

Digital Sustainability:

Weaving a Tapestry of Interdependency to Advance Digital Programs

Tyler O. Walters

Associate Director, Technology & Resource Services
Georgia Tech Library & Information Center

Emory Sustaining Digital Libraries Conference
October 6, 2006



UGH!... Another Talk on Sustainability?

- *How are we going to sustain all this digital library activity?*
 - Correct path: Developing the concept of Digital Sustainability
 - Social networking: Organize DL work via interconnectedness
 - My “in progress” thoughts, observations, experiences
 - Diffusion of Innovations Theory – Everett Rogers, 1962
- Suggest a Model
 - Organizational
 - Technology
 - Economic
 - Collections
 - *Underlying Characteristic*:
 - Internetworking of technologies, content, and organizations
- ***MetaArchive Cooperative:**
 - **Case study on DL sustainability**

1st Mode of Sustainability: Organizational

“Strategies /approaches that advance collaborations between org. units or sub-units to perform a program’s functions and achieve a particular goal”

Four Levels of Organizational Sustainability:

1). Single Institution

- Resources (finances/infrastructure) and expertise

2). Consortium

- Institutions create to form a program of digital library activity
- Generates additional value and elements of sustainability that individual institutions cannot generate on their own.
- Sum of the whole (i.e., program) is greater than its parts (i.e., the institution)
- Long-term collaboration between projects, users, sponsors, agencies, other stakeholders is present

3). Non-profit Management Entity

- Further provide strategic guidance and support for organizational sustainability and program development
 - Broader organizational support and direction to consortia
 - Facilitates relationships with other organizations and consortia
 - Provides low-cost, low-overhead conduit to manage fiscal resources

4). National / International Strategic Initiative Partnerships

- Consortium links itself with larger strategic agendas
 - Fosters proper strategic alignment, funding, and additional access to expertise and new knowledge generation
 - Consortium interweaves itself with many insts., consortia, private orgs., government agencies, and expansive strategies
 - Bring to bear a wealth of resources, financial and otherwise, as well as people with a diversity of knowledge, skills, interests

2nd Mode of Sustainability - Technology

“strategies and approaches that advance collaborations to create, disseminate, and maintain technologies”

Open Source Community Model:

- Hope to achieve long-term sustainability, not unlike the model for organizational sustainability
- In the technology model, there are layers of interconnections as well

1). Initial Development Partners

- Give birth to new technology/software
- Nurtured, brought to market, where early adopters utilize it

2). Adopting Developer and User Groups

- Broaden, include new developers from early adopting insts.
- Bring expertise and resources to further technology's development
- Technology stabilizes and matures. Gains critical mass of users.
- Adopting groups of developers and users form

3). Governing Organizations

- Trend: Organizational development around governing and coordinating the developer and users groups
 - like non-profit management entity hosting the consortium in the organizational sustainability model

■ Examples:

- Apache Software Foundation
- DSpace Consortium Inc.
- LOCKSS Alliance
- Fedora Project
- Sakai Foundation
- Internet2 Consortium management of Shibboleth

Intent of technology organizational structures:

- Conduits for innovation, expertise, financial and infrastructural resources to develop new technologies

These and other technology development organizations bear studying as they evolve and attempt to sustain their technologies

3rd Mode of Sustainability - Economic

“relates to revenues / investments necessary to support digital libraries”

...As with earlier modes of sustainability, economic mode matures through successful construction of layers of interdependency

- Economic inputs: Finances, Infrastructure, and Expertise
 - ...all have monetary value
- 1). Individual Institutions and Initial Development Partners Group:
 - Provide a foundation of these economic inputs
- 2). Consortia and Adopting Developer and User Groups:
 - Inputs combine to generate **new ideas** and **new infrastructures**
 - Also, seek funding to apply to existing economic resources

3). Non-profit Management Entity and Governing Organizations:

- bring more partners, projects, consortia together in pursuit of generating funds (and new knowledge) to carry out objectives

4). National and International Strategic Partnerships alignment:

- helps to identify further revenues to infuse programs/projects

Goods and Services:

- In between layers: goods/services that each entity can provide directly to interested consumers to generate additional revenue
 - Consumers: entities that utilize digital library tech. and related svcs.
- Revenues are not meant to meet all costs incurred by the technology developing organization, but rather to provide one of several necessary revenue streams

Sources of Funding:

- From partnerships and other relationships, to fees for goods and services...

all must combine to meet economic needs organizations incur while developing and sustaining digital libraries

4th Mode of Sustainability - Collections

“approaches/strategies used to ensure that defining qualities characterizing information objects and collections persist”

Information resources managed in DLs have three major characteristics:

- 1) **Content**, 2) **Context**, 3) **Structure**
- **Provenance**: refers to social and organizational processes from which the information, data, or records are created and used
 - **Example: data management field - concept of “data provenance”**
- **Authentic**: not changed or manipulated after it has been created or received or migrated
- **Reliability**: refers to the trustworthiness of the content itself
 - Scholars/researchers: need to **verify the genuineness of data**

Ensuring that information is both authentic and reliable, and that provenance is tracked, are inherent qualities that must be sustained

DIGITAL SUSTAINABILITY MODEL

Economic

General Resource Inputs:

- 1) Financial
- 2) Infrastructure
- 3) Expertise

#4 Inputs: National/International Government Agencies & Organizations
 More Resources, Strategic Alignment
 High-Level Communication/Coordination
 New Economic Models / Technology Standards
 International Collaboration

#3 Inputs: Organizations, Projects, Technologies
 Oversee, Guide, Coordinate
 More Strategy / More Resources
 Sustain Innovation (Ideas/Technologies)

#2 Inputs: Add & Combine
 More Finances,
 New Infrastructure,
 New Ideas / Technologies Begin & Broaden

#1 Base of Inputs:
 Some Finances, Infrastructure, Expertise,
 and Initial Technology

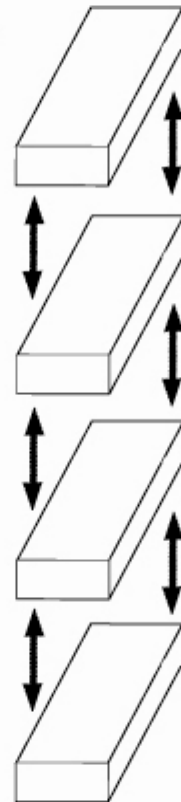
Organization

National / International Strategic Partnerships

Non-Profit Management Entity

Consortia

Single Institution



Technology

National / International Strategic Partnerships

Governing Organization

Adopting Developer/ User Groups

Initial Development Partners

Collections

Sustain:
 Content (Reliability)
 Context (Provenance)
 Structure
 Authenticity



Case Study: MetaArchive Cooperative

■ Project Summary:

- Six partner institutions:
 - Emory - Georgia Tech - Florida State
 - Virginia Tech – Auburn - Louisville
- Collaborate w/ LC/NDIIPP – \$1.4M initial effort to develop cooperative for preservation of digital content, 2004-2010

■ Goals:

1. **Distributed** preservation network infrastructure (LOCKSS)
 2. **Conspectus** of digital content held by the partner sites
 3. **Harvest** a body of most critical content to be preserved (3 TB)
 4. **Cooperative charter** model for collaboration and sustainability
- **Communications:** Internet2: Abilene, Southern Crossroads (SoX), Mid-Atlantic Crossroads (MAX) networks

MetaArchive as Consortia: Cooperative Charter

Level 2: Consortia - Cooperative Charter addresses four areas:

1. Types of partnerships in the MetaArchive Cooperative:
 - **Development Partners** – responsible for developing and testing the preservation network technology and operating a preservation node
 - **Preservation Partners** – responsible for operating a preservation node, ingesting collections from partner institutions, and making the node available for testing
 - **Contributing Partners** – cultural memory organizations that possess digital content to preserve via the MetaArchive Preservation Network. They contribute fees for this service, do not operate node
 - **Sponsored Partners** – cultural memory organizations that own or possess digital content they wish to preserve, but cannot contribute service fees. A Development Partner sponsors the preservation of their digital content

Cooperative Charter Cont'd

- 2). How MetaArchive is organized, governed, and communications
 - Steering Cmte.
 - Content Cmte.
 - Preservation Cmte.
 - Technical Cmte.

 - Partner institution reps can serve terms on the committees, ensuring a broad participation in governance and operations

- 3). Cooperative services to members
 - Digital preservation (network dev./maint., content ingest/retrieval)
 - Format migration
 - Digital collection disaster recovery
 - Digital preservation network consulting
 - LOCKSS services

- 4). A. Technical specs for preservation network
B. Copy of memo of understanding between the six universities

MetaArchive and Educopia Institute

Level 3: Non-profit management entity – three issues:

- Continuing need for financial resources
- Expose MetaArchive to new dig. projects, inform development
- Economically efficient, catalytic structure to bring #1 & 2 about

▪ Educopia Institute:

- Provide oversight to MA Cooperative and other digital projects
- Low-cost, low overhead conduit for digital library, scholarly communications technology projects (Mellon funds)
- Advance cyberinfrastructure needed to drive research, teaching and learning in contemporary digital era
- NSF (2003) and ACLS (2006) Cyberinfrastructure reports:
- Scholarly activity – teaching, research, learning, knowledge transfer via scholarly communications – need rational, strategic cyberinfrastructure

EI: Generate DL technology projects to support mission and goals

Nat'l Strategic Partnership: LC/NDIIPP

Level 4: Access to Resources in Nat'l Digital Preservation Context:

- LC's digital preservation partners and their approaches
- Expertise within LC itself
- Driver to develop MA models: Organizational, Technology, Services
- Facilitate communication on dig. sustainability, gather know-how
- Economic models development for digital library activity
 - Consultant, Paul Courant – Econ. Prof., Univ. of Michigan
- Align programs w/ Nat'l/Int'l initiatives
 - important to generating new revenues
- Nat'l/Int'l approaches are emerging to guide funding (US:NDIIPP)
 - UK: JISC and the DCC
 - Australian Partnership for Sustainable Repositories [APSR]
 - *** Int'l Strategy?**
Digital Preservation Coalition – embed MA's activities overseas*

MetaArchive: Technology Sustainability

MA Technology Path is Clear: Support LOCKSS Alliance

- Step 1) Initial Development Partners:
 - Original LOCKSS Development Team, Partners: Emory

- Step 2) Adopting Developer and User Groups:
 - MA as technology adopters/adapters and as user group

 - Georgia Tech / MA as early adopter and developer:
 - Integrate Typed Object Model framework for format emulation
 - Incorporate framework for automated metadata generation
 - Technology Diffusion: LOCKSS Private Networks

- Step 3) Governing Organization:
 - Educopia Institute, new techs, and MetaArchive Cooperative

- Step 4) Nat'l / Int'l Strategic Initiative Partnerships:
 - LC/NDIIP and the DPC (LOCKSS services via DCC?)

MetaArchive: Economic Sustainability

Steps 1,2) Institution, Consortium / Development Partner, User Group:

- Foundation: server rm., labor time, expertise, collections, network
- Expand members: more resources, interest in developing models

Step 3) Non-profit Entity / Governing Org.: Educopia Institute

- Revenues from synergies w/ new partners, related projects

Step 4) Nat'l / Int'l Strategic Partnerships: LC/NDIIPP, Others?

- **Funding:** \$650k (2004-07) – Servers, Project Staff (tech & non-tech)
- **Economic models, Nat'l initiatives alignment:** UK/JISC & Aus./APSR

Revenues from Service Fees:

- Content ingest/retrieval, format migration, digital collection disaster recovery, digital preservation network consulting, LOCKSS services
- Fees contribute to network maintenance
- Not in this phase yet, difficult to estimate amount of revenues

MetaArchive: Collections Sustainability

"...approaches and strategies used to ensure that the defining qualities of information objects/collections persist"

Early steps to sustain collections' original character:

- LOCKSS' built-in routines for checking a file's technical integrity
- Incorporate format emulation technology
- Record contextual and technical metadata

▪ Authentic nature of digital collections / Reliability of content

"a digital preservation archive consists of digital originals and digital derivatives resulting from format migration" (Jantz & Giarlo, 2005)

- Format emulation tools - Typed Object Module system (Ockerbloom)
- Generate current renderable digital versions w/o altering original coll.

▪ Trace handling / use over time, across technologies

- Metadata: MA Conspectus – track changes, how/when/why occurred
 - "Custodial History," "Creator,"
 - "Format Characteristics," "Accrual Period,"
 - "Accrual Periodicity," "Manifestation"

MetaArchive Digital Sustainability Model

Economic

General Resource Inputs:

- 1) Financial
- 2) Infrastructure
- 3) Expertise

#4 Inputs: National/International Government Agencies & Organizations
 More Resources, Strategic Alignment
 High-Level Communication/Coordination
 New Economic Models
 International Collaboration
 NDIPP Funding
 DPC Projects

#3 Inputs: Organizations, Projects, Technologies
 Oversee, Guide, Coordinate
 More Strategy / More Resources
 Sustain Innovation (Ideas/Technologies)
 Mellon Funding

#2 Inputs: Add & Combine
 Resource inputs from six universities
 New Infrastructures
 New Ideas / Technologies begin and broaden
 NDIPP Funding

#1 Base of Inputs:
 Universities' Infrastructure,
 People & Expertise,
 Initial Technology

Organization

LC / NDIPP
 DPC



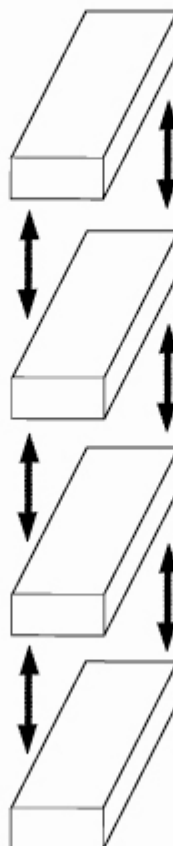
Educopia
 Institute



MetaArchive Cooperative
 (GT, FSU, Auburn, VT,
 Louisville, with Emory)



Emory University



Technology

LC / NDIPP
 DPC

Educopia
 Institute
 (plus, expand
 MetaArchive
 membership)

MetaArchive
 Cooperative,
 i.e. Georgia Tech,
 other developers

Stanford / LOCKSS
 Development Group

Collections

Sustain:

Content (Reliability)
 Context (Provenance)
 Structure
 Authenticity

Services:

Preservation
 (Content ingest /
 retrieval)

Format migration

Digital Collection
 Disaster Recovery

Digital Preservation
 Network Consulting

LOCKSS services



Next Steps / Summary

Keep networking!

...Branch out at each level of organization-building

...Reach more institutions, organizations, consortia, partners *globally*

- “In the flat world, more and more business will be done through collaboration...the more the flattening of the world connects all the knowledge pools together, the more new specialties will be spawned, and the more innovation will come from putting these specialties together in new and different combinations.”

-- Thomas L. Friedman, author, The World is Flat, 2006

MetaArchive work should result in a dynamic and developing organization that productively addresses distributed digital preservation networking issues

Thank You for Your Time!

- Feel free to contact me at:

Tyler Walters
Associate Director,
Georgia Tech Library & Information Center

- Tyler@gatech.edu
 - 404-385-4489