Building on to the Digital Preservation Foundation at Harvard Library

Andrea Goethals
ABCD-Library Meeting
June 27, 2016
• What do we already have?
• What do we still need?

Where I’ll focus
DIGITAL PRESERVATION SERVICES

• Organizationally within HL Preservation Services
  • Along with Media Preservation Services, Imaging Services, Weissman Preservation Center, Collections Care

• 2014-2020 Digital Preservation Roadmap

• Services
  • Monitor and maintain long-term usability of DRS content
  • Represent user preservation needs
  • Consultations, advice, guidelines
  • DRS oversight and management
WHAT IS THE DRS?

• Harvard-maintained service for digital collections & content for:
  • preservation
    • keep the content safe
    • keep the information usable long-term on modern platforms
  • delivery to users

• A service - not storage or a tool
  • Includes preservation & IT staff actively monitoring the content and systems
  • Includes documented policies, practices & preservation plans
  • Uses technology & systems but these change over time
# DRS ROLES

## DRS Business Owner
- Digital Preservation Services

### Key Responsibilities:
- Preservation & usage policies, strategies
- Manage & communicate about service
- Represent users’ & content’s preservation needs
- Define high-level enhancement roadmaps
- Preservation plans
- Preservation outreach, consulting, guidelines

## DRS Technology Owner
- Library Technology Services

### Key Responsibilities:
- Technology & security policies, strategies
- Manage hardware, software, development
- Bug fixes & enhancements
- System monitoring & scaling
- Refine roadmaps based on resources
- System testing & documentation
- User support & training on systems
KEY POLICIES

• What can be deposited into the DRS
• Who can deposit to the DRS
• Obligations of collection managers
• Responsibilities of DRS staff
• Retention policies
• Discovery & access policies
• Delivery services
• Preservation services

HIGHLIGHTS

• In production since 2000
  • Integrated with Library infrastructure & services
  • Content from over 50 Harvard units
• Supports curatorial management
• Rich set of metadata
• Accepts any format but provides incentives to deposit in recommended "preservable" formats
• Supports HRCI content
• Proven track record
  • Experience with several metadata migrations, maintenance of persistent names, several storage refreshes, planning for format migrations this year
  • Never lost a file!
• Spawned well-used open-source tools (JHOVE, FITS)
WHAT’S IN THE DRS?

• Quantity: 61.3 million files, 196 TB per copy
• Many formats
  • Images, audio, text, digitized books, websites, documents, biomedical image stacks, email and soon video
  • But primarily digitized images and text
WHO OWNS THE CONTENT?

DRS Content by Owner

Year

TB


0 20 40 60 80 100 120 140 160

Does not include content from the Harvard-Google project
Ongoing Enhancements, Upgrades and Project Support
(2+ developers + systems and user support + management)

LARGE PARALLEL PROJECTS

- IMLS-funded DRS Audit
  (1 resident through 5/16)

- Arcadia-funded Easier DRS Deposits Project
  (1 term position through 11/16)

- Arcadia-funded Long-Term Preservation Project
  (2 term positions through 11/16)

- DRS2 Metadata Migration Project
  (1 term position through 3/17)

- Planning for Format Migrations Project

- Format Migration Project

- Migration to Next-Generation DRS Storage Project

- Planning for Next-Generation DRS Storage Project

Have

Need

We are here

6/2016

12/2016
Have
Need
Ongoing Enhancements, Upgrades and Project Support (2+ developers + systems and user support + management)

- IMLS-funded DRS Audit Project (1 resident through 5/16)
- Arcadia-funded Easier DRS Deposits Project (1 term position through 11/16)
- Arcadia-funded Long-Term Preservation Project (2 term positions through 11/16)
- DRS2 Metadata Migration Project (1 term position through 3/17)
- Planning for Format Migrations Project
- Format Migration Project

Prepared for eventual certification as a trustworthy repository; have direction for future improvements
Have a plan for easier & more efficient deposit workflows; more content preserved
Most-requested formats supported by the DRS for preservation and access; more content preserved
Refreshed storage with greater geographic distribution; less preservation risk
All DRS content described in consistent, high-quality, standard ways to support preservation planning
All DRS audio and image files in accessible, modern formats for our users; framework in place for future format migrations

We are here
6/2016
12/2016
BUILDING ONTO THE DRS

• Additional workflows into the DRS
• Additional access from the DRS
• Related infrastructure
• Support for additional formats
ADDITIONAL WORKFLOWS INTO THE DRS

imports from third-party services and vendors

common workflows and tools for processing and depositing born-digital collections

web-based deposits

cleaner deposit packages

automated transfer from Harvard systems

deposit service

Have

Need
ADDITIONAL ACCESS FROM THE DRS

Have

Need

new forms of delivery for sensitive content, emulated content, packaged content

researcher platforms for analysis, annotation, collection-building

online exhibits, digital library

aggregators and peer institutions

usage data, content “health” metrics
What is needed:
• Harvard policies, strategies and guidelines for when to use which
• Transfer mechanisms between them
SUPPORT FOR ADDITIONAL FORMATS

Have

Need

SUPPORT FOR ADDITIONAL FORMATS

DRS

- databases
- spreadsheets
- presentations
- social media
- CAD/CAM
- complex art
- datasets
- computer games
- camera RAW images
- image sequences
- vector graphics
- apps and executables

long tail of formats in collections