



Geospatial Multistate Archive and Preservation Partnership

Investigating Storage Architectures for Long-term Preservation: Channeling the Archival Data Deluge

Alec Bethune- NC Center for Geographic Information and Analysis

North Carolina Electronic Records Forum | 3 November 2011 | Raleigh, NC



We're doing backups, isn't that archiving?

- Backups – a means to save and recover current records
- Personal Archives- “Keeping stuff” locally, on drives /DVDs, or on an online file share
- True Archiving – formally preserving important data **permanently** in a structured, dynamic digital repository



Preservation Repository Architecture

➤ Staging Area:

- Processing, Review and Verification space
- Checksums/ Virus Scan



➤ Preservation Environment:

- Secure area where preservation copy(ies) will be stored/managed
- Hashing for long-term integrity verification

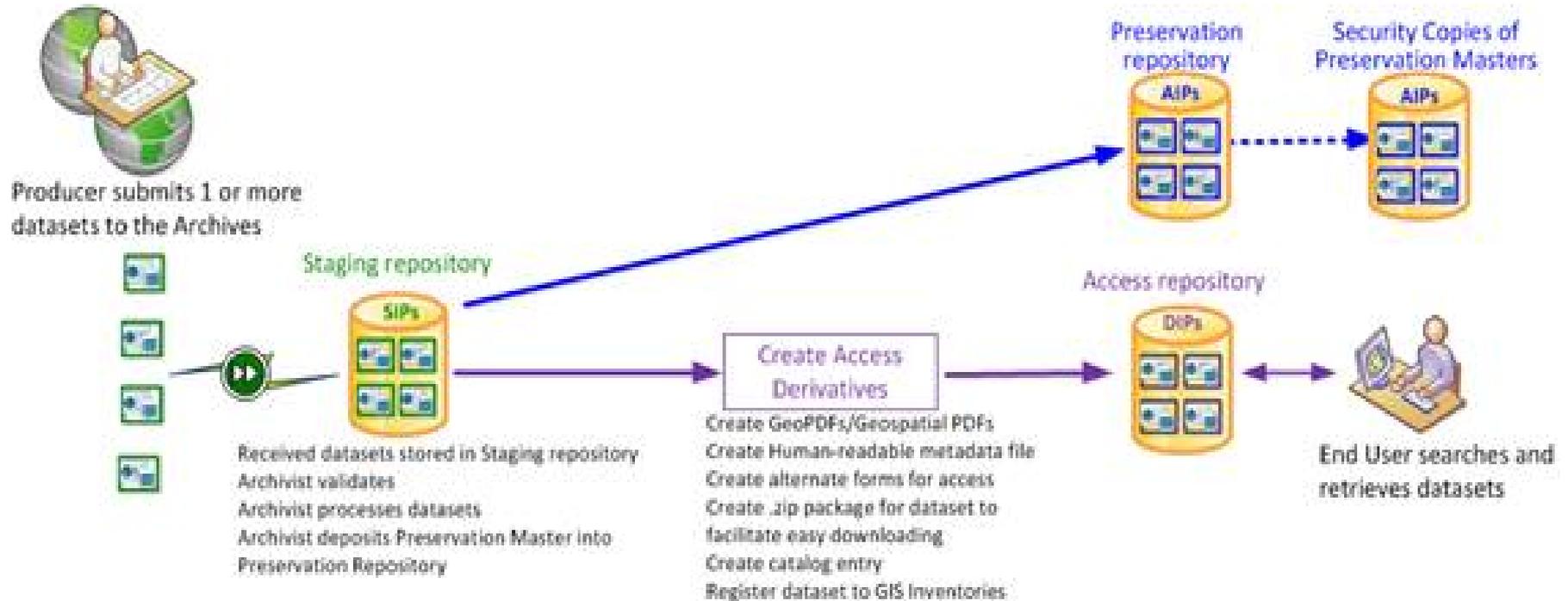


➤ Access Environment:

- Home for copy of data for public use and access



Preservation Architecture/ OAIS



Preservation Storage: Storage Media Options



Technologies- Storage Area Networks (SAN)

- High availability/ performance/ scalability
 - Supports data replication
- Costly
- Durable
- Well-supported by IT shops
 - Though complex enterprise architecture
- Disk block based storage/fibre channel



Technologies- Network Attached Storage (NAS)

- Slower than SAN, but still “online”
- Less costly than SAN
- Durable
- Ease of implementation/ support
- File based system/ TCP/IP based

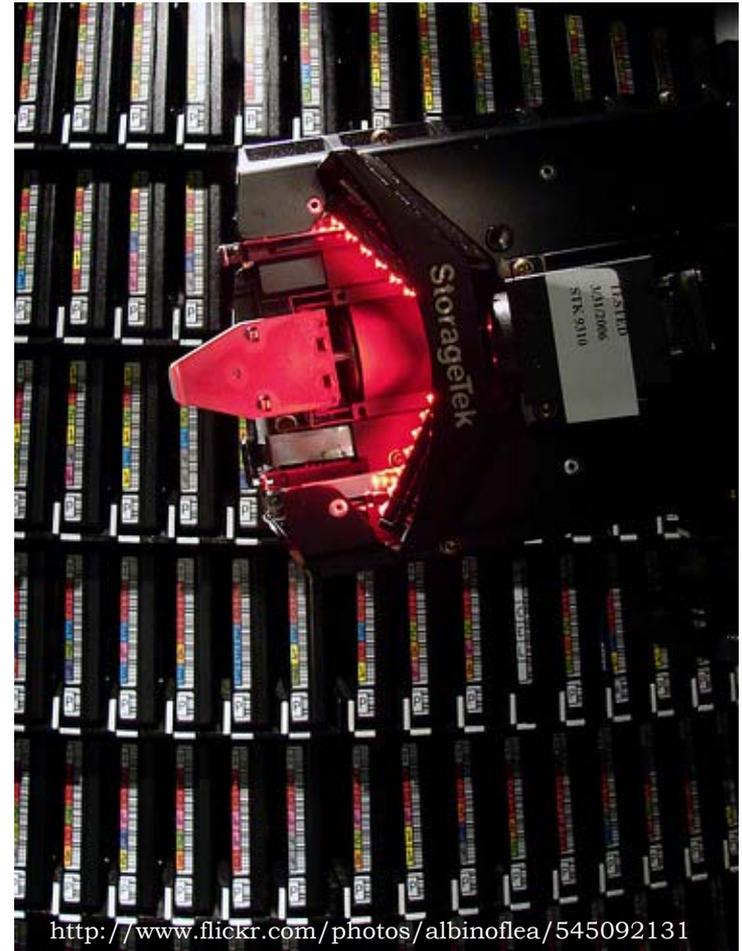


<http://www.flickr.com/photos/evansfam/3881882662>



Technologies- Tape Storage

- Slower Access
 - Near-line and off-line options
- Lower cost than SAN
 - Though large initial infrastructure investment
 - Requires significant maintenance
- Durable
- Not all IT shops offer the service
- Cumbersome file access



Technologies- Cloud Storage

- Variable Performance
 - Scalability & Ease of Implementation
 - Sexiness vs Reality
 - No local control of media
 - May present security/authentication risks
 - Complex Pricing
 - Different vendor pricing models
 - Requires monthly fee, cannot be purchased outright
 - Provides services local IT may not offer
-
- ➤ Checksum maintenance



Technologies- Portable Hard Drives

- Fast Load/Unload
- Very inexpensive
- Not durable
 - Sensitive to corruption and data loss
- Easy for Archives to use
- Not easily networked
- Security risk: drives can walk out the door if not secured
- Possible uses:
 - Temporary backup to more robust solution
 - Data transfer to the archives from the field



Technologies- Thumb Drives, DVDs, CDs

- Storage capacity is low
- Very Inexpensive
- Flash drives not durable/ scratching-cracking for CD/DVDs
- Security risk: media can walk out the door if not secured
- May be used to transfer data to an archives from the field



<http://www.ubergizmo.com/2008/09/usb-thumb-drive/>



<http://www.flickr.com/photos/eyebee/2267435575/>



Preservation Storage Tiers/ Considerations

- Data transfer/ingest space
- “Dark” storage for preservation copy
- Online use for access copies of records



- Managing active archival processing
- Audit, Replication and Transformation functions across storage units
- Don't forget Backups!

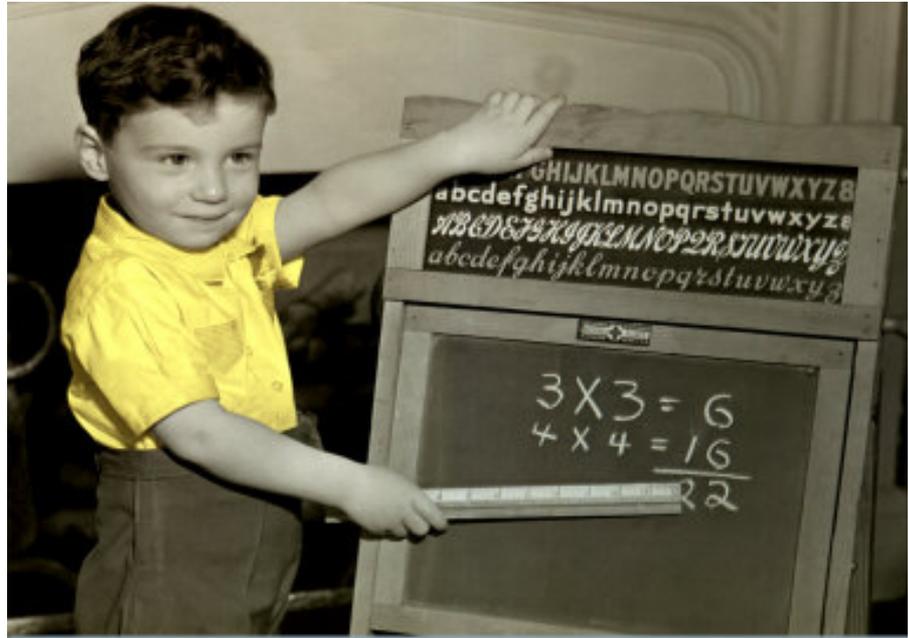


Recap

- Back-ups and Digital Archiving aren't the same thing
- Preservation Architecture: Staging, Preservation, and Access Environments
- Lots of storage options in the marketplace
 - Factors to consider: cost, data access, reliability, ease of implementation
- Matching storage architecture with preservation need



Questions???



There are
NO STUPID QUESTIONS
or stupid answers.

<http://www.insidesocal.com/tomhoffarth/27640-No-Stupid-Questions-Posters.jpg>

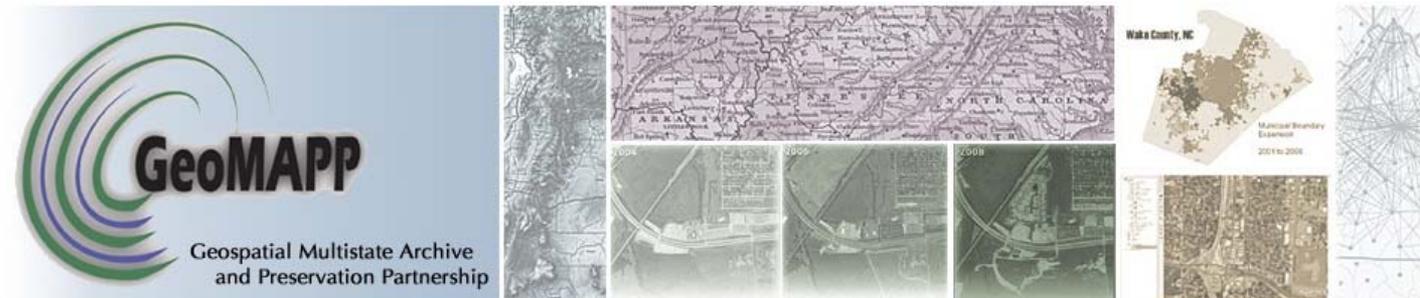
Thanks!!



<http://imagecache2.allposters.com/>

- ▶ Alec Bethune (NC CGIA)
alec.bethune@nc.gov

Follow GeoMAPP on Facebook and Twitter!



-
- ▶ www.geomapp.net