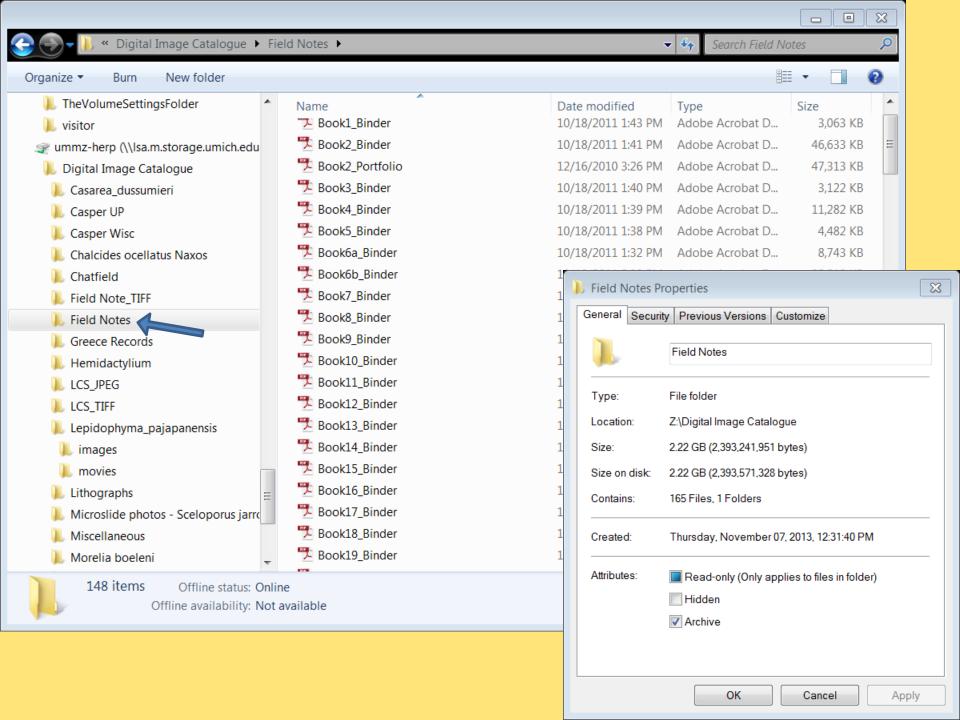


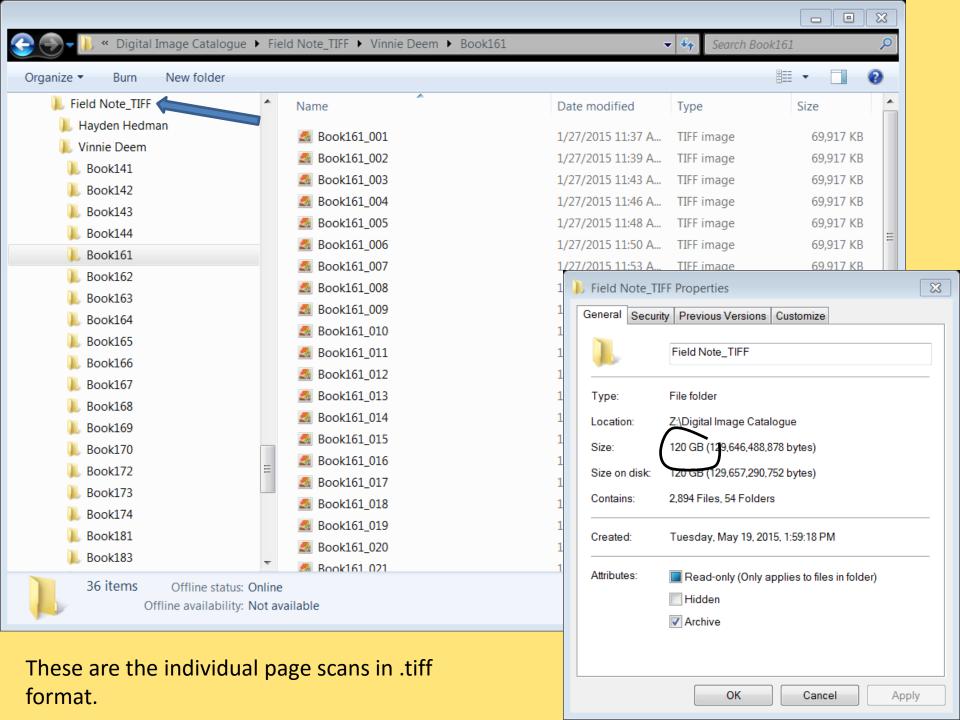


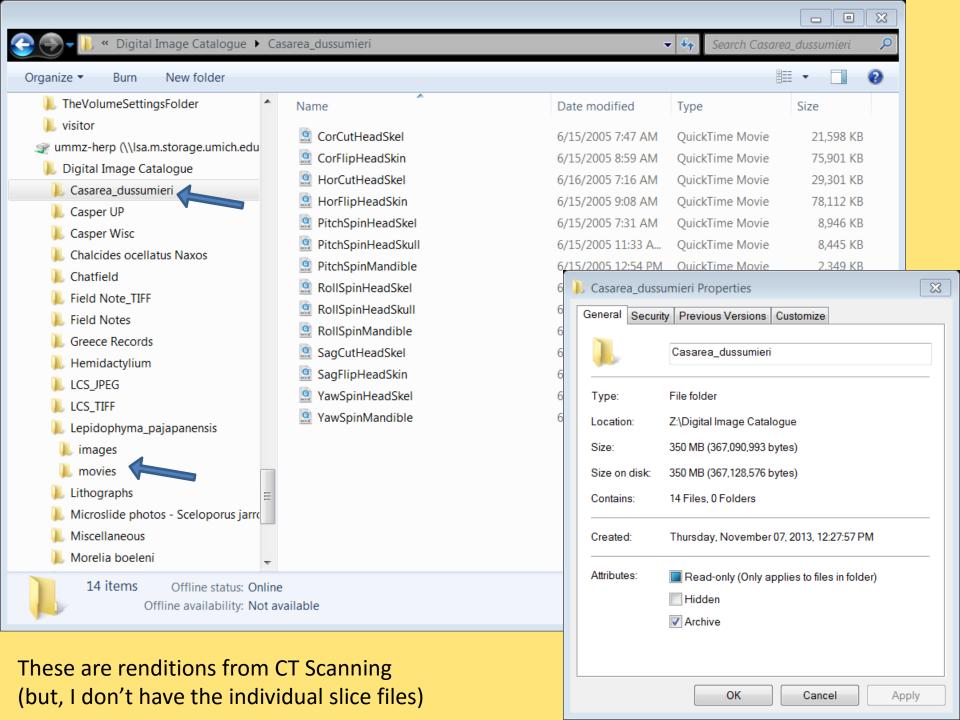


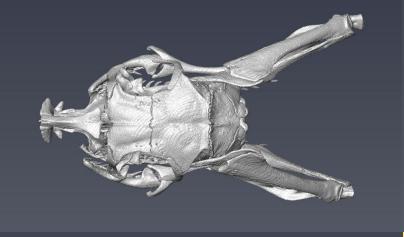


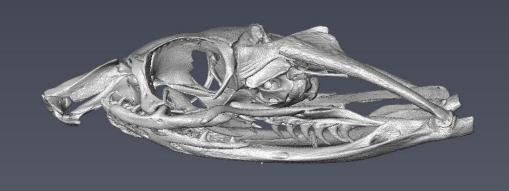


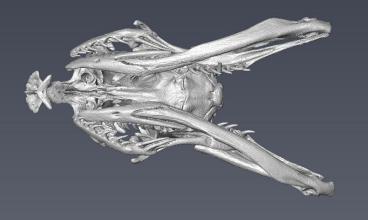






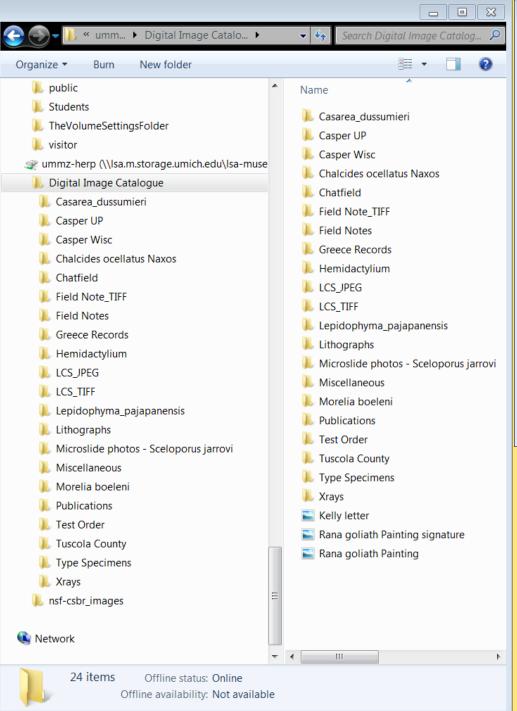


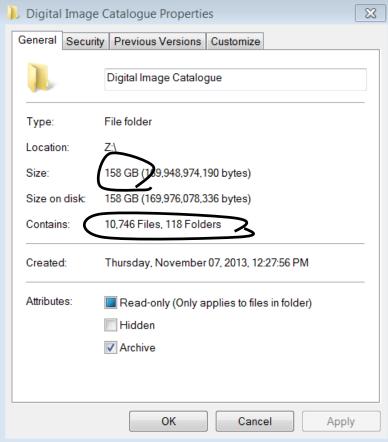




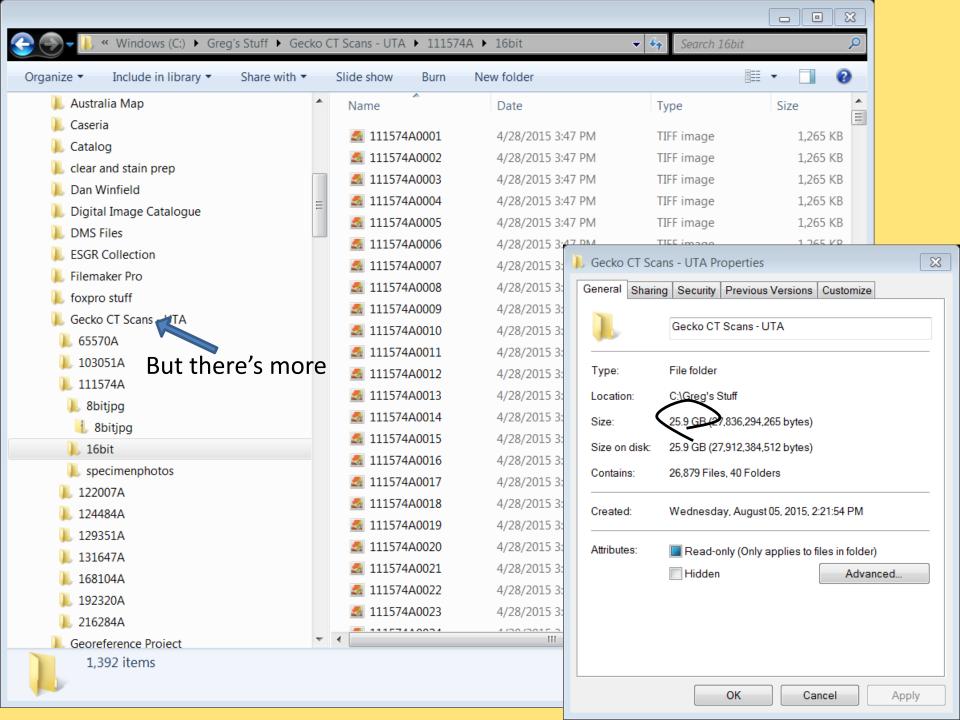
CT Scan Pilot Project for UMMZ snake skulls

- The resolution was at 20 microns.
- The surface rendition model is a file size of 543 MB.
- There are 503 slices, each is 4.34
 MB, thus the total size of the folder is 2.17 GB

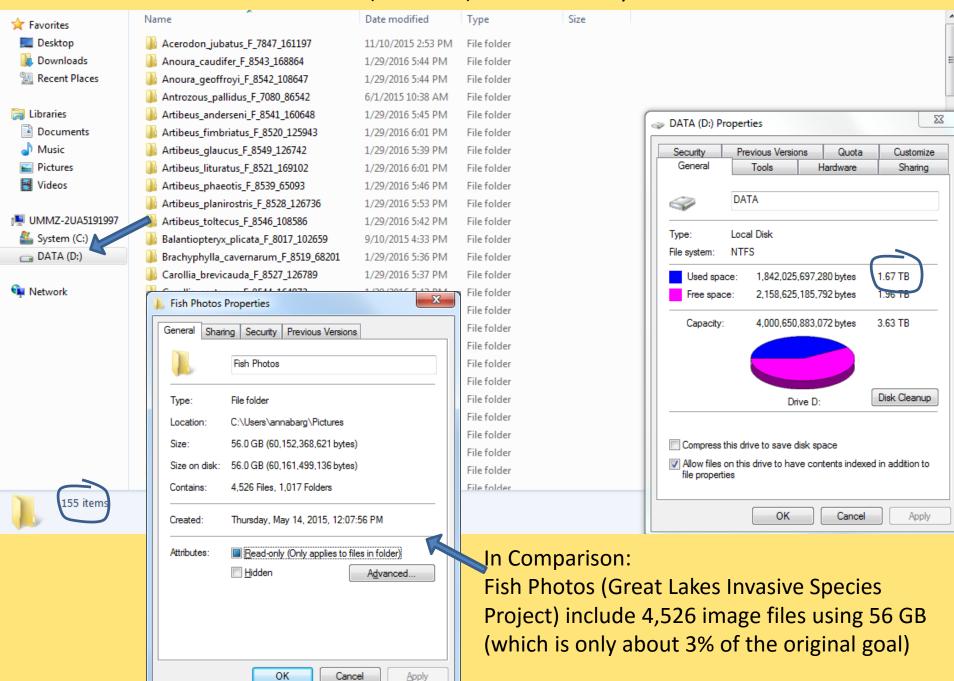


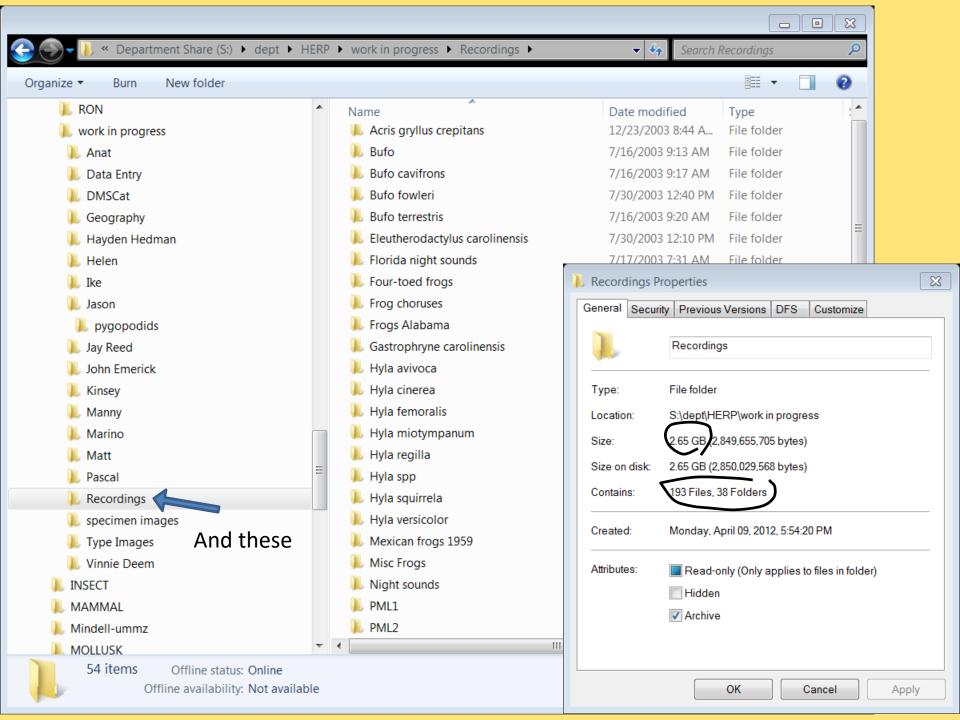


These are the digital images that reside on our LS&A Server that has automatic backups daily and weekly.

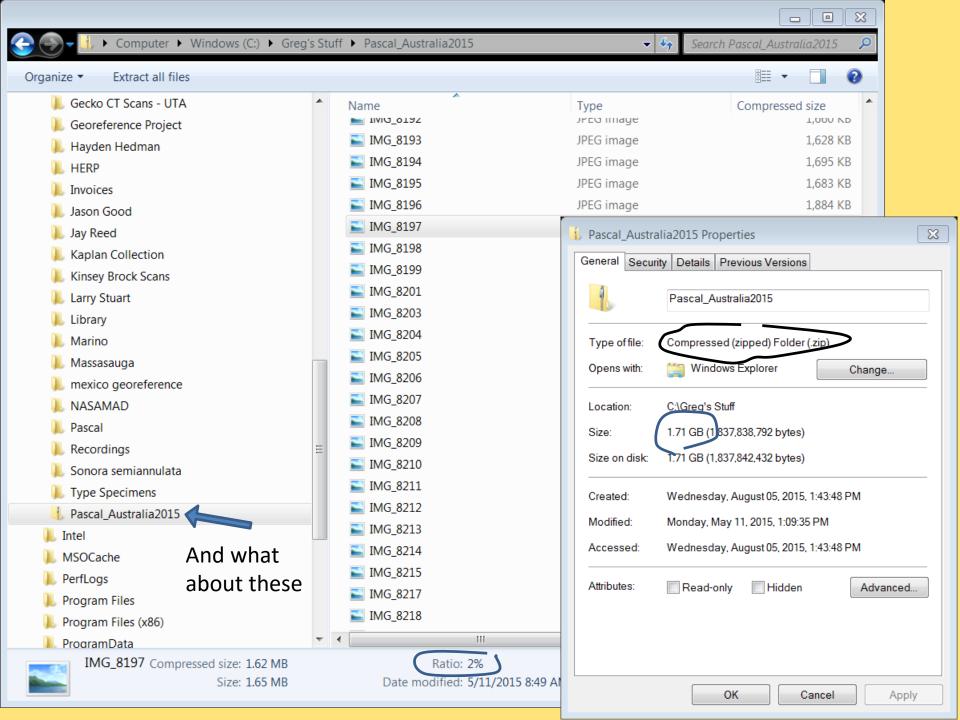


These 155 folders of CT Scan data (bat skulls) use 1.67 Terabytes on an external hard drive.



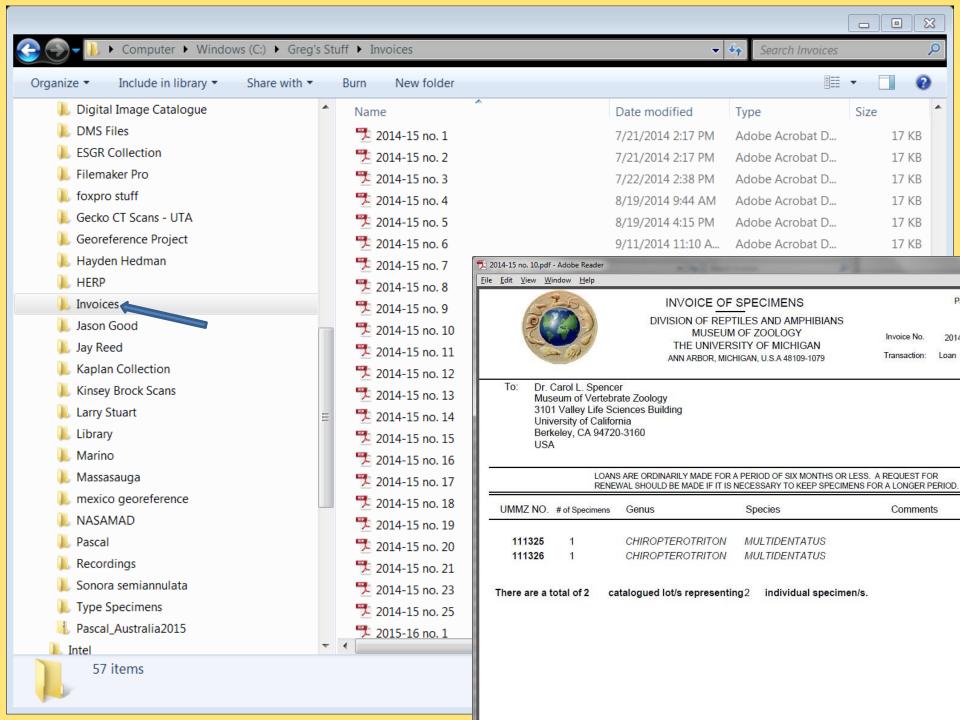








This one image (by Pascal Title) from our Australia trip last year is 19 MB (.jpg). We have a lot of these to process.



Concluding Remarks

- Space Requirements a lot
 - Database data maybe up to 1 GB
 - Image data up to 250 GB (and rapidly growing, especially with CT imaging)
 - Redundancy How many copies / backups ?
- Data Management Plan DMP Tool
 - Everyone should have one I wish I did
 - Where will we store it and at what cost
 - Who is responsible

