

Digital Library Curriculum Development

Module 3-b: Digitization

(Draft, Last Updated, 09/03/08)

1. Module name: Digitization

2. Scope

This module covers the general principles and application of the digitization process to build a collection for digital libraries.

3. Learning objectives

By the end of this lesson, the student will be able to:

- a. Explain the standard process of digitization projects, from initiating the project, to selecting and creating materials, making them accessible to users, and maintaining the collection of digitized materials.
- b. Demonstrate the critical issues and challenges of the digitization project (e.g., the potential uses, legal and financial considerations, preservation, and technical feasibility).
- c. Practice digitization, by creating a small-scale collection of digital objects.

4. 5S Characteristics of the module

- **Stream:** Digitization creates a stream of data entering the digital library.
- **Structure:** The concept of structures may apply to deal with the technical standards related to the digitization process, and to manage the digitized resources.
- **Spaces:** The physical storage issues, such as where the digital resources will be stored, and where the network server will be located, can be discussed relative to spaces. Also, images and video work with 2-D spaces, digitization of sculpture and buildings relates to 3-D spaces, and digitization of archaeological sites relates to 3-D as well as 4-D (considering spatial-temporal connections).
- **Scenario:** Digitization is a process, with workflows. There are a variety of scenarios related to capturing, transforming, and managing the resulting data.
- **Society:** N/A

5. Level of effort required

- a. Class time: 1 1/2 hour
- b. Student time outside class: 4 hours

- Reading before the class starts: 2 hours
- Homework assignment: 2 hours

6. Relationships with other modules

- 2-a: Text Resources, 2-b: Multimedia

2-a and 2-b can be taught prior to 3-b. The nature, structure, composing factors, and formats of various types of digital objects (e.g., text, images, video, etc.) are reviewed in 2-a and 2-b, while 3-b covers the general process of digitization regardless of the object types, and related issues to consider.

- 4-b: Metadata, cataloging, metadata mark-up, metadata harvesting

4-b covers uses of metadata and metadata standards related to the context of digital libraries in general. In 3-b, the scope is narrowed and covers the issues about assigning metadata to digitized materials.

- 8-a: Preservation

One of purposes of digitization is to make materials accessible and usable for a long-term period of time. 8-a covers the related technology, standards, and policies concerning the preservation of digital objects.

- 9-a: Project Management

While 3-b explains the administrative decision-making processes, mainly focusing on activities related to digitization, 9-a deals with the issues of the overall process of building and maintaining a digital library.

- 9-e: Intellectual Property, 9-f: Cost/Economic Issues

The comprehensive review of legal and economic issues regarding the overall aspects of digital libraries is introduced in 9-e and 9-f.

7. Prerequisite knowledge required: None

8. Introductory remedial instruction: None


9. Body of knowledge

1. Definition

- Born digital vs. being digitized
- Digitization: “The conversion of an analogue signal or code into a digital signal or code” (Chowdhury & Chowdhury, 2003; Lee, 2001)
 - Analogue examples

- Clocks, or speed indicators, with the hands showing the continuous change of moments
- Natural vision, voice, or hearing
- o Digital examples
 - Digital Images: “Electronic snapshots taken of a scene or scanned from documents, such as photographs, manuscripts, printed texts, and artwork” (Cornell University Library, 2000: A more detailed explanation about digital images is available at <http://www.library.cornell.edu/preservation/tutorial/intro/intro-01.html>)
 - Digital clocks, digital speed indicators, and digital photos, videos and sounds (More examples and a detailed explanation about digital objects will be introduced in the 2-a: Text Resources, and 2-b: Multimedia modules.)
- o Digital Conversion/Representation of Analog
 - An electronic process that transforms the continuous tones, waves, lines or images into segments, dots or bit streams, with assigned values, without changing the original contents
 - Benefits:
 - a. Easy to duplicate
 - b. Easy to edit, or reformat (Flexibility)
 - c. Easy to store and maintain
 - Drawbacks & Risks:
 - a. Possible change of the value of the original object. (e.g., the physical form of the historic documents)
 - b. Authenticity, version control
 - c. Migration
 - d. Reader or viewer software dependency

2. Digitization Process

 Figure: Steps involved in digitization (Chowdhury & Chowdhury, 2003, p. 106, Fig. 6.1)


A. Potential and Intended uses

- o Expecting frequency of use
- o User needs to access digital resources
- o Security or access to use issues
- o Control unauthorized access and use
- o Shared collection, collaboration, and consortium

B. Considering issues before digitization

- Intellectual nature of the source materials
 - Enhancing the intellectual value of the resources
- Legal restrictions
 - Copyright protection of the resources
 - Resource collection from the public domain/electronic databases
 - 'Fair use'
 - Using for educational purpose
 - Collecting resources which are no longer under copyright
 - Orphan works
- Finance
 - Available funds
 - Staff resources (skills, experiences, training costs)
 - Time cost
 - Cost for digitizing, maintaining and updating materials
- Preservation considerations
 - Possible damage to the original resources from digitization
 - Protection when handling
- Technical feasibility
 - Technical infrastructure for institutes
 - Hardware and software
 - Usable equipment, facilities, and tools
 - Standards (file formats, metadata schema, indexes, storages, etc.)

C. Selecting materials for digitization

 Figure: Selection for Digitizing: A Decision-Making Matrix (Hazen, Horrell, & Merrill-Oldham, 1998, available at: <http://www.clir.org/PUBS/reports/hazen/matrix.html>)

- Types of materials (texts, images, photos, videos, etc.)
- Vulnerability of the source materials
- Physical attributes of materials (sizes, conditions, colors, etc.)

D. Actions for digitizing

- Scanning
 - Resolution, color, file formats, display requirements

- File format standards:

- Table: Common Image File Formats (Connell University (2000), Available at: <http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html>)

- Table: Digital Master Images-Sample Technical Specifications for Photograph Collections (Peterson (2004), Available at: http://www.loc.gov/rr/print/tp/DgtlMastersSamplSpecsSelctdRcmndFinal7_2004.pdf)

- Quality control
 - Conversion / Compression

E. Processing for use

- Metadata assignment
 - Indexing (metadata vs. full-text)
 - Searching and browsing

3. Digitization Projects

- Google Books Library Project
 - Partnership with about 18 libraries including Harvard University, Oxford University, Stanford University, and the University of Michigan (MBooks - Michigan Digitization Project, <http://www.lib.umich.edu/mdp/>)
 - Digitizing the full text of out-of-copyright books of libraries and making them available with no charge through Google Book Search (<http://books.google.com/>)
 - Library Partners: <http://books.google.com/googlebooks/partners.html>
 - University of Michigan Library/Google Digitization Partnership FAQ: <http://www.lib.umich.edu/mdp/overview.pdf>
- Open Content Alliance (OCA) (<http://www.opencontentalliance.org/>)
 - An international consortium among cultural, technology, nonprofit organizations to build a permanent archive of a digital collection of text and multimedia content.
 - Announced in October 2005 by the Internet Archive
 - Scanning books and uploading them to the Open Library
 - Copyrighted books: getting permissions from copyright holders
 - Operating the Open Library (<http://www.openlibrary.org/>)
 - About 200,000 scanned books are currently available to the public for free.

- Comparing to the Internet Archive (<http://www.archive.org/>): offering text, audio, moving images, web content and software for public use
- o Contributors & Partners: university libraries in U.S., Canada, the European Archive, the National Archive of U.K., HP Labs, MSN, O'Reilly Media, Yahoo!, etc.
- o Video: Libraries Going Open!
(http://www.archive.org/details/oca_2007_movie)
- The Library of Congress: American Memory
(http://memory.loc.gov/ammem/collections/habs_haer/index.html)
 - o 1990-1994: Nation's Memory
 - Digitizing some of the Library of Congress's unparalleled collections of historical documents, moving images, sound recordings, and print and photographic media
 - o 1994: American Memory historical collections
 - Received \$13 million in private sector donations to establish the National Digital Library Program
 - Partnership with \$45 million in private sponsors from 1994 through 2000.
 - o Storing more than 9 million documents about U.S. history and culture
 - o Organizing documents with about 100 thematic categories based on their original format, their subject matter, or who first created, assembled, or donated them to the library.
 - o Including manuscripts, prints, photographs, posters, maps, sound recordings, motion pictures, books, pamphlets, and sheet music
 - o Related resources:
 - Library of Congress Technical Standards for Digital Conversion of Text and Graphic Materials
(<http://memory.loc.gov/ammem/about/techStandards.pdf>)
 - Technical Q&A about copyright, metadata, preservation, scanning, conversion, text-markup, etc.
(<http://memory.loc.gov/ammem/about/techIn.html>)

10. Resources

a. Required readings for students

Chowdhury, G.G., & Chowdhury, S. (2003). Chapter 6, Digitization (pp. 103-119). In *Introduction to Digital Libraries*. London: Facet Publishing

Cornell University Library. (2000). Moving theory into practice: Digital imaging tutorial. Retrieved September 3, 2008, from <http://www.library.cornell.edu/preservation/tutorial/contents.html>

Smith, Abby. (1999). Why Digitize? Washington, DC: Council on Library & Information Resources. Retrieved November 2, 2007, from <http://www.clir.org/pubs/abstract/pub80.html>

b. Recommended readings for students

Liu, Y.Q. (2004). Best practices, standards, and techniques for digitizing library materials: A snapshot of library digitization practice in the US. *Online Information Review*, 28(5), 338-345.

Humanities Advanced Technology and Information Institute (2002). III. Selecting Materials: An Iterative Process, In the NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials. Retrieved January 31, 2008, from <http://www.nyu.edu/its/humanities/ninchguide/III/>

Hazen, D., Horrel J., & Merrill-Oldham, J. (1998). Selecting Research Collections for Digitization. Washington, DC: Council on Library & Information Resources. Retrieved November 2, 2007, from <http://www.clir.org/PUBS/reports/hazen/pub74.html>

c. Suggested readings for instructors

o *Introduction to Digitization/Digitization Handbooks*

Baxes, G. (1994). *Digital Image Processing: Principles and Application*. New York, NY: Wiley.

Besser, H. (2003). *Introduction to Imaging* (rev. ed.). Los Angeles, CA: Getty Research Institute. Retrieved November 2, 2007, from http://www.getty.edu/research/conducting_research/standards/introimages/index.html

Lee, S. (2001). *Digital Imaging: A Practical Handbook*. New York: Neal-Schuman Publishers, Inc.

Lesk, M. (2004) Chapter 3, Images of pages. In *Understanding Digital Libraries*. (2nd ed) (pp. 61-90). San Francisco, CA: Morgan Kaufmann.

Peterson (2004) Digital Master Images-Sample Technical Specifications for Photograph Collections, available at: http://www.loc.gov/rr/print/tp/DgtlMastersSamplSpecsSelctdRcmndFinal7_2004.pdf

Puglia, S. (2000). VI, Technical primer. Andover, MA: Northeast Document Conservation Center (NEDCC). Retrieved November 2, 2007, from <http://www.nedcc.org/oldnedccsite/digital/vi.htm>.

Vogt-O'Connor, D. (2000). IV, Selection of materials for scanning. Andover, MA: Northeast Document Conservation Center (NEDCC). Retrieved November 2, 2007, from <http://www.nedcc.org/oldnedccsite/digital/iv.htm>.

o *Standards/Rationale*

Conway, P. (2000). II, Overview: Rationale for digitization and preservation. Andover, MA: Northeast Document Conservation Center (NEDCC). Retrieved November 2, 2007, from <http://www.nedcc.org/oldnedccsite/digital/ii.htm>.

Peterson, K. A. (2004). Digital Master Images-Sample Technical Specifications for Photograph Collections, Retrieved January 31, 2008, from http://www.loc.gov/rr/print/tp/DgtlMastersSamplSpecsSelctdRcmndFinal7_2004.pdf

Wisser, K. M. (2007). North Carolina ECHO guideline for digitization. Retrieved January 31, 2007, from <http://www.ncecho.org/guide/toc.html>

Technical Advisory Service for Images (2006). File naming. Retrieved January 31, 2008 from <http://www.tasi.ac.uk/advice/creating/filenaming.html>

○ *Practices/Projects*

Brancolini, K.R. (2000). Selecting research collections for digitization: Applying the Harvard Model. *Library Trends*, 48(4), 783-798

Macklin, L.L., & Lockmiller, S .L. (1999). Digital Imaging of Photographs, A Practical Approach to Workflow Design and Project Management. LITA Guides #4. American Library Association, Chicago.

University of Michigan, Digital Library Services (2001). Assessing the Costs of the Conversion: Making of America, The American Voice, 1850-1876. Retrieved November 2, 2007 from http://www.umdl.umich.edu/pubs/moa4_costs.pdf

○ *Digitization for Special Resources*

Brown, M.S. & Seales, B. (2000). Beyond 2D images: Effective 3D imaging for library materials. *Proceedings of the Fifth ACM Conference on Digital Libraries*, 27-36.

Gertz, J. (2000). Digitization of maps and other oversize documents. In Skitts, M. (Ed.), *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, MA: Northeast Document Conservation Center (NEDCC). Retrieved November 2, 2007, from <http://www.nedcc.org/oldnedccsite/digital/intro.htm>.

11. Concept map (created by students)

12. Exercises / Learning activities

- **Homework assignment: Building a digital image collection**

This assignment provides an opportunity for the students to create digital objects and process the objects to be used as a part of an art image collection of a hypothetical digital library that the class members will build together.

This is a class project to build a small scale digital library of art images publicly available in the local area. Students are asked to create digital images, taking pictures of art sculptures in the area and building a photo collection. (If the university offers digitization equipment such as a scanner, the students would scan images of art sculptures to create the digital files, instead of taking pictures.) The guidelines for this assignment are as follows.

- 1) Take a picture of any 3 art sculptures in the local area. It can be a school statue, local art work, historic clock, or any other types of sculptures, available to the public. Any type of digital cameras can be used for this project. You can use yours or borrow one from the university lab or library.
- 2) With the picture of the images, create an archival master file and two derivative images - one for full screen viewing and one thumbnail.
- 3) After creating the three files of each art work, assign core metadata for each master file.
- 4) Upload the images and related metadata to the web space provided by the instructor.
- 5) Write a short report describing the digitization and documentation procedure.
- 6) View the images and related metadata of others' additions to the collections and compare them to yours considering the following issues.
 - i. The image creation and metadata description of an art sculpture submitted by multiple students
 - ii. Copyright, intellectual property right issues: Who has the intellectual property rights to the images?

The class will have a discussion session on the assignment at the beginning of the next class.

Instructors need to provide specific guidelines to the students about how to create digital files, the file formats, sizes or resolution for the collection, core metadata elements, etc. Instructors can use the discussion section of Blackboard or similar courseware, or create a simple version of a digital library with applications, like Greenstone. It is important that the database of the collection is available to the students to view their own works as well as those of others.

13. Evaluation of learning achievement

14. Glossary

- Analog: Describes a device or system that represents changing values as continuously variable physical quantities (Webopedia: <http://www.webopedia.com/TERM/a/analog.html>)
- Digital: Describes any system based on discontinuous data or events. (Webopedia: <http://www.webopedia.com/TERM/d/digital.html>)
- Metadata: Data about data. A schema for describing data objects, or the data that describes a specific data object (see Module 4-a: Metadata, for detailed explanation.)
- Thumbnail: A reduced-size digital file of an image or picture for easy browsing and recognizing of the brief impression or content of the original file

15. Additional useful links

16. Contributors

- a. Initial author: Sanghee Oh
- b. Jeff Pomerantz, Barbara Wildemuth, Carol Perryman, Eunyoung Yoo, Stacy Kowalczyk, Edward A. Fox