What to Retain: a framework for managing change in the library organization

Constance Malpas
Program Officer, OCLC Research
malpasc@oclc.org
Reorganizing the Legacy Print Collection

Matej Krén “Book Cell” Installation at Centro de Arte Moderna, Lisbon (2006)

An illusion of imprisonment

artfully positioned looking glass

(It only looks infinite)
Shift to digital has transformed scholarly landscape, yet *academic library operations still dominated by print paradigm*

Format migration has introduced new levels of complexity into collection management as the *scholarly function of print is revised*

Decisions about what to withdraw, what to retain are fraught with *uncertainty about future of the library mission*

For books, especially, a *fear of loss to academic reputation*
E-Formats: Increase in Research Productivity?

Session length & gateway access

Deep log analysis of ScienceDirect

Mean session length (seconds)

Bangor 6.4%
Swansea 12.7%
Strathclyde 21.4%

Aberdeen 21.6%
CEH 17.0%
Manchester 27.1%

Rothamsted 15.2%
Edinburgh 34.8%

Cambridge 35.0%
UCL 36.3%

Journal spend, use & research outcomes

... a correlation between e-format consumption and institutional research reputation

Article downloads: 100
Research papers: 636
PhD awards: 622
Research grants and contracts: 21,956

1 million downloads:
2 million downloads:
3 million downloads:
4 million downloads:

Source: (UK) Research Information Network E-journals: their Use, Value and Impact (2009)
Framework for assessing preservation risks, proposes criteria for identifying print journals suitable for withdrawal

- optimal number of copies (2 - 4 in dark archives)
- reliability of digital access (quality, business continuity)
- Image-intensive titles an excluded class (retain in print)

Print as ‘back-stop’ to digital preservation

Retention horizon of 20-100 years, depending on digital preservation status

Decision support tool for JSTOR titles
Investment in Academic Print Collections

Academic Library Expenditures on Purchased and Licensed Content

E-book Margin is Increasing

$169.5M in 2009

$9.3M in 2004

Source: American Association of Publishers
Shift in Pattern of Library Investment

US Academic Library Spending on Selected Categories
as a % of Total Library Expenditures


Declining library investment in preservation
Shared Infrastructure: Journals v. Books

- E-Journals Committed
- E-Journals Preserved
- E-Books Committed
- E-Books Preserved

Margin of confidence?

Source: Portico, Growth of Archive
Dematerialization of the Scholarly Record

Scholarly journals: ~26,000 titles in 2010
i.e. refereed academic journals in
Ulrich’s knowledge-base

Est. 80-90% titles online (Cox, 2008)

ARL aggregate collection: ~50M titles in 2010
i.e. titles held by one or more ARL member library

Est. 6-7 million (12-14%) titles digitized
(extrapolated from analysis of Hathi archive and based on current estimates of 12 million volumes scanned by Google, February 2010)
Premise: emergence of large scale shared print and digital repositories creates opportunity for strategic externalization* of core library operations

- Reduce costs of preserving scholarly record
- Enable reallocation of institutional resources
- Model new business relationships among libraries

* increased reliance on external infrastructure and service platforms in response to economic imperative (lower transaction costs)
Key Findings

• Scope of mass-digitized corpus in Hathi is already sufficient to replace at least 20-30% of most academic print collections
  • Ratio of replaceable inventory independent of collection size
• Most content also held in trusted print repositories with preservation and access services (CRL, UC Regional Library Facilities, ReCAP, Library of Congress)
  • Distribution of resource still suboptimal for shared service model
• If limited to titles in the public domain, shared service offering may not be sufficient to mobilize significant resources
  • Fewer titles, smaller audience: demand is low
Hathi Growth Trajectory - 12 months

Equal in size to median ARL collection (2008)

Equal in scope to very large ARLs (Columbia, Washington, etc)

Equal in scale to LoC?

NB: average holdings per book (title) in WorldCat = 11

Data current as of February 2010
Hathi Trust: Subject Distribution

N=3.2 million titles

Humanities content (literature, history) dominates - presages shift in scholarly practice?

Data current as of February 2010
Distribution by Date of Publication

N=3.2 million titles

>75% of titles in repository published after 1949;
~50% of titles published since 1976
~10% of titles published since 2000

A recent corpus, hence likely to be more broadly relevant to scholars

Date of Publication

Data current as of February 2010
Copyright Status: What Counts?

Volumes in Hathi Library

- 4,521,058 (85%)
- 797,833 (15%)

N=5.3M volumes

Titles in Hathi Library

- 2,859,380 (89%)
- 369,433 (11%)

3,208,813 (85%)

Optimistically, additional copyright determination on orphan works might increase yield by ~600K titles

Based on Hathi profile February 2010
Distribution by WorldCat Library Holdings

N=3.2 million titles

Collective priority

Sweet spot for shared service offering?

Local mandate

Commercial viability

Data current as of February 2010
How Much is Enough?

• If limited to titles currently in the public domain, average academic research library might regain *space equivalent to ~2% of local collection* (based on WorldCat holdings)

• Since public domain collections (excepting government documents) typically not growing, *replacement value a ‘one time’ proposition*

• Roughly equivalent to median annual growth rate in ARL libraries (~2% based on volume count); *at best, enables steady-state for a single year*

*Public domain corpus inadequate to mobilize large-scale shift in library resources*
If Scope is Expanded to In Copyright Titles...

Data current as of February 2010

*R Spheres are scaled to size of institutional collection based on WorldCat holdings
Why (Re) Organize Now?

- Uncertainties about outcome of GBS settlement should not hold us back
  - Many (majority?) of print books currently represented in Hathi are *low-use titles for which aggregate demand can be met with reduced inventory*, even without a licensed provision
  - There is sufficient redundancy to enable *space savings for a significant number academic libraries*; adequate scale
  - By progressively increasing reliance on shared print collections, *libraries create economy in which further externalization becomes possible* and shared asset gains in value
  - Increased confidence in long-term preservation will enable broader base of institutions to participate in licensed offering, *increasing library negotiating power*
Recycling Some Ideas about Sustainability

Kristian Bjornard *Principles of Sustainability (and where they come from)* MFA thesis installation, Maryland Institute College of the Arts, 2009
Overexploitation of common-pool resources (‘tragedy of the commons’) is not inevitable.

Multi-institutional ownership of non-commercial assets is viable and may increase sustainability.

Cooperative governance can be modeled scientifically.

E. Ostrom, *Governing the Commons* (1990)
Can CPR be Applied to Libraries?


Figure 1. Types of Goods

<table>
<thead>
<tr>
<th>EXCLUSION</th>
<th>SUBTRACTABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>difficult</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>public goods</td>
</tr>
<tr>
<td></td>
<td>sunset</td>
</tr>
<tr>
<td></td>
<td>common knowledge</td>
</tr>
<tr>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>common-pool resources</td>
</tr>
<tr>
<td></td>
<td>irrigation systems</td>
</tr>
<tr>
<td></td>
<td>libraries</td>
</tr>
<tr>
<td></td>
<td>toll or club goods</td>
</tr>
<tr>
<td></td>
<td>day-care centers</td>
</tr>
<tr>
<td></td>
<td>country clubs</td>
</tr>
<tr>
<td></td>
<td>private goods</td>
</tr>
<tr>
<td></td>
<td>doughnuts</td>
</tr>
<tr>
<td></td>
<td>personal computers</td>
</tr>
</tbody>
</table>

[Yes]
Empower ‘rational appropriators’ (regional and national consortia) to undertake systematic redistribution and rationalization of low-use monographic collections. Efforts underway in WEST, CRL, CIC etc.

Systematically assess carrying capacity of aggregate resource, i.e. system-wide supply/demand dynamics.

Leverage OhioLINK and other findings.

Monitor change in demand over time; enjoin participants to act as monitors.

CRL audit role might be extended.

Adopt contingent strategies for print preservation.

Embrace de-sacralization of codex.

---

Table 3.1. Design principles illustrated by long-enduring CPR institutions

1. Clearly defined boundaries: Individuals or households who have rights to withdraw resource units from the CPR must be clearly defined, as must the boundaries of the CPR itself.
2. Congruence between appropriation and provision rules and local conditions: Appropriation rules restricting place, technology, and/or quantity of resource units are related to local conditions and to provision rules requiring labor, material, and/or money.
3. Collective-choice arrangements: Most individuals affected by the operational rules can participate in modifying the operational rules.
4. Monitoring: Monitors, who actively audit CPR conditions and appropriator behavior, are accountable to the appropriators or are the appropriators.
5. Graduated sanctions: Appropriators who violate operational rules are likely to be assessed graduated sanctions (depending on the seriousness and context of the offense) by other appropriators, by officials accountable to these appropriators, or by both.
6. Conflict-resolution mechanisms: Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials.
7. Minimal recognition of rights to organize: The rights of appropriators to devise their own institutions are not challenged by external governmental authorities.
8. Nested enterprises: Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

E. Ostrom Governing the Commons
Where to Start?

Actively seek to replace low-use print inventory with reliance on digitized and shared print collections; shift economic model toward cooperative management

- Low-risk public domain titles; institutional risk tolerance will dictate whether regional print copy is needed
  - ~370K titles in Feb ‘10; approx. 250K (67%) held by >9 libraries
- In-copyright digitized monographs already in large-scale stores, for which there is adequate duplication to create a market for service
  - e.g. .5M titles held in UC SRLF and by >99 libraries
What to Retain (locally)

Distinctive institutional assets that demonstrably contribute to university’s research mission

Print monographs already digitized and in copyright, for which aggregate supply is relatively low (<10 to 25 libraries)

*ongoing demand will indicate whether long-term local stewardship is a logical choice and where relegation is advantageous*

*Neither scarcity of supply (‘uniqueness’) nor present ownership are reliable indicators of scholarly value*
Academic print: it’s not the end . . .

but it’s no longer the means

Ongoing redefinition of scholarly function and value of print

will entail some loss

and some gain in library relevance

“Archive of the available past” by Joguldi
Abandoned books at the Detroit Central School Book Depository (6 May 2009)