Preservation Reformatting—
Use, Sustainability and Affordability
Stephen Chapman
Weissman Preservation Center
Harvard University Library

March 27, 2003

Harvard Practices:
Multiple Technologies and Services

- Film
  - 35mm microfilm, film negatives, slides, photographic prints
- Digital
  - b/w, grayscale, color images; full text (OCR); metadata
- Hybrid (microfilm + digital) (film/print surrogate + digital)
  - text: 35mm microfilm, 1-bit digital images (and OCR)
  - pictorial: film or print surrogate, 24-bit digital images
- Hybrid (microfilm + digital) + Direct Digital
  - 35mm microfilm and 1-bit (b/w) digital images: all pages,
    + grayscale or color (direct) digital images: selected pages

Preservation & Access Mission

Perpetuate use.

"Preservation … shall not come at the expense of usability."

… goal is to perpetuate access to that portion of source materials identified as essential to continued usefulness for stated purposes.
Weissman Preservation Center

Reformatting Infrastructure:
Time Capsule

To preserve information content…

- Production
  - Make copies
- Acquisition
  - Move copies from production to storage
- Storage
  - Maintain copies and environments
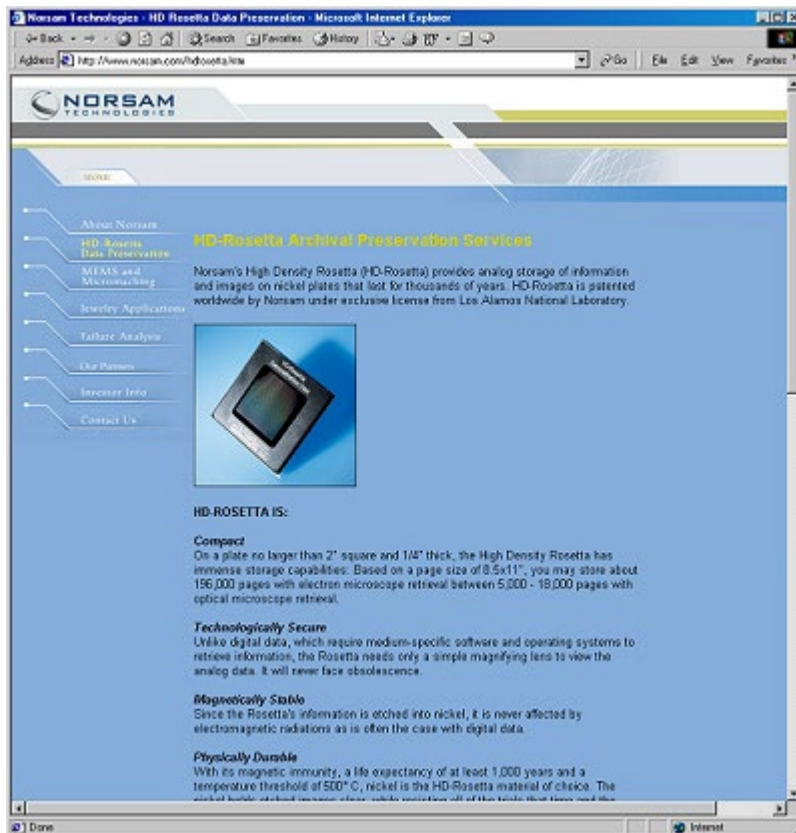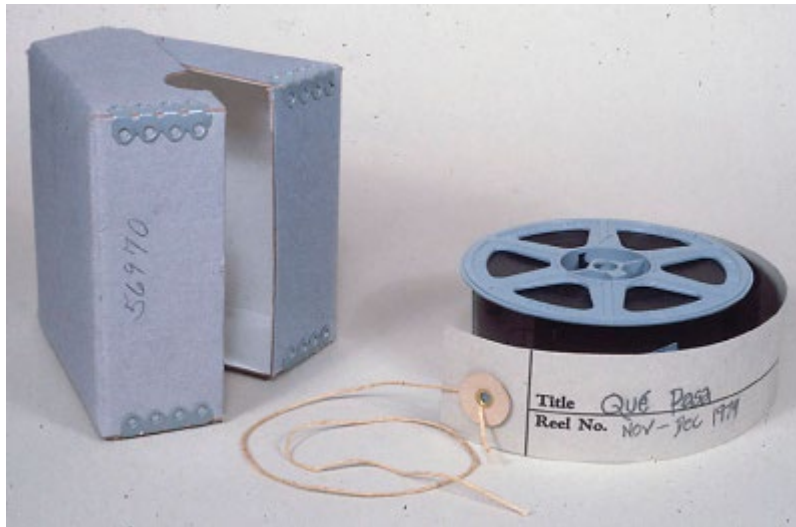
Reformatting Infrastructure:
[No] Time Capsule

To preserve information content...

- Production
  - Make copies
- Acquisition

- o   Move copies from production to storage
- Storage
  - o   Maintain copies and environments

Reformatting Infrastructure:
Library

To preserve use of information content…

- Production
- Acquisition
- Storage
- Discovery
  - o   Catalogs
- Delivery
  - o   Reader/printers, slide projectors, networked PCs

Two Contexts of Usability

- Usable by machines
  - o   film duplicators, enlargers, reader/printers
  - o   image processing software, web browsers
- Usable by people
  - o   preservation managers and technical specialists
  - o   scholars and other users

Analog, Digital or Hybrid?

Users and use requirements

- Owner, to manage risk of loss
  - o   copy mss for preservation backup, not distribution
- Librarian, to build collection
  - o   491 reels duped in FY02, including 63 titles by NLM to fill gaps
- Users, to "access" and "study" item(s)
  - o   definitions dictate delivery format

Affordability: Pragmatic Solutions

- Production
- Acquisition and storage
- Discovery
- Delivery
- Standards and quality metrics

- Centralized repositories
- Scalable delivery services that meet user needs
- Maximum intervals between interventions
    - When obsolete, you must repeat (reformatting).
- Considered uses of production technology(ies)
    - Masters
    - Deliverables

Obsolescence

- Industry-created
    - Product discontinuation (mfrg and/or support)
        - Example: KODAK EKTAPAN Film
        - Example: color-matched printing from transparency film
    - User-created
        - Change in expectation: abandonment of format (service model) or device
        - Example: shift from reserves (photocopies) to e-reserves (digital) to distribute course materials
        - Example: shift from PCs to PDAs

Role of Analog or Digital

Meet requirements for production, storage and delivery in most affordable way

- Production
    - Once-per-? costs for masters: optimize for automated production of range of deliverables
    - Periodic costs for delivery
- Storage
    - Annual costs, regardless of use
- Delivery
    - Costs largely determined by market factors: users/user needs, availability of technology

Harvard Infrastructure

[Library Preservation Web Page](#)

Imaging Equipment Photographs
[1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#)

[Harvard Depository Web Page](#)

[Digital Repository Service Web Page](#)

Harvard Products

Product Specifications Tailored to Project

- Formats and versions
  - Analog and digital
    - Standards, best practices, funding & repository policies
    - Required output(s) [for use] in near-term
- Technical specifications and quality control
  - Analog
    - Conventions for "preservation": standards, best practices, funding policies, quality requirements
  - Digital
    - No conventions for "preservation": informed by extent of vendor services, empirical review of samples, price

Production Strategies: Text

- Filming of bound and unbound materials
  - Use requirements satisfied with analog copies
- Filming of bound volumes: scanning of film
  - Retain bindings, 1-bit output, direct bound scanning not yet viable for high production
- Direct digitization of print
  - 8-bit, 24-bit output
- Autofeed scanning of brittle pages (or photocopies)
  - Multiples, 1-bit output, film quality not uniform
- Legacy microfilm scanning -- *exploring*
  - Depends upon quality of film

Production Strategies: Photographs

- Direct digitization of prints, film, plates
  - HUAM exclusively digital
- Photographic services
  - Film masters
  - Darkroom (chemical) b/w prints
- Digitization of legacy film
  - Post processing varies according to budget
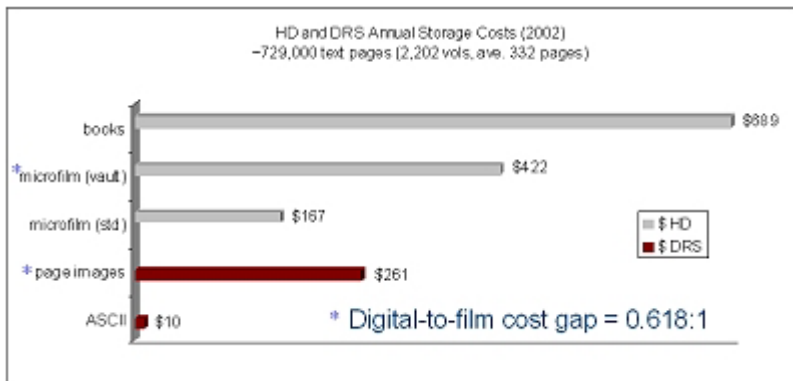
- o File sizes range from 15-200 MB for color work…

Harvard Prices
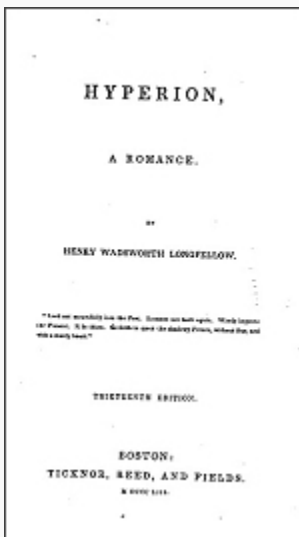**repository storage, (some) delivery**

Repository Storage Annual Charges, by Size

- Harvard Depository
    - o $3.91 per Billable Square Foot (standard)
    - o $9.85 per Billable Square Foot (film vault)
- Harvard Digital Repository Service
    - o $5.00 per GB (1,024 MB)

Text Storage (Masters):
Harvard Repository Costs Compared

HD and DRS Annual Storage Costs (2002)
~729,000 text pages (2,202 vols. ave. 332 pages)

- books $689
- *microfilm (vault) $422
- microfilm (std) $167
- *page images $261
- ASCII $10

$ HD
$ DRS

* Digital-to-film cost gap = 0.618:1

Scanned Images of Henry Wadsworth Longfellow's *Hyperion*

HYPERION,

A ROMANCE.

HENRY WADSWORTH LONGFELLOW.

THIRTEENTH EDITION.

BOSTON:
TICKNOR, REED, AND FIELDS.

BOOK THE FIRST.

CHAPTER I.

THE HERO.

In John Lyly's "Endymion," Sir Topas is made to say : "Dost thou know what a poet is ? Why, fool, a poet is so much as one should say, — a poet !" And then, reader, dost thou know what a hero is ? Why, a hero is as much as one should say, — a hero! Some romance-writers, however, say much more than this. Nay, the old Lombard, Matteo Maria Bojardo, set all the church-bells in Scandiano ringing, merely because he had found a name for one of his heroes. Here, also, shall church-bells be rung, but more solemnly.
The setting of a great hope is like the setting of the sun. The brightness of our life is gone. Shadows of evening fall around us, and the world seems but a dim reflection, — itself a broader

Black and White Images

Color Images

Text Storage (Masters):
Harvard Repository Costs Compared

HD and DRS Annual Storage Costs (2002)
~729,000 text pages (2,202 vols, ave. 332 pages)

$ HD
$ DRS

microfilm (vault) $422

1-bit images $261   Digital-to-film cost gap = 0.618:1

8-bit images $21,353   Digital-to-film cost gap = 51:1

24-bit images $64,060

Digital-to-film cost gap = 152:1
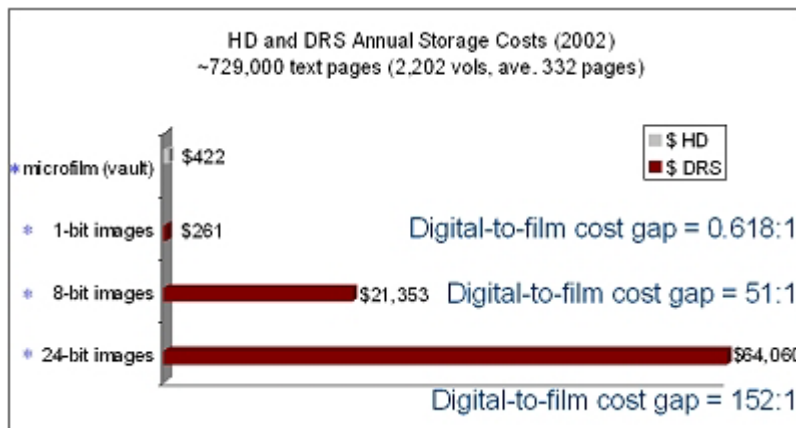
Photographs Storage (Masters):
Harvard Repository Costs Compared

HD and DRS Annual Storage Costs (2002) for 3,000 Photographs

35mm — $5
4 x 5 — $49
PCD (10.7 MB) — $157
TIFF (200 MB) — $2,930

Digital-to-film cost gap = 31:1
Digital-to-film cost gap = 60:1
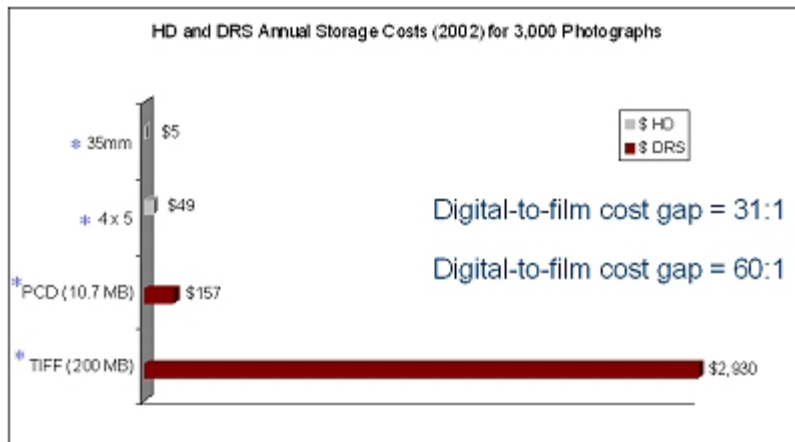
Delivery Services: Newspapers

Microfilm	**Digital**

• Periodic cost
  o $7,800 per reader/printer
• Annual cost
  o $ 950 per reader/printer
	• Periodic cost
	  o $ 49,000 per license (min.) or $ .218 per page + $1,000
	• Annual cost
	  o $ 7,412 per license (min.) or $ .218 per page + $1,000

Analog, Digital or Both?

Cost

- Production + storage + discovery + delivery
  - Factoring use requirements
  - Factoring volume
  - Factoring infrastructure
  - Factoring business models
  - Factoring time (between interventions)

Thank you

Stephen Chapman
Preservation Librarian for Digital Initiatives
Weissman Preservation Center