WGBH
DuraCloud Pilot

Audio and video in the cloud

NSDL Annual Meeting
Tuesday, November 2, 2010

Peter Pinch
Director of Technology for Interactive
What is DuraCloud? Storage and more

QuickTime™ and a decompressor are needed to see this picture.
DuraCloud Services

- Duplicate on Upload
- Duplicate on Demand
- Image Server
- Media Streamer
- Bit Integrity Checker
- Bulk Bit Integrity Checker
- Image Transformer
- Bulk Image Transformer
- System Transformer Utility
- System WebApp Utility
DuraCloud Pilot Goals

- **Access**
  - Streaming video
  - Integration with [http://openvault.wgbh.org](http://openvault.wgbh.org)
  - Cost savings?
  - Improved sustainability?

- **Preservation**
  - Uncompressed audio and video storage
  - Cost Savings?
  - Improved reliability?

- **Future Services**
DuraCloud Use Case: American Archive Pilot

- CPB pilot project, 20 stations including WGBH
  - civil rights era and World War II
  - Stations responsible for preservation & hosting

- Preservation
  - 110 hrs of video, 8.5 TB
  - 120 hrs of audio, 150 GB

- Access (streaming)
  - 12GB of H.264 video
  - 4GB of mp3 audio
Preservation using DuraCloud

**Preservation support:**
- File validation
- Replication management
- Administrative access
- Error checking
- File
- migration/transformation
- Monitoring
Access using DuraCloud

Access services:
- Streaming
- File format transformation
- File access collaboration

Digital Access Management system

Open Vault Fedora Repository

WGBH DuraCloud Instance

Access file

Access file streaming

Ingest

Amazon

EMC

RackSpace
Lessons Learned
Sending disks to the cloud

WGBH Delivery to DuraCloud via Hard Drive (.7 TB)
(.7 TB transferred, 1TB drive, ext3 format)

Collect Assets

Create Amazon Manifest

Chunk Assets

Transfer to Shipping Drive

Calculate Checksums

Pack and Ship Drive

Files available at DuraCloud

2 days 1.75 days up to 7 days
Using tubes and wires

WGBH Delivery to DuraCloud via Internet
(.7 TB transferred, 300Mbps pipe)
Costs

- Gathering data
- Sunk costs
  - DAM (including hierarchical storage)
  - Bandwidth (to the cloud)
- Incremental costs
  - Off-line storage
  - Cloud storage
  - Streaming bandwidth (+1 for cloud)
# Cost Comparisons

<table>
<thead>
<tr>
<th></th>
<th>In-house</th>
<th>Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth to storage</td>
<td>n/a</td>
<td>“free”</td>
</tr>
<tr>
<td>Bandwidth for access</td>
<td>$1 per GB transferred</td>
<td>17¢ per GB transferred</td>
</tr>
<tr>
<td>Storage</td>
<td>8.8¢ per GB</td>
<td>15¢ per GB per month</td>
</tr>
</tbody>
</table>
Outcomes

- 5.5 TB of audio & video uploaded
  - Preservation and access files
- Still working with sync tool
- Streaming service works
  - But still need to integrate with Open Vault web site (for access)
The Future

- Complete integration with Open Vault site
- Dealing with file size limits
  - Editing (clipping)
- Transcode services?
  - Proposal with NCSA
- Speech to text?
  - Transcript alignment
- Recommend for American Archive when it moves to preservation phase of project
Questions?

http://openvault.wgbh.org
peter_pinch@wgbh.org