Goldmine.uncc.edu

Local & Commercial Content
Topics

- Context
- Goals
- Technical Framework
- Current State
- Next Steps
Context

- 2012 - Infrastructure 8-10 years old
- Purchases occurred from one-time money
- Digital assets- aging media; multiple uncontrolled locations
- Short-term: server replacement & local area
- 2 Years to investigate plan and migrate
CONCERN

- 80% of drives last 4 years
- >4 years = +12%/year
- Add in human error/departure
Goals

• Affordable >> Utility Bills
• Durability, scalability, reliability, sustainability, security
• Disaster Recovery
• Management - anywhere, anytime
• Began journey thinking about locally owned/created content
• Springer agreed to our e-book principles (commercial content):
  ○ Provision of irrevocable perpetual access and archival rights.
  ○ Allowance for unlimited simultaneous users.
  ○ Freedom from any Digital Rights Management (DRM), including (but not limited to) use of proprietary formats, restricted access to content, or time-limited access terms.
Platform Base
Strong, Serious Community
Islandora on AWS

Loading Dock
Ingest Workflow
Islandora on AWS

- **Closed Loop System**
  - Everything ingested is preserved
  - BagIt Bags are updated on every metadata edit, object touch
  - PREMIS audit logging

- **Checksums on everything**; all objects and metadata are fingerprinted on ingest

- **Constant Checksum Checking**
  - Whole Repository Check Completed Monthly

- **FITS Technical Metadata** (TECHMD)

- **SOLR Full Text Index** ∞
  - All Metadata Indexed
  - eBook Full Text Indexed
    - OCR/HOCR/FULLTEXT
Real World Cost

• Digital Repository **Hardware** (EC2 “r3.2xlarge” Reserved Instance) :
  $6,130 / 3 years

  • **Preservation Storage (Glacier)** :
    1 TB @ $0.004 per GB / month ($4.00 per month, $48.00 per year)

• **Access Storage (Elastic File System)** :
  1 TB @ $0.30 per GB / month ($300 per month, $3600 per year)

• **Backup Storage (S3)** :
  1 TB @ $0.023 per GB / month ($23.00 per month, $276 per year)
Media Type Agnostic

- Large Images
  - TIFF

- Large Audio
  - Oral Histories

- Manuscript Collections
  - Bound, Closely Related Objects

- Electronic Theses and Dissertations
  - Dipping Our Toe Into The IR Water

- Research Datasets, Genomes

- Newspapers

- Disk Images
  - Politician Hard Drives

- Website Archiving (WARC)

- Video
  - Streaming
Next Steps

• Cross Region Failover

• Loading Dock XML Transformations
  - Publisher XML -> Standards-based XML
    - This Week’s Example:
      - University of Michigan Press ONIX -> MARC(XML)

• Scaling
  - Autoscaling
    - OCR, HOCR
  - Triplestore Upgrade (Object Relationships)

• Full Automation
  - Automated Object Repair
  - Outlive Humans
Long Term Viability

• Endowment for Long Term Funding
  
  ○ 100 Years
  
  ○ Working with AWS on long term storage funding models
    ■ Pay Ahead?
    ■ Bank Draft?
  
  ○ Succession of Trustworthy Human Resources
    ■ Archivists
    ■ Programmers
    ■ Administrators
  
  ○ Cost Prediction Is Key
“Libraries must meet publishers where they are.”
-- Charles Watkinson, Director, University of Michigan Press

- e-Procurement
- e-Delivery to Loading Dock
- Automated Metadata Transformations
- Machine Readable Licenses, Fully Automatic Implementation of Afforded Rights
  - Automated Ingest on Machine Agreement/Handshake
- Post-successful ingest: OAI pull or API push from Repository into ILS and Discovery Systems
- Highly Advanced Library/Publisher Systems Integration
- Kumbaya and S’mores

- Thank You!