UC Santa Barbara

Core Curriculum-Geographic Information Science (1997-2000)

Title
Unit 191 - Digital Libraries

Permalink
https://escholarship.org/uc/item/9f234240

Authors
191, CC in GIScience
Yeung, Albert K.

Publication Date
2000

Peer reviewed

Unit 191 - Digital Libraries

By Albert K. Yeung
Ontario Ministry of Northern Development and Mines, Canada

Advanced Organizer

Topics covered in this unit

This unit describes the use of World Wide Web technology for the development of digital libraries. Topics covered in this unit include the objectives, concept and architecture of digital libraries including spatial digital libraries.

Learning Outcomes

After learning the materials covered in this unit, students should be able to:

- explain the differences between a digital library and a conventional library
- explain the impacts of digital libraries on the future use of the World Wide Web

Full Table of Contents

Metadata and Revision History

Digital Libraries

1. Digital Libraries

- digital libraries are also referred to as *virtual libraries* and *electronic libraries*

1.1. What is a Digital Library?

- the advent of the computer has revolutionized the ways conventional libraries are organized and operated
    - early efforts in computerization focused mainly on the automation of end-user support functions such as cataloging and loan records
- networking and computing technologies have now become sufficiently advanced to support the design and deployment of large digital libraries which are not only capable of supporting the conventional end-user functions, but also networked access to printed and non-printed materials, including images as well as audio and video files
- the objective of digital libraries is not merely to automate traditional library functions
- it represents a new approach to the operation of libraries that encompasses
    - new types of information resources
    - new methods of acquiring and distributing information resources
    - new means of storing and preserving information resources
    - new modes of interaction with and for library users
    - new ways of using computing and network technologies
- the issue of digital libraries is a high priority government agenda because they represent a way to
    - reduce the cost of government
    - enforce standards for information collection and management
    - meet public access laws mandated by the federal/state legislatures
    - support interoperability between different computer systems
- the concept of digital libraries has greatly expanded the traditional role of libraries as a repository of printed scientific, literary and artistic materials
    - digital libraries are not only operated as a general public service for the advancement and dissemination of information
    - they represent an environment that brings together the collection, services and people in support of the full like cycle of creating, disseminating, using and preserving data, information and knowledge
    - they constitute an essential component of the *national information infrastructure* (NII) that aims to bring knowledge to every citizen
        - NII is a United States federal government initiative that aims to physically network the nation to create an information superhighway (see Unit 190)
        - it also aims to connect all citizens to the network
        - similar initiatives are also found in other countries (e.g. Australia, Canada and European Union countries)

1.2. The World Wide Web and Digital Libraries

- the Internet, and World Wide Web in particular, is a critical component of digital libraries because
    - as the Internet expands, more people begin to recognize the need to search indexed collections of information
        - the Internet is like a library without a card catalog

- without the necessary mechanisms it is hard to find information on the Web effectively and efficiently
- the desire to develop the necessary infrastructure to effectively mass-manipulate information on the Internet is one of the primary driving forces behind the digital library initiatives in the United States and many other countries
  - the capabilities of the World Wide Web protocol make it an ideal tool to develop the interface of digital libraries (see Unit 148)
  - the wealth of digital information of the World Wide Web makes it the most valuable source of information for digital libraries
  - the large number of people who now use the World Wide Web regularly makes it the most widely used mechanism to access digital information
- on the other hand, trying to search for information in the Web is a daunting experience for many users
  - despite the advances in search engine technology, current methods are not intuitive
    - many users are not used to elaborating an artificial text string to match their requirements
    - information returned by search engines is practically unorganized
    - there is no mechanism to assist users in interpreting information retrieved from the Web
    - there is no mechanism to assist users in managing information retrieved from the Web
  - the objective of digital libraries is to help users obtain knowledge by filtering and cross-referencing information they obtain from the Web
- the objective of digital libraries is to help users obtain knowledge by filtering and cross-referencing information they obtain from the Web

1.3. Components and Architecture of a Digital Library

- the following sections describe the components and architecture of an *ideal* digital library
  - not all the examples of digital libraries noted later in this unit are built on these concepts and technologies

1.3.1. The building blocks of digital libraries

- digital libraries are made up of three components: *digital objects*, *handles* and *repositories* (Figure 1)
- a *digital object* is the basic unit of the digital library architecture
  - a digital object is composed of two key elements
    - digital materials or data, including library materials, their structure and associated information such as intellectual property rights
    - key-metadata, i.e. information required to access the digital object in a networked environment, including conditions of use and a handle
- a *handle* is a general-purpose identifier that uniquely identifies a specific digital object in a repository
  - it is also called *Uniform Resource Name* (URN) because it identifies Internet

resources by name, in contrast to the Uniform Resource Locator (URL) that identifies Internet resources by location
- handles are created by naming authorities that are mandated to create and edit handles
- handles are managed by a distributed computer system called the *handle system*
    - the handle system stores handles and associated data used to identify and access data objects named by handles
- a *repository* is a system for network-based storage and access to digital objects
    - users interact with repositories using a simple protocol, known as *Repository Access Protocol* (RAP)
        - RAP has a small number of fundamental operations such as "deposit", "access", "verify" and "delete", that allow a user to access the contents of a digital object or its key-metadata

1.3.2. The architecture of digital libraries

- the architecture of a digital library is made up of four components: *user interface*, *repository*, *handle system* and *search system* (Figure 2)
- the user interface has two parts
    - a standard Web browser for the interaction between the user and the library
    - client services providing intermediary functions between the browser and the other parts of the library (e.g. deciding where to search, interpreting information structured as digital objects, managing relationships between digital objects and converting among the protocols used by the various parts of the system)
- the repository stores and manages digital objects and associated information
    - a digital library may have different types of repositories: modern repository, legacy databases and Web servers
- the handle system provides a distributed directory service for handles of digital library resources
    - input to the handle system is the handle (identifier) of a digital object of interest
    - output of the handle system is the identifiers of the repositories where the digital object of interest is stored
- the search system houses various indexes and catalogs that can be searched in order to discover information before retrieving it from a repository
    - searching is carried out by specially designed Web-based retrieval systems that are capable of accessing and retrieving digital objects across distributed repositories
    - distributed searching involves *federating* (i.e. mapping together) similar digital objects from different sources in a way that makes them appear as one organized collection

1.4. Current Status of Development of Digital Libraries

1.4.1. Examples of Digital Libraries

- many digital libraries are now under development that aim to support educational, research and government activities
    - digital libraries are currently in the early stage of development

- many existing digital libraries are prototypes or testbeds associated with research initiatives that aim to develop cataloging models, search methodology and query protocols
- all existing digital libraries are under continuous development as the concepts evolve and technologies advance
  - different organizations may define digital library in different ways
    - as a result, the objectives and scopes of different digital libraries can vary considerably between one another
  - examples of digital libraries
    - Library of Congress Online Catalog
    - Library of Congress American Memory
    - NOAA Library and Information Network
    - National Library of Australia
    - The British Library
    - Electronic Collection, National Library of Canada
    - Bibliotheque nationale de France
    - Digital Library Network, Japan
    - National Institute for Informatics, Japan

1.4.2. The Digital Libraries Initiative (DLI)

- this is a United States government initiative sponsored by the National Science Foundation (NSF), the Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA)
  - the focus was to advance the means to collect, store and organize information in digital forms, and make it available for searching, retrieval and processing across communication networks
  - between 1994 and 1998, six DLI projects were funded as research testbeds (Table 1)
    - a testbed is a prototype system with real collections and real users, but support as a research project rather than an operational application

1.4.3. The Digital Libraries Initiative Phase 2 (DLI-2)

- DLI-2 is built on the success of DLI
- in addition to the original sponsors of DLI, DLI-2 is also supported by the National Library of Medicine (NLM), the Library of Congress and the National Endowment for the Humanities (NEH) and others
- the objectives of DLI-2 are
  - to provide leadership in research fundamental to the development of the next generation of digital libraries in such areas as education, engineering and design, earth and space sciences, biological sciences, geography, economic, and the arts and humanities
  - to advance the use and usability of globally distributed, networked information resources
  - to encourage existing and new communities to focus on innovative application areas
- DLI-2 includes an international digital libraries initiative that aims to foster international

cooperation in the development of systems that can operate in multiple languages, formats, media and social and organizational contexts

---

2. Digital Libraries for Geospatial Information

2.1. The Needs for Spatial Digital Libraries

- geospatial data are an important component of the information holdings of many academic, research and government organizations
    - geospatial data have been increasingly used in decision making in business, resource planning and environmental management
    - geospatial data are now increasingly available in the digital form
        - it is not always easy for users to know where digital geospatial data exist
        - it is not always easy for users to know whether existing geospatial data meet their application needs
        - it is not always easy for users to know how they can combine their own geospatial data with external sources for value-added applications
    - the objective of spatial digital libraries is to foster the use of geospatial data by helping potential users to access, evaluate and retrieve geospatial in public domain or obtainable from commercial data suppliers
- geospatial data require special treatment (indexing and federated searching) in digital libraries because
    - they are both graphics- and text-based
    - they are searchable by key words (e.g. place names) and by location (e.g. coordinates)
    - they can be combined and cross-referenced only if they meet certain compatibility requirements (e.g. scale, classification, cartographic symbology, revision cycle)
- two of the six original DLI testbeds had a significant geospatial data focus
    - Environmental Electronic Library, University of California, Berkeley
    - Alexandria Digital Library, University of California, Santa Barbara
    - the objective of these two testbeds is to determine the optimum data model, cataloging and indexing methodology as well as query protocols for geographically referenced data, including maps and remote sensing imagery

2.2. Examples of Digital Spatial Libraries

- many spatial digital libraries have been developed or are currently under development
    - most of these spatial digital libraries are currently have a strong focus on cataloging
        - they are mainly designed to serve the purpose of telling users what is available in a certain server or repository
        - they offer relatively limited capability in federated spatial searching
        - search can usually be conducted for specific geographical areas only

examples of spatial digital libraries in operation
- international cooperative project
  - <u>Consortium for International Earth Science Information Network</u> (CIESEN), Columbia University
- spatial digital libraries in the United States
  - National Archives and Records Administration
  - USGS Earth Observation System (EROS) Data Center Distributed Active Archive Center (EDC DAAC)
  - USGS National Biological Information Infrastructure
  - USGS Global Land Information System
  - USGS National Mapping Information
  - National Atlas of the United States
  - Microsoft TerraServer
- Spatial digital libraries in other countries
  - National Atlas of Canada
  - University of Waterloo Electronic Library
  - University of Edinburgh Data Library
  - KINDS (Knowledge-based Interface for National Data Sets), United Kingdom
  - Geographical Survey Institute, Japan

3. Summary

- the development of digital libraries will make the use of the Web more efficient and intuitive
- digital libraries represent a major component of the national information infrastructure that aims to connect every citizen to the information superhighway
- spatial digital libraries will greatly facilitate public access to geographic information

4. Review and Study Questions

1. The following table lists the general characteristics of traditional libraries. Complete the table by noting the characteristics of digital libraries:

Characteristics	Traditional libraries	Digital libraries
Location of library collections	Centralized	
Mode of use	User must visit library physically	

Access to document	A document can be accessed by one user only at a time	
Rare and fragile documents	Restricted use due to preservation concerns	
Combining and cross-referencing documents	Difficult to combine and cross-reference documents	
Scope of library collections	Library may limit scope of collections to specific topics and/or geographical areas	
Size of library collections	Limited by physical size of library	

2. Explain the role that the World Wide Web plays in the development of digital libraries. What impacts do you expect digital libraries will have on the use of the World Wide Web in future?

5. References

Arms, William Y. (1995) *Key Concepts in the Architecture of the Digital Library, D-Lib Magazine*, July, 1995.

Blanchi, Christophe, William Y. Arms, Edward A. Overly, (1997) An Architecture for Information in Digital Libraries *D-Lib Magazine*, February 1997.

Rusbridge, Chris. (1998) Towards the Hybrid Library, *D-Lib Magazine*, July/August, 1998.

ESRI (Environmental Systems Research Institute) (1994) *GIS Approach to Digital Spatial Libraries*, White paper Series, Environmental Systems Research Institute, Redlands, CA.

Goodchild, M.F. (1995) Alexandria Digital Library (Report on a Workshop on Metadata), Santa Barbara, CA, posted at http://www.alexandria.ucsb.edu/public-documents/metadata/metadata_ws.html

Griffin, S.M. (1988) NSF/DARPA/NASA Digital Libraries Initiative: A Program Manager's Perspective, *D-Lib Magazine*, July/August 1998.

Harder, C. (1998) *Serving maps on the Internet*, Environmental Systems Research Institute, Redlands, CA.

Lopez, X.R. (1997) The Network as Organization: Digital Libraries for Spatial Information, paper presented at UCGIS Annual Assembly and Summer Retreat, posted at http://www.spatial.maine.edu/ucgis/testproc/lopez/xlopez.html

Miller, J.S. (1996) W3C and Digital Libraries, *D-Lib Magazine*, November 1998.

Pinfield, Stephen, Jonathan Eaton, Catherine Edwards, Rosemary Russell, Astrid Wissenburg, Peter Wynne (1998) Realizing the Hybrid Library, *D-Lib Magazine*, July/August 1998.

Schatz, B. and Chen, H. (1999) Digital Libraries: Technological Advances and Social Impacts, *IEEE Computer*, vol. 32, no. 2, pp. 45-50.

Schatz, B., Mischo, W., Cole, T., Bishop, A., Harum, S., Johnson, E., Neumann, L., Chen, H. and Ng, D. (1999) Federated Search of Scientific Literature, *IEEE Computer*, vol. 32, no. 2, pp. 51-59.

Wiederhold, G. (1995) Digital Libraries, Value and Productivity, *Communications of the ACM*, vol. 38, no. 4, pp. 85-86.

Citation

To reference this material use the appropriate variation of the following format:

Albert K. Yeung. (1999) Digital Libraries, *NCGIA Core Curriculum in GIScience*, http://www.ncgia.ucsb.edu/giscc/units/u191/u191.html, accessed [today's date].

Created: January 15, 1999. Last revised: August 6, 2000.

Unit 191 - Digital Libraries

Table of Contents

Back to the Unit

Unit 191 - Digital Libraries

Metadata and Revision History

Page contents

---

1. About the main contributor

- author: Albert K. Yeung, Ontario Ministry of Northern Development and Mines, Canada

2. Details about the file

- unit title: Digital Libraries
- unit key number: Unit 191

3. Key words

4. Index words
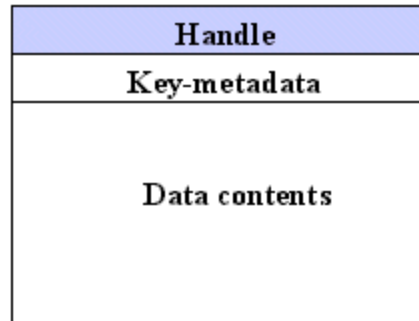
5. Prerequisite units

6. Subsequent units

7. Revision history

- First draft completed: January 15, 1999
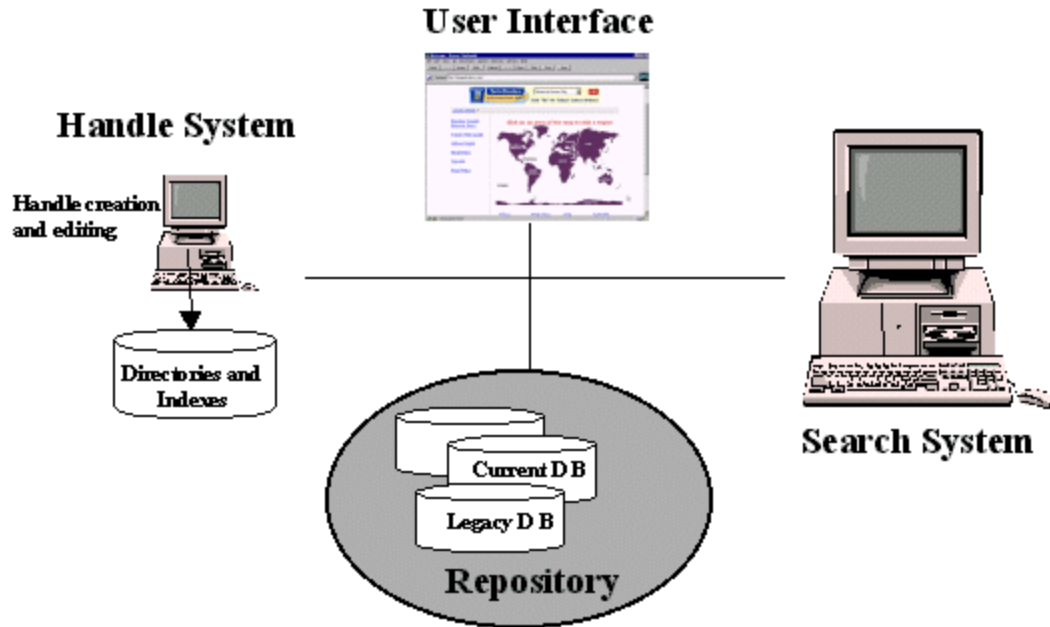- Posted to website: August 6, 2000

Back to the Unit

Figure 1: Components of a Digital Object

Handle
Key-metadata
Data contents

A.K. Yeung 1999-02-05 u148-18

Figure 2: Architecture of an Ideal Digital Library

A.K. Yeung 1999-02-05 u148-19

Table 1: DLI Testbeds

Testbed sites	Focus and objectives of investigation
<u>University of California at Berkeley</u>	Environmental planning and geographic information systems (GIS)
<u>University of California at Santa Barbara</u>	Spatially-referenced map information
<u>Carnegie-Mellon University</u>	Informedia digital video library
<u>University of Illinois at Urbana-Champaign</u>	Federating repositories of scientific literature
<u>University of Michigan</u>	Intelligent agents for information location
<u>Stanford University</u>	Interoperation mechanisms among heterogeneous services