GeoMAPP

Geospatial Multistate Archive and Preservation Partnership

Storage Options

February 24, 2011

Business Need for Storage

➢ GIS geospatial data submitter to the archives (e.g. GIS Clearinghouse)

　➢ Space for actively used datasets and imagery.

　➢ Staging area to prepare the datasets for transfer.

　➢ Space requirements: 2x the expected total datasets size, as you'll have a copy of the dataset that you're preparing, plus the final prepared package of the datasets to transfer.

　➢ Note if the data submitter is also producing uncompressed orthos for transfer to the archives, you would also need to take that into account.

Business Need for Storage

➢ Geospatial archival repository

  ➢ Staging area to receive the dataset

  ➢ Preservation space to retain the originally received dataset

  ➢ Staging space to process the received dataset

  ➢ Preservation space for the preservation "master" version of the dataset

  ➢ Space for copy(ies) of the "preservation" master (may choose to store to offline media such as tape)

  ➢ Space for the "access" copy, plus any other potential derivatives that may be used to facilitate access (e.g. .zip of the dataset files, PDF of a shapefile, HTML of the geospatial metadata)

Informational Partner Feedback

➢ What are the current storage technologies or services used by your state for GIS?

  ➢ Hardware used for orthoimagery

  ➢ Hardware used for vector data

➢ What are the current storage technologies or services used by your state for archives and preservation?

  ➢ Storing or plan to store orthos?

  ➢ Hardware or services used for vector data

State's Approaches

➢ Architecture and Points of Pain

   ➢ Kentucky

   ➢ North Carolina

   ➢ Utah

Technologies

- SAN
  - High availability
  - Costly
    - Archives-owned vs. monthly IT fee
    - Sensitive to economic disruption
  - Durable
  - Well-supported by IT shops

Technologies

- Tape
  - Near-line and off-line options
    - Less prone to risk from economic downturns
  - Slower
  - Lower cost than SAN
  - Requires significant maintenance
  - Not all IT shops offer the service
  - Durable
  - Difficult to access individual files

Informational Partner Feedback

➢ What does the term cloud computing/storage mean to you?

➢ How has your state implemented cloud computing/ storage solutions?

Technologies

- Cloud
  - Scales well
  - Less costly than SAN
    - Requires monthly fee, cannot be purchased outright
  - No local control of media
  - May present security/authentication risks
  - Provides services local IT may not offer
    - Checksum maintenance

Technologies

➢ NAS

  ➢ Less costly than SAN

   ➢ No monthly fee, unless local IT requires it

  ➢ Fairly durable

  ➢ Slower than SAN, but still "online"

  ➢ Can be managed by Archives in-house

  ➢ Easy access to individual files

Technologies

➢ Solid State Disk

  ➢ Durable, no moving parts

  ➢ Very expensive

  ➢ Not a practical choice for large datasets, especially Ortho

  ➢ If prices come down, could be used to store vector data

  ➢ Primary purpose of SSD is not storage, but to facilitate inner workings of other devices

Technologies

➢ Portable Hard Drives

  ➢ Very inexpensive

  ➢ Not very durable

    ➢ Sensitive to corruption and data loss

  ➢ Easy for Archives to use

  ➢ Not networked

  ➢ May be used as temporary backup to more robust solution

  ➢ May be used to transfer data to the archives from the field

  ➢ Security risk: drives can walk out the door if not secured

Technologies

➢ Flash Drives, CDs, DVDs

   ➢ Consumer media known to fail

   ➢ Storage capacity is low

   ➢ May be used to transfer data to an archives from the field

   ➢ Security risk: media can walk out the door if not secured

Configuration Options

➢ RAID

➢ Blocks and LUNs

➢ File System Limits

Informational Partner Feedback

➢ What do you know about storage now that you wish you knew earlier?

Storage Options Per Use

➤ Online use for access copies of records

➤ "Dark" storage for preservation copy

➤ Storage needed during active archival processing

➤ Replication and Audit functions between storage units (LOCKSS, IRODS, ACE)

➤ Backups for any of the above

Common Findings

➢ Tiers of Infrastructure

    ➢ Worst case

    ➢ Workable solution

    ➢ Ideal solution

➢ Challenges

    ➢ Procurement

    ➢ Management

    ➢ Sustainability

    ➢ Organizational Policies (Politics)

Informational Partner Feedback

➢ What challenges have you encountered with storage procurement and management (cost, organizational policies, span of control)?