Partner to Preserve: Digital Preservation Solutions, Networks, and Collaboration

Liz Bishoff, The Bishoff Group
Tom Clareson, LYRASIS

April 15, 2013

1

Caring for Digital Materials Goals

1. Participants will have a better understanding of the inherent fragility of digital objects
2. Participants will acquire information to help them select preservation formats, metadata, and backup systems for digital objects
3. Participants will be able to identify one or more actions that can be taken to improve their institution's digital preservation efforts

2

Caring for Digital Materials Sessions

Overview of digital preservation	Lauren Goodley	Tues., April 2, 2013 2:00 – 3:30 EDT
Convert it to preserve it: Digitization and file conversion	Jacob Nadal	Thurs., April 4, 2013 2:00 – 3:30 EDT
Describe it so you can find it: Metadata, finding aids, and asset management	Danielle Plumer	Tues., April 9, 2013 2:00 – 3:30 EDT
Practice safe archiving: Backups, copies, and what can go wrong	Jefferson Bailey	Weds., April 10, 2013 2:00 – 3:30 EDT
Partner to preserve: Digital preservation networks and collaboration	Liz Bishoff and Tom Clareson	Mon., April 15, 2013 2:00 – 3:30 EDT

We are here:

3

Session Outline

- Why we need digital preservation solutions
- Collaborative approach to digital preservation
- Collaborative digital preservation services/solutions
- Use of multiple solutions
- Weighing your options
- Planning for the future

4

Getting to know you

- How many of you have less than 1, 1-5TB of data? 5-15TB, 15+, don't know?
- How many of you back-up your files at least weekly? Monthly? Annually? Don't know?
- How many locations do you store your copies in?

5

Why we need digital preservation solutions: a quick review

6

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

Digital Preservation: Why is it an issue?

- Technological Changes
  - File formats change
  - Approaches—local vs. distributed
- Organizational challenges
  - Resources
  - New partnerships and expectations
  - Long-term access to digital resources
- System architecture
  - Use non-proprietary systems
  - Use standards that allow content that can be migrated

7

Backup vs. Digital Preservation

Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. A backup is short-term data recovery solution following loss or corruption and is fundamentally different to an electronic preservation archive.
- JISC. Digital Preservation: Continued Access to authentic digital assets.

8

Digital Preservation Risks

- Can be technical
- Physical
- Organizational
- Socio-cultural
- Legal
- Financial
- Political
- Contractual
- Force majeure

9

The Impact of Risk on Digital Collections

- Impact on repository staff or public well-being
- Impact of damage to or loss of digital assets
- Impact of statutory or regulatory breach
- Damage to organization's reputation
- Damage to financial viability
- Deterioration of product or service quality
- Loss of ability to ensure digital object authenticity and understandability is ultimate expression of impact

10

Policies Demonstrate Commitment

- What commitments have your organizations made?
  - Does your mission statement include preservation of digital collections?
  - Have you included digital resources in your collection development policy?
  - Does your emergency plan incorporate digital collections?
  - Does your deed of gift include rights to allow modifications supporting digital preservation?
- How long do you intend to make your digital collections available?

11

Policy Poll

- Digital holdings are incorporated into
  - Mission statement
  - Long range plan
  - Collection plan
  - Emergency plan
  - Exhibits
  - Preservation

12

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

Questions from the audience

13

Why collaborate?

14

15

Digital preservation and collaboration

- Collaboration is required to meet our public purpose
- Collaboration may be at the international, national, state, local level
- Geographic distribution is a key strategy
- Share the risk
- Expand our capacity to implement a digital preservation repository
- Two core principles in survivability/preservation
  - The more copies, the better
  - The more locations, the better
- The more heterogeneous, the better

16

Collaboration poll

- How many of you participate in collaborative digital initiative(s), either regional, statewide, or national?
- Are you currently participating in
  - Portico
  - LOCKSS
  - HathiTrust
- Are you a member of the National Digital Stewardship Alliance?

17

Collaborative digital preservation: services and software

18

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

Today's Digital Preservation Services/Solutions

19

Today's Services and Solutions

- Not all options are equal
- Not all options are suitable for all materials
- Multiple solutions will be needed for materials at most institutions
- "Long term" is a long, long time…all solutions must be based in flexibility and extensibility to succeed
- The perfect is the enemy of the good—cultural memory can't afford for us to wait for the perfect!

20

Who does what?

- Digital preservation services
  - Archivematica
  - Chronopolis
  - DuraCloud integrates with CDM, D-Space, Fedora
  - Private LOCKSS Network
  - OCLC Digital Archive
  - Tessella

- Digital/institutional repository software
  - D-Space
  - Fedora Digital
  - Islandora/Fedora
  - Content DM
- AND then there's Hathi Trust

Categories of Service

- Data Preparation/Management services
  - Regardless of what service(s) you engage, you will have data preparation work to do in house, at least at this point.
- Preservation services
  - In house
  - Centralized
  - Distributed
    - Back-ups distributed
    - Heterogeneous storage options
    - Distributed as overall approach

22

Categories of Service

- Data Preparation/Management services
  - Digitization services
  - Existing tools that are helpful and may be used in the preservation process
    - JHOVE2
    - DROID
    - PREMIS-in-METS (PIM)
   - Various microservices (CDL, LC, UNT, FIT, etc)

23

Preservation Services/Solution

- Locally managed solutions
  - Solution that can be installed in a local environment
  - Tessella, Ex Libris Rosetta
- Centralized/hosted solutions
  - Solution that is housed in one main location but used by partners/clients/members
  - OCLC Digital Archive, PORTICO, Tessella's Preservica, DuraSpace's DuraCloud, HathiTrust
- Distributed Services
  - Definition still contested
  - Three main "flavors":
    - Back-ups distributed
    - Heterogeneous storage options
    - Distributed as overall approach

24

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

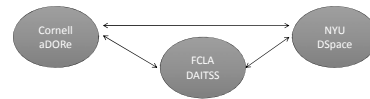## Distributed Preservation Services

- Back-ups Distributed
  - Back ups of preservation repository stored in multiple locations (including cloud servers)
  - Good practice; likely need at least three copies so that if two copies disagree, there is a "tiebreaker" to help establish authentic and viable copy
  - Keeping the back-ups synched is of utmost importance
  - Glacier, DuraCloud
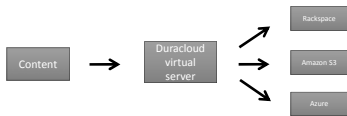

25

Preservation Services

- Heterogeneous storage options
  - Rather than banking on one storage mechanism, place AIPs in multiple preservation repositories
  - Great practice given the new-ness of most solutions
  - Challenging to keep the copies synched
  - TIPR project

26

Preservation Services

- Distributed as overall approach
  - Redundant copies that are regularly (and ideally, in automated fashion) compared to ensure that no file degradation has occurred
  - Can be network-based or handled through less automated (human-based) checks
  - LOCKSS, Private LOCKSS Networks (MetaArchive, PEDALS) Chronopolis
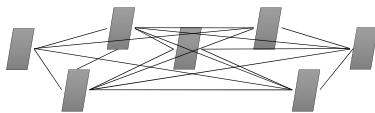
27

Multiple Solutions Approach

- Different materials may require different solutions
  - Born-digital materials vs. Digitized materials
  - Items that are restricted vs. freely available items
  - Unique vs. commonly held items
  - Owned vs. leased items

- For example, think about the differences between the following:
  - Images, E-journals, Datasets, Websites and blogs

28

Questions from the participants

29

Digital Preservation Network (DPN)

- Federated approach built on existing local repositories
- Creating replicating nodes, redundant dark archive
- Objects and metadata are replicated across notes
- Partnership of higher education institution—54 universities, CDL, Texas Digital Library, American Council of Education
- Membership fees

30

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

DuraCloud

- Not for profit organization built on merger of Fedora Commons and D-Space Foundation
- Taking advantage of 'Cloud' excitement
- Offers both hosted and locally managed application
- Integrating with Fedora and D-Space
- Pricing--$4500/500gb/year for one cloud storage service additional fees for second cloud, streaming service, etc. (2011)

31

Hathi Trust

- Partnership of research institutions & libraries
- Certified Trustworthy digital repository
- Provides access to 10m+ volumes, 31% public domain
- API's available to support reuse
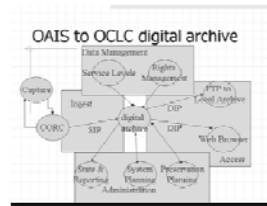- Membership fees

32

LOCKSS Private Network

- Multiple replications, geographically dispersed
- Cross different power grids, climate environments
- Builds off experience preserving e-journals
- Open source software
- Low cost hardware
- LOCKSS membership
- MetaArchive, ADPnet, Pedals, etc

33

OCLC Digital Archive—Hosted Service

- Brand recognition
- Price decrease in fall--$500/100gb/year; $200/100gb if depositing 1TB
- 10 years experience with digital archive; 40-50 customers
- Integrated with CONTENTdm; 1000+CDM customers

34

Tessella's Preservica—Hosted service

- Built on Tessella's Safety Deposit Box (SDB), OAIS compliant system
- Cloud based
- Active preservation
- Upload content from variety of systems
- Full search, browse, download
- Content hosted on Amazon Web Services
- Currently on stored in one geographic location
- Annual subscription fee

35

Central questions to ask of any model/service

- What type of preservation repository do you need? Open or closed or dim?
- How does it work? Are there checks-and-balances?
- What is the approach(es) to preservation? Bit-level, migration, emulation?
- What is required to prepare data and who is responsible for doing it?(How must data be structured for ingest?)
- What formats are supported? (restricted or open?)
- What metadata is required? What can be automatically populated vs what is required by depositor? Is metadata at the item or collection level?
- What will your organizations role? How actively engaged will you need to be? What technical expertise will be required? How will you build in-house expertises?
- How is geographic distribution addressed?
- What is the exit strategy? What if you want out or if they end the service?
- Cost

36

Caring for Digital Materials Webinar 5:
Partner to preserve: Digital preservation
networks and collaboration

Instructors: Liz Bishoff and Tom Clareson

Make sure you

- Evaluation the following areas
  - Safety—Is the system working and has it been evaluated?
  - Security—"dark," "dim," or "open," which do you need?
  - Longevity—Is the system flexible, extensible, and sustainable
  - Breadth—Does the system accept the formats you need to store?
  - Authenticity—Can the system authenticate digital objects?
  - Reporting—Does the system provide ongoing reports on collections
- Assess technical infrastructure (yours and theirs)
  - In-house vs. 3rd party
  - Inventory your technical skills
  - Understand your ability to meet your long term commitment

37

Remaining questions on solutions/services

38

Next steps for collaboration

39

So what's next in collaborating?

- Identify potential partners
- Establish the long term commitment to digital collections
- Establish a planning team
- Determine roles and responsibilities
- Define leadership & governance
- Determine funding strategies
- Explore technology options
- Define content
- Identify metadata requirements
- Develop a digital preservation plan

40

Questions and Discussion

- Please contact us with follow-up questions:
  - Tom Clareson, LYRASIS, tom.clareson@lyrasis.org
  - Liz Bishoff, The Bishoff Group, liz.bishoff@gmail.com

41