

1 **Digital Library Curriculum Development**

2 **Module 3-b: Digitization**

3 (Draft, Last Updated, 11/16/07)

4  
5 **1. Module name: Digitization**

6  
7 **2. Scope**

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

10  
11 **3. Learning objectives**

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

- 13 a. Explain technical standards, selection criteria for digitization, and the digitization  
14 process.
- 15 b. Discuss the critical issues and challenges of digitization (e.g., their potential uses,  
16 legal and financial considerations, preservation, and technical  
17 feasibility/standards)
- 18 c. Develop and manage small scale digitization projects

19  
20 **4. 5S Characteristics of the module**

- 21 • Stream: Digitization creates a stream of data entering the digital library.
- 22 • Structure: The concept of structures may apply to deal with the technical  
23 standards related to the digitization process and manage the digitized resources.
- 24 • Spaces: The physical storage issues, such as where the digital resources will be  
25 stored, where the network server will be located, can be discussed related to  
26 spaces.
- 27 • Scenario: N/A
- 28 • Society: N/A

29  
30 **5. Level of effort required**

- 31 a. Class time: 1 1/2 hour
- 32 b. Student time outside class: 4 hours
- 33 • Reading before the class starts: 2 hours
- 34 • Homework assignment: 2 hours

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

## **6. Relationships with other modules**

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

The Text Resources module and the Multimedia module can be taught before the Digitization module is introduced. The nature, structure and composing factors of different types of digital objects are reviewed in these modules, while the Digitization module covers the technical formats and standards of various types of resources (e.g., text, images, video, etc), specifically related to the digitization process.

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

The basic principles and elements of metadata of digital materials are covered in this module. The Digitization module discusses about assigning the metadata to the digitized resources and the core elements to describe the resources.

- 8-a: Preservation

One of the most benefits of digitization is that it enables to access resources in long-term period of time. The related technology, standards, policies to preservation in digital libraries is reviewed in the Preservation module.

- 9-a: Project Management

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

- 9-e: Legal Issues, 9-f: Cost/Economic Issues

The comprehensive review of legal and economic issues regarding the overall aspects of digital libraries is introduced in the Legal Issues and Cost/Economic issues module.

## **7. Prerequisite knowledge required: None**

## **8. Introductory remedial instruction: None**

## **9. Body of knowledge**

### **1. What is digitization?**

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


- 1           - Clocks, or speed indicators with the hands showing the continuous change
- 2           of moments
- 3           - Natural vision, voice, or hearing
- 4           ○ Digital examples
- 5           - Digital Images: “Electronic snapshots taken of a scene or scanned from
- 6           documents, such as photographs, manuscripts, printed texts, and artwork”
- 7           (Cornell University Library, 2000)
- 8           - Computers which break data up into 0s and 1s and put together in a binary
- 9           code
- 10          - Digital clocks or speed indicators which represents times or speeds with
- 11          discrete numbers
- 12          - Digital photos, videos or sounds.
- 13          ○ Digital Conversion/Representation of Analog
- 14          - The continuous tones, waves, lines or images are divided into segments,
- 15          dots or bit streams, with assigned values and mapped, simulating the
- 16          original analog objects.
- 17          - Benefits:
- 18              a. Easy to duplicate
- 19              b. Easy to edit, or reformat (Flexibility)
- 20              c. Easy to store and maintain (Permanence)
- 21          - Drawbacks:
- 22              a. Not exactly the same as the original analog object
- 23                  For some purposes, the value of the original object is its physical
- 24                  form, e.g., the study of historic documents
- 25              b. Version control
- 26              c. Authenticity
- 27              d. Reader or viewer dependent
- 28              e. Migration

## 30   2. Benefits of Digitization for Users

- 31          • Enabling the remote access to resources
- 32          • Enabling simultaneous access of multiple people to resources
- 33          • Easy access to various versions of reference surrogates (e.g., thumbnails, low-
- 34          resolution images, etc)
- 35          • New scholarly use

- 1 • Powerful teaching materials
- 2 • Flexible modification, restoration, integration
- 3 • Preservation
- 4

### 5 3. Digitization Process

6  Figure 6.1: Steps involved in digitization (Chowdhury & Chowdhury, 2003, p.  
7 106)

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


#### 10

#### 11 A. Potential and Intended uses

- 12 ○ Expecting frequency of use
- 13 ○ User needs to access digital resources
- 14 ○ Security or access to use issues
- 15 ○ Control unauthorized access and use
- 16 ○ Shared collection, collaboration, and consortium

#### 17 B. Considering issues before digitization

- 18 ○ Intellectual nature of the source materials
  - 19 - Enhancing the intellectual value of the resources
- 20 ○ Legal restrictions
  - 21 - Copyright protection of the resources
  - 22 - Resource collection from the public domain/electronic databases
  - 23 - 'Fair use,'
  - 24 - Using for educational purpose
  - 25 - Collecting resources which are no longer under copyright
  - 26 - Orphan works
- 27 ○ Finance
  - 28 - Available funds
  - 29 - Staff resources (skills, experiences, training costs)
  - 30 - Time Cost
  - 31 - Cost for digitizing, maintaining and updating materials
- 32 ○ Preservation consideration
  - 33 - Possible damage to the original resources from digitization

- 1           - Protection for handling
- 2           ○ Technical feasibility
- 3           - Technical infrastructure of institutes
- 4           - Hardware and software
- 5           - Usable equipment, facilities, and tools
- 6           - Standards (file formats, metadata schema, indexing, storages, etc)
- 7        C. Selecting materials for digitization
- 8           ○ Types of materials (texts, images, photos, videos, etc.)
- 9           ○ Vulnerability of the source materials
- 10          ○ Physical attributes of materials (sizes, conditions, colors, etc.)
- 11        D. Actions for digitizing
- 12          ○ Scanning
- 13           - Resolution, color, file formats, display requirements
- 14           - File format standards:
- 15            Table: Common Image File Formats (Connell University (2000),
- 16           Available at:
- 17           [http://www.library.cornell.edu/preservation/tutorial/presentation/table7-](http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html)
- 18           [1.html](http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html))
- 19          ○ Quality control
- 20          ○ Conversion / Compression
- 21        E. Processing for use
- 22          ○ Metadata
- 23          ○ Indexing (metadata vs. full-text)
- 24          ○ Searching and browsing
- 25
- 26        5. Digitization Projects
- 27          ▪ Google Books Library Project
- 28           ○ Partnership about 18 libraries including Harvard University, Oxford
- 29           University, Stanford University and University of Michigan (MBooks -
- 30           Michigan Digitization Project (<http://www.lib.umich.edu/mdp/>)
- 31           ○ Digitizing the full text of out-of-copyright books of libraries and making
- 32           them available with no charge through Google Book Search
- 33           (<http://books.google.com/>)
- 34           ○ Library Partners: <http://books.google.com/googlebooks/partners.html>

- 1           ○ University of Michigan Library/Google Digitization Partnership FAQ:  
2           http://www.lib.umich.edu/mdp/overview.pdf
- 3        ▪ Open Content Alliance (OCA) (<http://www.opencontentalliance.org/>)
- 4           ○ An international consortium among cultural, technology, nonprofit  
5           organizations to build a permanent archive of digital collection of text and  
6           multimedia content.
- 7           ○ Announced in October 2005 by the Internet Archive
- 8           ○ Scanning books and uploading them to the Open Library
- 9           - Copyrighted books: getting permissions from copyright holders
- 10          ○ Operating the Open Library (<http://www.openlibrary.org/>)
- 11          - About 200,000 scanned books are currently available to the public for  
12          free.
- 13          - Comparing to the Internet Archive (<http://www.archive.org/>): offering  
14          text, audio, moving images, web content and software for public use
- 15          ○ Contributors & Partners: university libraries in U.S., Canada, the  
16          European Archive, the National Archive of U.K., HP Labs, MSN,  
17          O'Reilly Media, Yahoo!, etc.
- 18          ○ Video: Libraries Going Open!  
19          ([http://www.archive.org/details/oca\\_2007\\_movie](http://www.archive.org/details/oca_2007_movie))
- 20        ▪ The Library of Congress: American Memory  
21        ([http://memory.loc.gov/ammem/collections/habs\\_haer/index.html](http://memory.loc.gov/ammem/collections/habs_haer/index.html))
- 22          ○ 1990-1994: Nation's Memory
- 23          - Digitizing some of the Library of Congress's unparalleled collections  
24          of historical documents, moving images, sound recordings, and print  
25          and photographic media
- 26          ○ 1994: American Memory historical collections
- 27          - Received \$13 million in private sector donations to establish the  
28          National Digital Library Program
- 29          - Partnership with \$45 million in private sponsors from 1994 through  
30          2000.
- 31          ○ Available more than 9 million items that document U.S. history and  
32          culture
- 33          ○ About 100 thematic collections available based on their original format,  
34          their subject matter, or who first created, assembled, or donated them to  
35          the Library.
- 36          ○ Including manuscripts, prints, photographs, posters, maps, sound  
37          recordings, motion pictures, books, pamphlets, and sheet music

- 1           ○ Library of Congress Technical Standards for Digital Conversion of Text  
2           and Graphic Materials  
3           (<http://memory.loc.gov/ammem/about/techStandards.pdf>)
- 4           ○ Technical Q&A about copyright, metadata, preservation, scanning,  
5           conversion, text-markup, etc.  
6           (<http://memory.loc.gov/ammem/about/techIn.html>)

7

## 8 **10. Resources**

### 9 **a. Required readings for students**

- 10           Chowdhury, G.G., & Chowdhury, S. (2003). Chapter 6, Digitization. (pp. 103-  
11           119) In *Introduction to Digital Libraries*. London: Facet Publishing
- 12           Cornell University Library. (2000). Moving theory into practice: Digital imaging  
13           tutorial. Retrieved October 29, 2005, from  
14           <http://www.library.cornell.edu/preservation/tutorial/contents.html>
- 15           Smith, Abby. (1999). Why Digitize? Washington, DC: Council on Library &  
16           Information Resources. Retrieved November 2, 2007, from  
17           <http://www.clir.org/pubs/abstract/pub80.html>

### 18 **b. Recommended readings for students**

- 19           Liu, Y.Q. (2004). Best practices, standards, and techniques for digitizing library  
20           materials: A snapshot of library digitization practice in the US. *Online*  
21           *Information Review*, 28(5), 338-345.
- 22           Hazen, D., Horrel J., & Merrill-Oldham, J. (1998). Selecting Research Collections  
23           for Digitization. Washington, DC: Council on Library & Information Resources.  
24           Retrieved November 2, 2007, from  
25           <http://www.clir.org/PUBS/reports/hazen/pub74.html>

### 26 **c. Suggested readings for instructors**

- 27           ○ *Introduction to Digitization/Digitization Handbooks*
- 28           Baxes, G. (1994). *Digital Image Processing: Principles and Application*. New  
29           York, NY : Wiley.
- 30           Besser, H. (2003). *Introduction to Imaging* (rev. ed.). Los Angeles, CA: Getty  
31           Research Institute. Retrieved November 2, 2007, from  
32           [http://www.getty.edu/research/conducting\\_research/standards/introimages/index.h](http://www.getty.edu/research/conducting_research/standards/introimages/index.html)  
33           [tml](http://www.getty.edu/research/conducting_research/standards/introimages/index.html)
- 34           Lee, S. (2001). *Digital Imaging: A Practical Handbook*. New York: Neal-  
35           Schuman Publishers, Inc.
- 36           Lesk, M. (2004) Chapter 3, Images of pages. In *Understanding Digital Libraries*.  
37           (2nd ed) (pp. 61-90). San Francisco, CA: Morgan Kaufmann.

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

4 Vogt-O'Connor, D. (2000). IV, Selection of materials for scanning. Andover,  
 5 MA: Northeast Document Conservation Center (NEDCC). Retrieved November 2,  
 6 2007, from <http://www.nedcc.org/oldnedccsite/digital/iv.htm>.

7 ○ *Standards/Rationale*

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

11 ○ *Practices/Projects*

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

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

17 University of Michigan, Digital Library Services (2001) Assessing the Costs of  
 18 the Conversion: Making of America, The American Voice, 1850-1876. Retrieved  
 19 November 2, 2007 from [http://www.umdl.umich.edu/pubs/moa4\\_costs.pdf](http://www.umdl.umich.edu/pubs/moa4_costs.pdf)

20 ○ *Digitization for Special Resources*

21 Brown, M.S. & Seales, B. (2000). Beyond 2D images: Effective 3D imaging for  
 22 library materials. *Proceedings of the Fifth ACM Conference on Digital Libraries*,  
 23 27-36.

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

29

30 **11. Concept map (created by students)**

31

32 **12. Exercises / Learning activities**

33 • **Ungraded homework assignment: Building a digital image collection**

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

37 This is a project to build a digital library of art images, in particular the sculptures  
 38 publicly available in the local area. Students are asked to take a picture of at least  
 39 5 art sculptures in the area and process the digital photos.



- 1) Take a picture of any 5 art sculptures in the local area. It can be a school statue, local art works, historic clocks 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 two different types of digital objects - one for the image display in the library, and the other for the image preview, a thumbnail.
- 3) After creating the two images of each art work, assign core metadata for each image.
- 4) Upload the images and related metadata to the web space where the instructor provides.  

(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 necessary that the database of the collection is available to the students to view their own works as well as those of others.)
- 5) View the images and related metadata of others' additions to the collections and compare them to yours considering the following issues.
  - i. File formats or other technical standards of images and thumbnails.
  - ii. The common or uncommon elements of metadata used to describe the images
  - iii. The image creation and metadata description of an art sculptures submitted by multiple students
  - iv. Standards issues
  - v. Copyright, intellectual property right issues: Who has the intellectual property rights to the images?

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

### 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).

- 1           ○ Thumbnail: A reduced-size digital file of an image or picture for easy browsing
- 2                   and recognizing of the brief impression or content of the original file.

3

4   **15. Additional useful links**

5

6   **16. Contributors**

7           a. Initial author: Sanghee Oh

8           b. Team evaluators: Jeff Pomerantz, Barbara Wildemuth