# Information Institutions

Session 21

INST 301

Introduction to Information Science

"Information Institutions"

- Schools
  - Primary, Secondary, Tertiary
- Libraries
  - Public, Academic, School, Special
- Archives
  - Government, Corporate, …
  - Special collections, Historical societies
- Museums
  - Art, Material culture, Natural history
- Search engines
  - Google, Bing, Baidu, Yandex

"LAM"
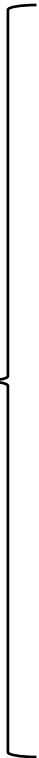
A Collection Development Policy

Customer use is the most powerful influence on the Library's collection. …The other driving force is the Library's strategic plan.

… selections are made to provide depth and diversity of viewpoints to the existing collection and to build the world-class Western History/Genealogy and African American Research Library collections. …

… The Library provides materials to support each individual's journey, and does not place a value on one customer's needs or preferences over another's. …

Materials for children and teenagers are intended to broaden their vision, support recreational reading …

Access models

- Ownership ("just in case")
  - Unlimited use for an unlimited period
  - Right of first sale vs. license restrictions

- Subscription
  - Unlimited (or limited) use for a defined period
  - Single vs. multiple users

- Pay-per-view ("just in time")

Zipf's Law
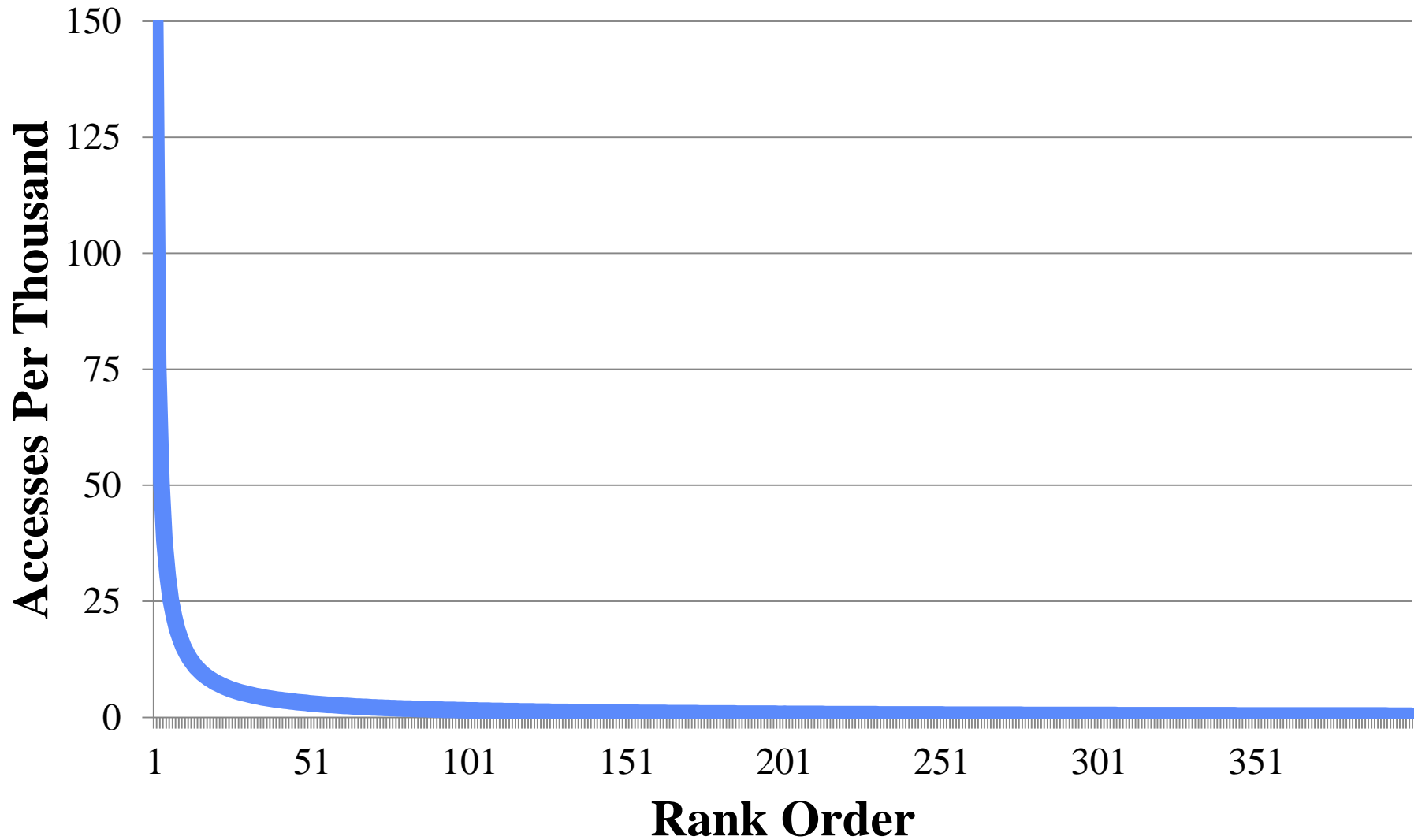
Physical Storage

- Segregate by:
  - Users (e.g., Chemistry library)
  - Type (e.g., audiovisual materials)
  - Usage frequency (e.g., offsite storage)
  - Size (e.g., folios)
- Arrange in a way that facilitates access
  - Topical shelf order (e.g., Dewey Decimal System)
- Foster preservation
  - Environment (temperature, humidity, light)
  - Access controls (closed stacks, gloves, …)

High-Density Shelving

http://www.kmhsystems.com/high-density-storage.html

Compact Storage Robot

Kyushu University, Japan

Access Points

- Originally designed for card catalogs
  - One card for every "authorized" access point
- Four types "dictionary" catalog access points
  - Title (uniform titles)
  - Author (name authority)
  - Subject (controlled vocabulary)
  - Series
- Other things can serve a similar purpose
  - Call number (shelf order)
  - "Keywords" (full-text search)

Library of Congress Subject <u>Headings</u>

- Controlled vocabulary for subject access points
  - Most commonly applied to books and serials

- Used when a subject describes $\geq$20% of the work

- Choose the most specific appropriate headings
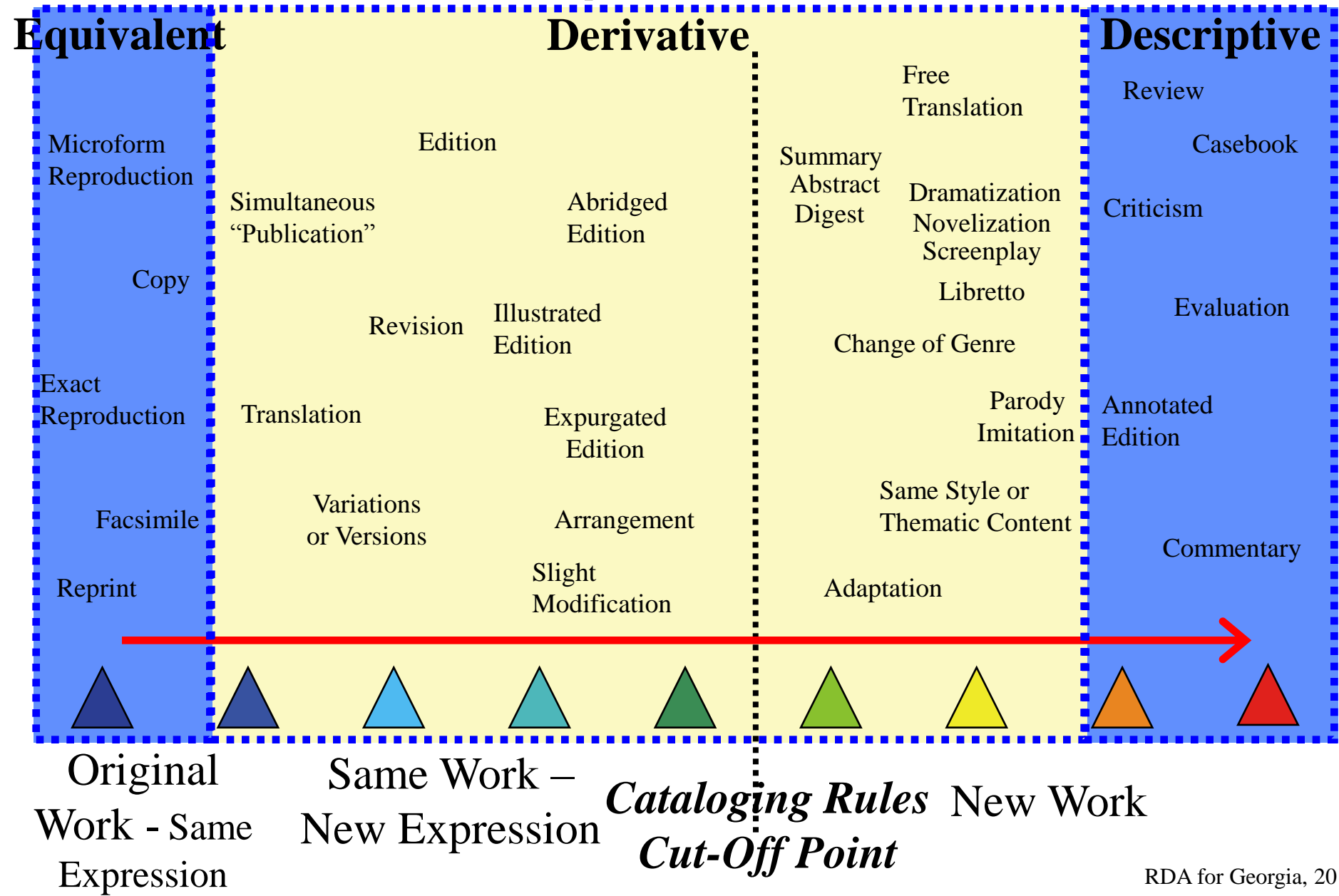 - But if more than 3 subtopics, choose a broader heading

A Framework: FRBR
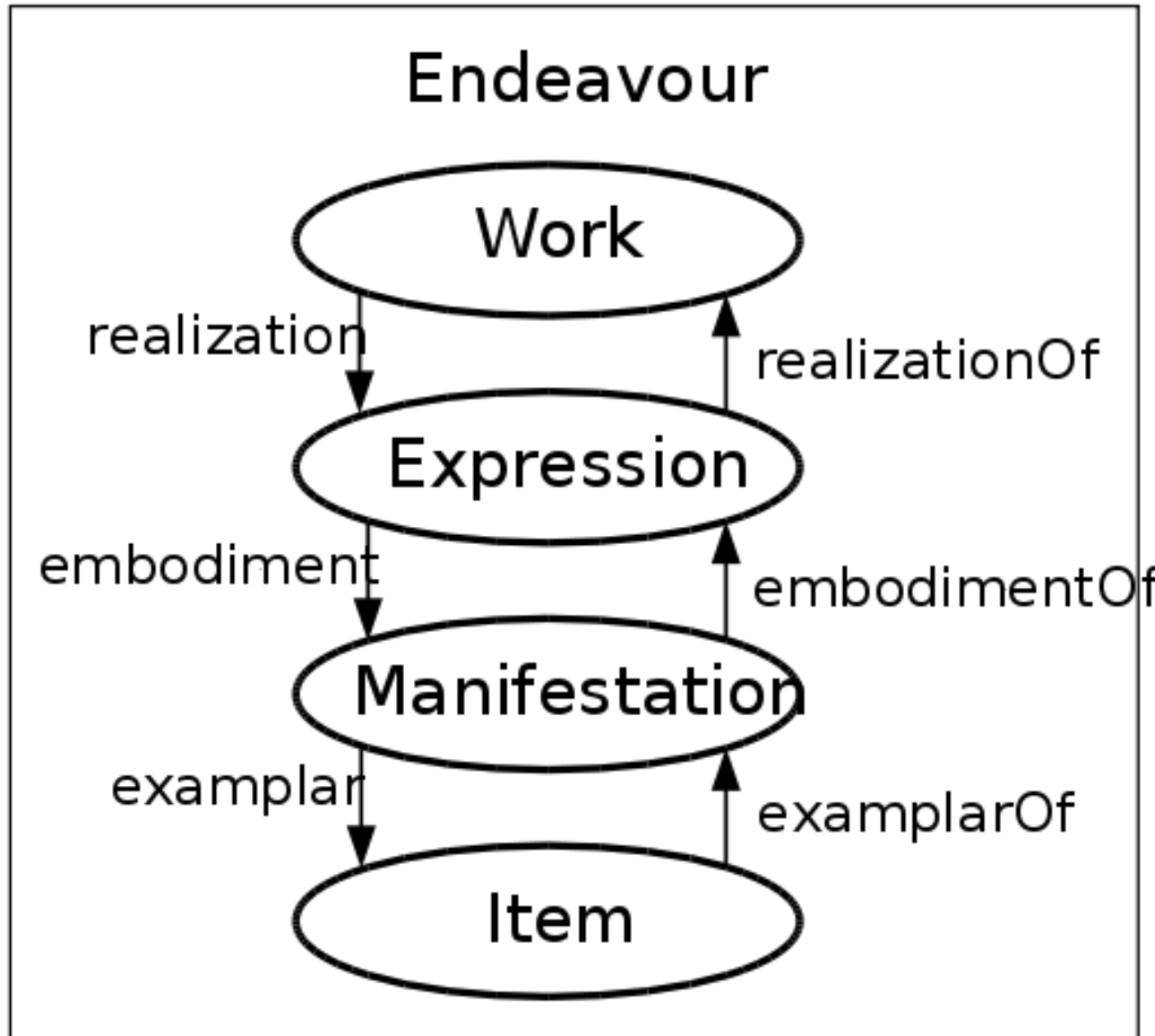
The Organization of Information

Third Edition

Paperback (ISBN 978-1-59…)

Copy 2 (barcode 102343…)

Family of Works

Equivalent

Microform Reproduction

Copy

Exact Reproduction

Facsimile

Reprint

Derivative

Edition

Simultaneous "Publication"

Abridged Edition

Revision

Illustrated Edition

Translation

Expurgated Edition

Variations or Versions

Arrangement

Slight Modification

Free Translation

Summary Abstract Digest

Dramatization Novelization Screenplay

Libretto

Change of Genre

Parody Imitation

Same Style or Thematic Content

Adaptation

Descriptive

Review

Casebook

Criticism

Evaluation

Annotated Edition

Commentary

Original Work - Same Expression

Same Work – New Expression

Cataloging Rules Cut-Off Point New Work

Metadata

	Material Culture Libraries Archives Museums	Bibliographic Libraries Archives Museums	Archival Libraries Archives Museums
Data Structure	CDWA	MARC	EAD
Data Content	CCO	AACR2 (RDA)	DACS
Data Format	XML	XML/ISO2709	XML
Data Exchange	OAI	OAI Z39.50 SRU/SRW	OAI

Adapted from Elings and Waibel, *First Monday*, (12)3, 2007

Some Types of "Metadata"

- Descriptive
  - Content, creation process, relationships
- Technical
  - Format, system requirements
- Administrative
  - Acquisition, authentication, access rights
- Preservation
  - Media migration
- Usage
  - Display, derivative works

Not in Taylor & Joudrey

Adapted from
Introduction to Metadata,
Getty Information Institute (2000)

Aspects of Metadata

- Framework
    - Functional Requirements for Bibliographic Records (FRBR)
- Schema ("Data Fields and Structure")
    - Dublin Core
- Guidelines ("Data Content and Values")
    - Resource Description and Access (RDA)
    - Library of Congress Subject Headings (LCSH)
- Representation (abstract "Data Format")
    - Resource Description Framework (RDF)
- Serialization ("Data Format")
    - RDF in eXtensible Markup Language (RDF/XML)

Some RDA Elements for Products

- Work
  - ID
  - Title
  - Date
  - etc.

- Expression
  - ID
  - Form
  - Date
  - Language
  - etc.

- Manifestation
  - ID
  - Title
  - Statement of responsibility
  - Edition
  - Imprint (place, publisher, date)
  - Form/extent of carrier
  - Terms of availability
  - Mode of access
  - etc.

- Item
  - ID
  - Provenance
  - Location
  - etc.

Library of Congress Classification

Book title: Uncensored War: The Media and Vietnam
Author: Daniel C. Hallin
Call Number: DS559.46 .H35 1986

First two lines { DS 559.45
Third line { .H35
Last line { 1986
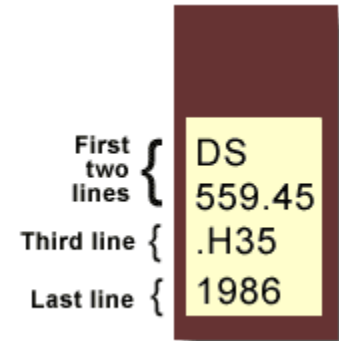
The first two lines describe the subject of the book.
DS559.45 = Vietnamese Conflict

D	History
DS1-937	History of Asia
DS520-560.72	Southeast Asia
DS556-559.93	Vietnam. Annam
DS557-559.9	Vietnamese Conflict

The third line often represents the author's last name.

H = Hallin

*After other initial **consonants***
for the second letter:

	a	e	i	o	r	u	y
use number:	3	4	5	6	7	8	9

*For **expansion***

for the letter:	a-d	e-h	i-l	m-o	p-s	t-v	w-z
use number:	3	4	5	6	7	8	9

The last line represents the date of publication.

The World Is Flat (in LCC)

HM846 .F74 2005

H	Social sciences
HM	Sociology
HM831	Social change – Causes
HM846	Technological Innovations. Technology.
.F74	Cutter number for Friedman, Thomas

The World Is Flat (in Dewey)

303.4833

300   Social science

300   Social sciences, sociology, & anthropology

30**3**   Social processes

303.4 Social change

303.48 Causes of change

303.483 Development of science and technology

303.4833 Communication (Information technology)

LCSH Subdivisions

- ## Topical

  Archaeology – Methodology

- ## Form

  Archaeology – Fiction

 - ## Chronological

  Archaeology – History – 18th century

- ## Geographic

  Archaeology – Egypt

Inter-Library Loan

- Users search "union catalog" to find books

- Remote library "ships" it to local library
  - Often by scanning it, where practical
  - Someone pays for this (local library or user)

- Local library manages circulation
  - Limited access period
  - Some "return" mechanism

Weeding ("Library Hygiene")

- Presumes some limited asset
  - e.g., shelf space, browsing time, …
- Anticipated future use
  - Reshelving and circulation statistics
  - Historical value
  - Sufficiency of single copies
  - Last copy doctrine
- Condition
  - Preservation costs

HathiTrust

- Centralized repository for digitized books
  - Google Books digitization (via owning libraries)
  - Microsoft book search (ran from 2006-2008)
  - Internet Archive
    - Million book project, project Gutenberg, contributions, …
  - Cooperative digitization

20%

80%
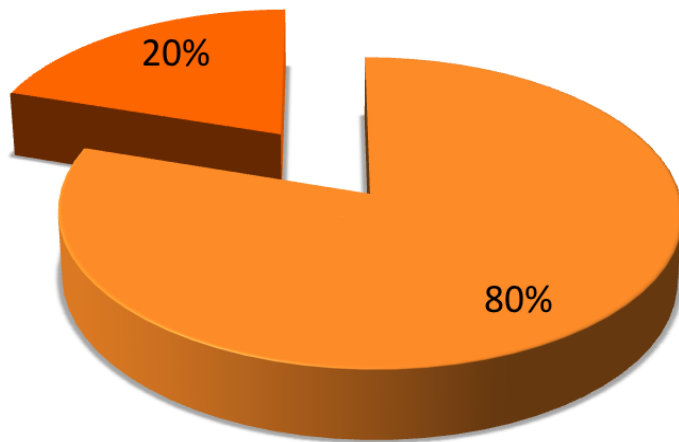
■ In Copyright
■ Public Domain

As of August 13, 2010
6,549,680 Total volumes
3,798,116 Book titles
   153,311 Serial titles
1,300,896 Public Domain

Archives

- Government
  - Legal, cultural

- Institutional
  - Liability, institutional memory

- Manuscript repositories
  - Research, preservation

Some Sources for Collections

- Institutional components
  - Transferred from records management

- Donors
  - Typically deed of gift specifies terms

- Purchase

National Archives Records Schedules

Schedule 1. Civilian Personnel Records
Schedule 2. Payrolling and Pay Administration Records
Schedule 3. Procurement, Supply, and Grant Records
Schedule 4. Property Disposal Records
Schedule 5. Budget Preparation, Presentation, and Apportionment Records
Schedule 6. Accountable Officers' Accounts Records
Schedule 7. Expenditure Accounting Records
Schedule 8. Stores, Plant, and Cost Accounting Records
Schedule 9. Travel and Transportation Records
Schedule 10. Motor Vehicle Maintenance and Operations Records
Schedule 11. Space and Maintenance Records
Schedule 12. Communications Records
Schedule 13. Printing, Binding, Duplication, and Distribution Records
Schedule 14. Information Services Records
Schedule 15. Housing Records
Schedule 16. Administrative Management Records
Schedule 17. Cartographic, Aerial Photographic, Architectural, and Engineering Records
Schedule 18. Security and Protective Services Records
Schedule 20. Electronic Records
Schedule 21. Audiovisual Records
Schedule 23. Records Common to Most Offices Within Agencies
Schedule 24. Information Technology Operations and Management Records
Schedule 25. Ethics Program Records
Schedule 26. Temporary Commissions, Boards, Councils and Committees
Schedule 27. Records of the Chief Information Officer

Appraisal

- Value
  - Evidential
  - Informational
- Costs
  - Storage, arrangement, description, preservation, …
- Stakeholder interests
  - Primary: Institutional needs
  - Primary: Accountability
  - Secondary: Other future record users

DACS Principles

1. Records in archives possess unique characteristics.
2. The principle of respect des finds is the basis of archival arrangement and description.
3. Arrangement involves identification of groupings within material.
4. Description reflects arrangement.
5. The rules of description apply to all archival materials regardless of form or medium.
6. The principles of archival description apply equally to records created by corporate bodies, individuals, or families.
7. Archival descriptions may be presented at varying levels of detail to produce a variety of outputs.
8. The creators of archival materials, as well as the materials themselves, must be described.

(Single-Level) DACS Elements

Required

- Reference code
- Name+location of repository
- Title
- Date
- Extent
- Name of creator(s)
- Scope and content
- Conditions governing access
- Languages and scripts
- Plus, for "Optimal"
  – Administrative/biographical history
  – Access points

Optional

- System of arrangement
- Physical access
- Technical access
- Conditions for reproduction and use
- (other) Finding aids
- Custodial history
- Immediate source of acquisition
- Appraisal, destruction, scheduling
- Accruals (anticipated additions)
- Existence+location of originals
- Existence+location of copies
- Related archival materials
- Publication note
- Notes
- Description control

Deaccessioning

- Space limits

- Policy changes

- Technology changes

Threats to Physical Collections

- Organic decay
- Intentional actions
  - Pilferage and vandalism
  - Official acts
- Disasters
  - Natural disasters
    - Flood, tornado, earthquake, …
  - Accidents
    - Fire, sprinkler malfunction, …
  - Armed conflict

Organic Decay

- Rag paper: 300-2,000 years

- Acidic paper: 25-50 years

- Acetate film: 40 years

- Nitrate film: 40-1-00 years
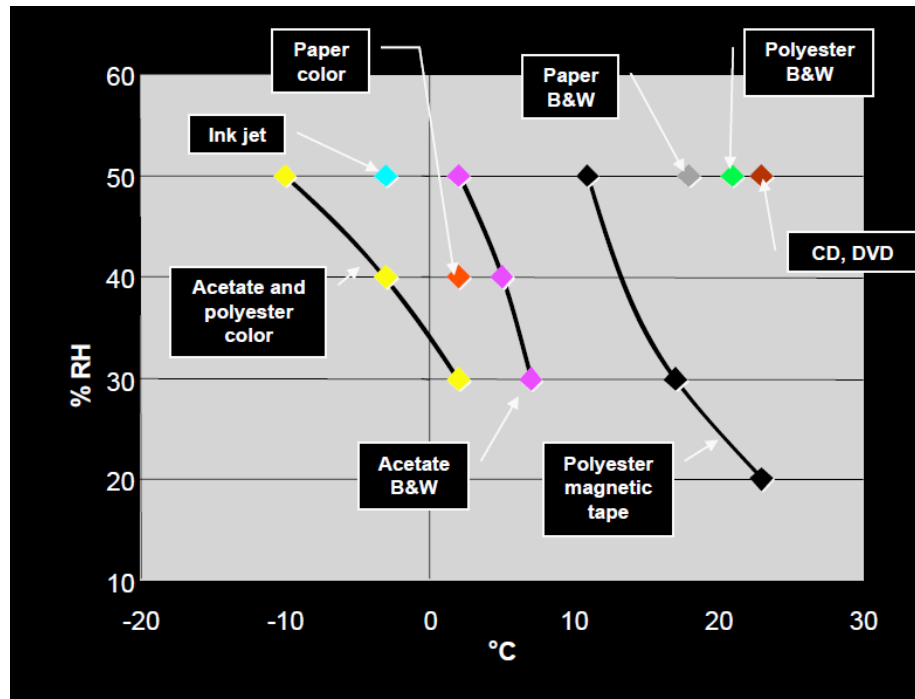
ISO 11799:2003

Digital Preservation

- Preservation of born-digital materials
  - Preserving appearance and interpretability
  - Preserving behavior

- Digitization for preservation
  - Scanning (of paper, of microfilm)
  - Audio digitization
  - Video digitization
  - Volumetric imaging
    - Digital holography, computational tomography

Indiana University Digitization

Table 6: Media Preservation Targets, 2013-2027

Target	Hours	Objects	% of Total Holdings
15 Years— all media types	317,000	408,000	71%
Audio	207,000	284,000	82%
Video	83,000	66,000	53%*
Film (access digitization)	27,000	58,000	69%

*IU Bloomington video holdings include a large number of non-archival, commercial VHS tapes and DVDs that circulate primarily to students. These are not included here.

■ Projected IU Storage Needs (PB)   ■ Cumulative MPI Data (PB)
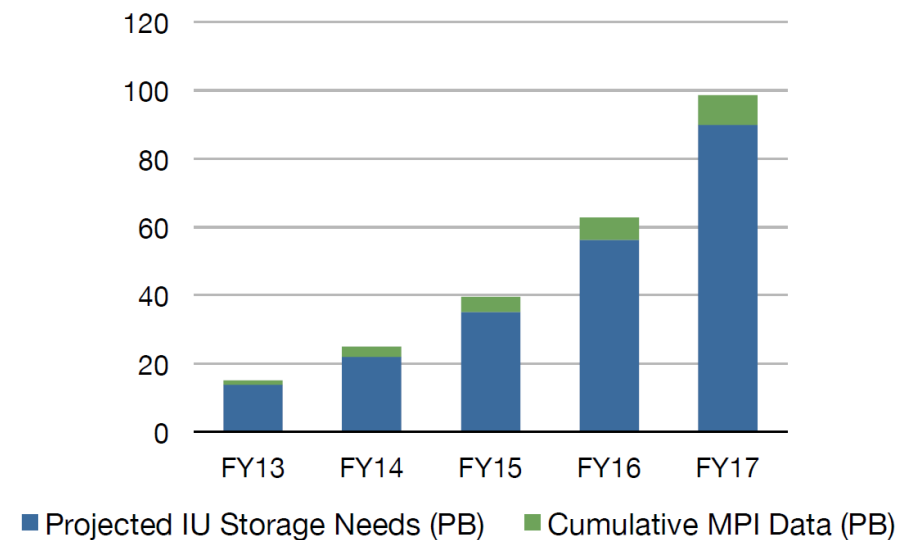
Preserving Behavior

- Word processors
  - Formatting, track changes, undo deleted text
- Spreadsheets
  - Formulas, visualizations
- Databases
  - Queries, forms, derived values
- Computer-Assisted Design (CAD)
  - Display, modification, manufacturing
- Software
  - Simulation, games, embedded systems, …

Behavior Preservation Strategies

- Format migration
  - For example, convert Word Perfect to PDF

- Emulation
  - Allows running old software on newer systems

ITHAKA

- JSTOR digitization
  - Back runs of journals
  - Recently expanded to books

- Portico preservation
  - Centralized management, originally for journals
    - Release triggers: discontinuation, loss of access
  - Also service for books and datasets