What is a 'collection' in digital libraries?

Changing:
collection concepts,
collection objects,
collection management,
collection issues

Tefko Saracevic, Ph.D.

… answer not simple

Library collections in the digital age are changing to new forms AND functions

- part is evolution
  - from old to new
- part is revolution
  - very, very new

Effects on libraries are profound and large

ToC

- Emergence & forces shaping

- Traditional and broader concept of collection

- Major characteristics

- Collection management

- Selection criteria

- Issues

Where did it all come from?
a bit of history

- From inside:

- Emergence of digital information systems
  - started with online bibliographic & numeric databases for searching
  - then online catalogs (OPACs) followed by
  - electronic reference sources
  - full-text digital articles & journals
  - e-books
  - **AND:** new forms of scholarly communication, e-publications, & e-archives in many fields

- All on networks; most distributed all over

From outside: What forces are shaping new collection concepts?

– Changing nature of **objects** for digital & hybrid libraries – e-resources, e-books

– New digital **technical & networked environment**

  • among others, many possibilities for **linkages & sharing**

– Changes in **publishing & distribution**

– Evolving **economic models** away from ownership

– Revolution in **access** ways & means

– Changes in users' **information seeking & needs**

Technology, users & use are changing each other

Traditional concept of a 'collection'

- From library point of view
  - sum total of library materials that make a holding of a particular library
    - organized for use

- Major characteristics:
  - objects: (documents) tangible, have a physical presence & location
  - economics: ownership of physical object, but not content
  - location: brick & mortar
  - access: tangible, physical, direct

Based on what a library has & owns

Broadening of the traditional concept

- Collection
  - a set of information resources selected, accumulated & developed for use by a user community or communities
- Major characteristics:
  - objects: tangible & intangible, have a physical presence, but also other media
  - economics: ownership less, licensing more, but also of access, sharing & cooperation
  - location: in-house and distributed (could be housed anyplace)
  - access: direct, but also remote

Digital 'collection'

- ## A set of digital & multimedia information resources
    - ### selected and/or created following defined criteria & policies
        - #### for a defined community of users
    - ### owned, licensed, or freely accessed, but not necessarily accumulated
    - ### with possibility of sharing - consortia
    - ### adapted for networked environment
        - #### distributed as to location
    - ### included access & search tools
    - ### parts with specified access rights
    - ### included portals to other collections

- ## Still evolving, not yet settled

Based on what a library provides access to

Major characteristics:
digital library **objects**

- A variety of intangible objects
  - **From outside:** e-publications, e-journals, e-books, multimedia
  - **From inside:** objects, collections, databases created within library
    1. tangible objects selected from traditional collections & then digitized & organized
    2. objects borne digital, new resources
  - **Gateways**, portals, directories, mashups
    - links selected, organized, evaluated,
    - vertical portals (vortals) in specialized areas, subjects, topics
    - digital bridges to distributed resources
    - mashups - combining data or functionality from two or more sources into a single integrated application

Characteristics (cont):
digital library economics

Main difference: Economics of ownership vs. economics of access

- – ownership declining a lot
  - counting what owned not relevant any more
- – replaced by access
  - licensing of e-publications - many variations
- – sharing costs & access through consortia & similar arrangements
- – charging distributed, subsidized
- Also entering into economics of e-publishing
  - – digitizing part of own collection

Characteristics (cont): digital library location

- Could be on in-house network
- More often, distributed to large number of locations
  - collection becomes collection**s**
  - e-publishers providing 'shelf space' to licensed collections
  - little or no control over distributed resources
- Persistence: a BIG issue
- Dependent on the Internet & networked arrangements

Characteristics (cont): digital library access

- Access arrangements are an integral part of collection
  - navigation, browsing arrangements
  - information retrieval for searching
    - search engines often provided with licensed collections
    - also federated searching (i.e. searching over a number of databases – using licensed software)
  - extended functionality

- Different levels of access
  - subject to policy & license
    - partial access universally open to all
    - full access dependent on belonging to a defined community

- Users usually self-directed

Approaches to collection decisions

- **Collection development** (narrower)
  - process of collection building
  - central issue for libraries & digital libraries caught by this quote:

- **Collection management** (broader)
  - collection development **plus** a broader range of policy, planning, analysis, and cooperative activities
    - including rights management – following copyright laws
  - with digital collections **plus** close involvement of access

  "It does not matter how many books you may have, but whether they are good or not."

  Lucius Annaeus Seneca (3 B.C.-65 A.D.), *Epistolae Morale*

Collection development

- A number of areas and processes:
  - selection
  - acquisition
  - weeding
  - based on criteria reflecting a number of aspects, user and community needs, and institutional mission

- examples of criteria for selection of digital resources from Indiana U

Indiana University Libraries Digital Projects & Services
⚹ Selection Guidelines

Collection management –
much broader & involved activity

> Often referred to as curation

- Includes a number of areas & activities
  - theory and practice of collection policy development
  - selection, acquisition, licenses
  - materials budget allocation,
  - collection analysis,
  - collection use and user studies,
  - staff training & organization
  - preservation, persistence
  - cooperative collection development; sharing resources
  - management of collection space: from physical to virtual
  - rights management & digital rights management in particular

Digital rights management (DRM)

- Administration of rights in a digital environment
  - involves legal, technological, access aspects
  - protect objects from unauthorized use
  - protects copyright use in libraries
  - provides control, statistics of use

- General controversy above & beyond libraires:
  - DRM technologies attempt to control use of digital media by preventing access, copying or conversion to other formats by end users, thus also called digital restrictions management.

Selection criteria for digital resources

- Many traditional criteria remain but with new interpretations

- A number of new criteria have emerged specific to digital nature of resources & access
  - many can be found on the Web, e.g. by Library of Congress

- Judicial & trusted selection:
  - a key value-added contribution by library
  - trust extends to digital collections
  - makes all the difference between a library & other collections

"A Framework of Guidance for Building Good Digital Collections"

(by National Information Standards Organization & Institute of Museum & Library Services)

"A digital collection consists of digital objects that are selected and organized to facilitate their discovery, access, and use. Objects, metadata, and the user interface together create the user experience of a collection."

emphasis added

developed in 2007, but still valid & adhered to

Principles that apply to good digital collections are:
(from Framework)

1. A good digital collection is created according to an explicit collection development policy.

2. Collections should be described so that a user can discover characteristics of the collection, including scope, format, restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity, and interpretation.

emphasis added

Principles (cont.)

3. A good collection is curated, which is to say, its resources are actively managed during their entire lifecycle.

4. A good collection is broadly available and avoids unnecessary impediments to use. Collections should be accessible to persons with disabilities, and usable effectively in conjunction with adaptive technologies.

5. A good collection respects intellectual property rights.

emphasis added

Principles (cont.)

6. A good collection has mechanisms to supply usage data and other data that allows standardized measures of usefulness to be recorded.

7. A good collection is interoperable.

8. A good collection integrates into the users own workflow.

9. A good collection is sustainable over time.

emphasis added

Web sites & cognitive authority

- Assessing & assigning credibility to Web information a BIG problem
  - sometimes even attribution difficult
    - identity? reputation? qualifications?
- Need to carefully asses:
  - **document, author, institution & affiliation** on criteria of:

    authority; accuracy; currency; objectivity; coverage

- Libraries provide an important service in criteria development & assessments of Web sites
  - important value added

Building digital collections

- Old approaches need to be revised and new devised

- Organizational & economic models VERY different
  - consortia, big deal, sharing
  - BIG change: ownership – not any more
    - now paying for access and when the deal is over no more access – nothing to show

- Measure; no more how much owned, but how much used –
  - e.g times accessed, downloads are major statistics

Consortia

- Library cooperatives emerging as answer to digital collection problems
  - economics of scale kick in
  - power in negotiation for licenses
  - enable sharing of collections
  - regulate access to their communities
- Libraries forced & want to work together by economic realities & technological possibilities
- Example of an international list of consortia by International Coalition of Library Consortia

The "big deal"

- Publishers bundling multiple journals into one package
  - labeled "big deal" in library lingo
  - efficient, but has drawbacks
    - single subscription to the whole basket, wanting some of the included journals or not

- Predominant way of subscription today
  - often through library consortia
    - Rutgers is a member of consortium VALE: Virtual Academic Library Environment

Building third party free collections

- Available from the Web
  - and incorporated in a library
  - content mostly domain specialized
  - what library otherwise does not have, but potentially enriching

- Require own set of practices, policies, & organizational models

- Issues: quality sustainability, scalability, cost-effectiveness, applicability

free collections cont....

- A number of resources are now open access
  - part of open access movement

Directory of Open Access Journals
  - lists over 10,000 journals (free access available)
  - 150+ journals in library & information science

Digital Library of Information Science and Technology
  - subject-based, open access digital archive
  - run by University of Arizona

Facing:
Realities within

- Assortment of forces in play
  - economic constraints
    - libraries are struggling to maintain two libraries: a physical library & a digital library, with resources that are inadequate for both.
  - difficulties in finding, training personnel with new competencies
  - balancing the demands of print and digital materials
  - difficulty in assessment of many digital resources
  - difficulty in assuring persistence
    - for many here today, gone tomorrow

Facing:
Realities without

- Dramatic changes in scholarly publishing & communication – digital scholarship
    - technical advances in digitization are truly revolutionizing the way scholarly information is published, organized, maintained, distributed, & accessed - new world emerging

- Increased competition
    - in provision of various library-like services by non-library institutions & commercial organizations

- Ever changing technology
    - hardware, software, networks
    - getting & keeping competencies

Issues for digital collections

- Traditions long tested in collection development & management:
  - What standards, values & approaches to keep? Modify?
  - What new standards to develop?
- Location of digital collections:
  - Where? does it matter?
- Boundaries: where are they?
    - Forward links - collection? directory?
    - Networks: a gigantic collection?
      - then, is a particular 'collection' meaningful or not to users?

Issues ...

- Creation:
  - How and where to become creator of digital collections?
    - Lines between libraries & publishers blurring
- Access:
  - How to integrate with collection?
 - How to provide, safeguard, & protect access?

Issues …

- Economics:
  - Where and how to adjust?
  - Trade-off? Sharing? Licensing?

- In many libraries (especially academic and research, and even school libraries) acquisition budgets for digital materials are predominating leaving print budgets behind

- Also: as archiving of digital materials grow, so do costs

Conclusions

- The concept of library 'collection' has changed dramatically & will change even more
  - includes many problems - should be viewed as opportunities for libraries to innovate & lead
- How radically will collection management change in the next five or ten years?
- How is it changing on a day-to-day basis right now?

Now that is a library collection!
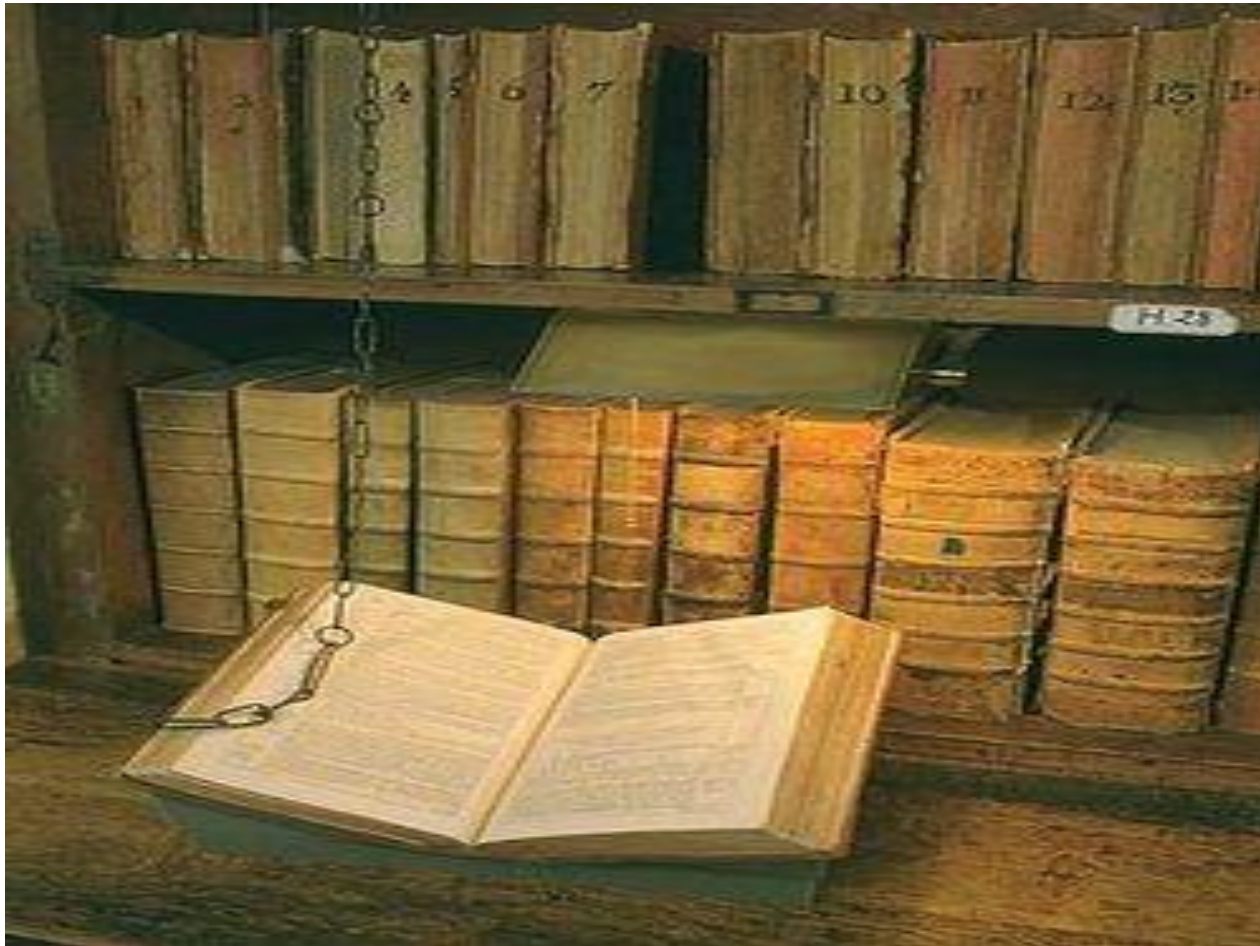
A chained book on exhibit at the Bodleian Library at Oxford University